AN INFORMAL APPROACH TO FORMAL INNER PRODUCTS

Richard Hammack

The fact that linear algebra is a distillation of our perception of physical space gives teachers and students of the subject a vast source of heuristic arguments. Almost every concept or result in finite-dimensional linear algebra can be illustrated or explained with a well-chosen geometric figure or schematic diagram.

The formal definition of an inner product seems to be an exception. Students who appreciate geometric reasoning are often mystified when, usually in their second linear algebra course, they first encounter this definition. They agree that an inner product gives a space a sense of length and orthogonality, but they are perplexed by the fact that its formal properties seem to have no geometric rationale. "Why," they ask, "is the definition phrased the way it is?"

In teaching a second-semester course recently, I devised an answer to that question. My strategy is the reverse of the standard one. I begin with intuitive notions about length and orthogonality, and derive the formal inner product from them, rather than the other way around. This article is an outline of my approach.

When introducing the formal definition of an inner product, I first remind the students of the standard inner product on $\mathbb{R}^n$. They have been exposed to it in their previous courses, and have an understanding of how it gives $\mathbb{R}^n$ metric properties. Then I give the formal definition of an inner product, pointing out that the standard inner product on $\mathbb{R}^n$ satisfies this definition.

An *inner product* on a vector space V over a field $\mathbb{F}$, where $\mathbb{F}$ is either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$, is a function $\langle\ ,\ \rangle\colon V \times V \to \mathbb{F}$ with the following properties. If $u, v, w \in V$, and $c \in \mathbb{F}$, then

1. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$
2. $\langle cv, w\rangle = c\langle v, w\rangle$
3. $\langle w, w\rangle > 0$ if $w \neq 0$
4. $\langle v, w\rangle = \overline{\langle w, v\rangle}$.

Further, from properties 1, 2, and 4, it follows that

5. $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$
6. $\langle v, cw\rangle = \overline{c}\langle v, w\rangle$.

The bar represents complex conjugation, which is superfluous when $\mathbb{F} = \mathbb{R}$ [1].

To explain the reasons behind properties 1–6, I tell my students to first consider the case of real vector spaces, that is to assume V is a finite dimensional vector space over $\mathbb{F} = \mathbb{R}$. (We momentarily postpone the complex case $\mathbb{F} = \mathbb{C}$.) I ask them to imagine that V is magically endowed with length and orthogonality. I then explain that we are going to make three intuitively reasonable assumptions about the behavior of this length and orthogonality, and that these assumptions will lead naturally to a function $\langle\ ,\ \rangle\colon V \times V \to \mathbb{R}$ satisfying properties 1–6.

The first assumption involves length. Intuitively, the length of a vector w in V should be a nonnegative real number which we will denote as $\|w\|$. If $w \neq 0$, we view the span of w as a copy of the real line in V. Multiplying w by a number $c \in \mathbb{R}$ scales w along this line, and we expect that the length of w gets scaled by a factor of $|c|$. This means $\|cw\| = |c|\ \|w\|$. This is our first assumption.

Assumption A: There is a *length function* $\|\ \|\colon V \to \mathbb{R}$ satisfying $\|w\| > 0$ whenever $w \neq 0$, and $\|cw\| = |c|\ \|w\|$ for all $w \in V$ and $c \in \mathbb{F}$.

Next, consider orthogonality. Our experience with physical space suggests that two vectors v and w can be orthogonal, and we express this relationship as $v \perp w$. Intuition demands that a nonzero vector w is not orthogonal to itself, nor to any of its nonzero scalar multiples. Moreover, the set $\operatorname{span}(w)^\perp$ of all vectors orthogonal to every element of $\operatorname{span}(w)$ is a subspace of V having dimension $\dim(V) - 1$. Therefore, $V = \operatorname{span}(w) \oplus \operatorname{span}(w)^\perp$, and the following assumption is reasonable.

Assumption B: There is a symmetric relation $\perp$ on V with the property that for any vector $w \in V$, the set $\operatorname{span}(w)^\perp = \{v \in V | v \perp cw \text{ for all } c \in \mathbb{F}\}$ is a subspace, and $V = \operatorname{span}(w) \oplus \operatorname{span}(w)^\perp$.

A *triangle* in V can be thought of as two independent vectors v and w (regarded as "sides"), plus the additional side $v - w$. A triangle is *right* if two of its sides are orthogonal. Our final assumption is that similarity of right triangles is a meaningful concept.

Assumption C: If two right triangles share a common (nonright) angle, then they are similar in the usual Euclidean sense. That is, ratios of corresponding sides are equal.

Suppose now that these assumptions hold for our finite-dimensional space V over $\mathbb{R}$. If $w \in V$ is a nonzero vector, then Assumption B guarantees the existence of a projection operator $E_w\colon V \to V$ whose range is the one-dimensional space span(w), and whose kernel is the hyperplane span$(w)^\perp$. Notice that for any $v \in V$, the vector $E_w(v)$ is a scalar multiple of w. This observation leads to the definition of a function $\langle\ ,\ \rangle : V \times V \to \mathbb{F}$. We will subsequently verify that this function satisfies Properties 1–6, so we call it the *inner product* on V.

Definition 1: Given Assumptions A, B, and C, the *inner product on* V is the function $\langle\ ,\ \rangle : V \times V \to \mathbb{F}$ defined as follows. If w is zero, then $\langle v, w\rangle = 0$. Otherwise,

$$E_w(v) = \langle v, w\rangle \frac{w}{\|w\|^2}.$$

Here is the intuitive picture. The vector $E_w(v)$ is in the one-dimensional space span$(w) =$ span$(w/\|w\|^2)$. The inner product $\langle v, w\rangle$ is the real number that $w/\|w\|^2$ must be multiplied by to equal $E_w(v)$. This is illustrated in Figure 1.

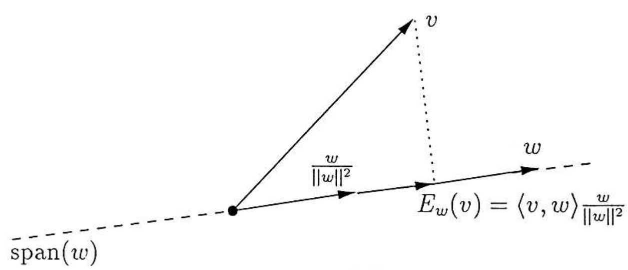

Figure 1.

The number $\langle v, w\rangle$ gives information about the spatial relationship between v and w. Notice that $\langle v, w\rangle = 0$ if and only if $v \in \ker(E_w)$, meaning $v \in$ span$(w)^\perp$, so v is orthogonal to w. Also, since $w = E_w(w) = \langle w, w\rangle w/\|w\|^2$, we have $\langle w, w\rangle = \|w\|^2$.

Now we verify that $\langle\ ,\ \rangle$ satisfies properties 1–4. Properties 1 and 2 are just rephrasings of the fact that the projection E_w is linear. By Definition 1, properties 1 and 2 are obviously valid when $w = 0$, so let us assume $w \neq 0$. To justify property 1, just notice that since $E_w(u + v) = E_w(u) + E_w(v)$, Definition 1 gives

$$\langle u + v, w\rangle \frac{w}{\|w\|^2} = \langle u, w\rangle \frac{w}{\|w\|^2} + \langle v, w\rangle \frac{w}{\|w\|^2},$$

so $\langle u+v,w\rangle = \langle u,w\rangle + \langle v,w\rangle$. Likewise, since $E_w(cv) = cE_w(v)$, we have

$$\langle cv,w\rangle \frac{w}{\|w\|^2} = c\langle u,w\rangle \frac{w}{\|w\|^2},$$

so $\langle cv,w\rangle = c\langle v,w\rangle$, which is property 2.

Property 3 states the intuitively obvious fact that the projection of a nonzero vector to its own span produces a nonzero vector. If w is nonzero, then

$$w = E_w(w) = \langle w,w\rangle \frac{w}{\|w\|^2},$$

which means $\langle w,w\rangle = \|w\|^2$, so $\langle w,w\rangle > 0$ by Assumption A.

We will next prove Property 4, which, in our case $\mathbb{F} = \mathbb{R}$, states that $\langle v,w\rangle = \langle w,v\rangle$. This property is certainly true if $v \perp w$, because both $\langle v,w\rangle$ and $\langle w,v\rangle$ are zero. Thus, assume that v and w are not orthogonal, that is neither $E_v(w)$ nor $E_w(v)$ is zero. There are two possible configurations that can happen here, illustrated in Figures 2A and 2B. In Figure 2A, $E_w(v)$ and $E_v(w)$ are both *positive* scalar multiples of w and v, respectively, and in Figure 2B they are both *negative* scalar multiples. Thus, $\langle v,w\rangle$ and $\langle w,v\rangle$ are either both positive or both negative. In either case, Assumption C gives

$$\frac{\|E_w(v)\|}{\|v\|} = \frac{\|E_v(w)\|}{\|w\|},$$

and using Definition 1 and Assumption 1, this becomes

$$\frac{|\langle v,w\rangle|}{\|w\|\ \|v\|} = \frac{|\langle w,v\rangle|}{\|v\|\ \|w\|}.$$

From this it follows that $\langle v,w\rangle = \langle w,v\rangle$, which is property 4 in the case $\mathbb{F} = \mathbb{R}$. (It should be mentioned that Assumption C suggests right triangles conform to our usual Euclidean expectations. For instance, in Figure 2A, $E_v(w)$ is not a *negative* scalar multiple of v. Assumption C can be made more mathematically precise, but its informal wording is better suited to classroom exposition.)

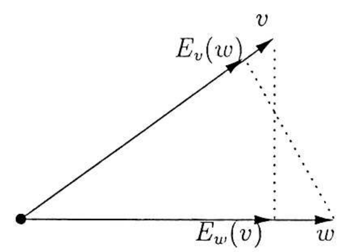

Figure 2A.

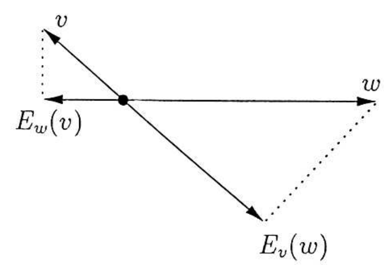

Figure 2B.

We have now justified properties 1–4 for real vector spaces (that is for $\mathbb{F} = \mathbb{R}$), and properties 5 and 6 follow immediately from these.

The remainder of this article concerns the complex case $\mathbb{F} = \mathbb{C}$. It turns out that in this case Assumptions A, B, and C are still reasonable, but we need to be thoughtful about their interpretation. If w is a nonzero vector, then we view span(w) as a copy of the complex plane V. Multiplication of w by a scalar $c = re^{i\theta} \in \mathbb{C}$ should be a rotation of w by *theta* radians in this plane, followed by a scaling by the factor $r = |c|$ (see Figure 3). Thus, as in the case $\mathbb{F} = \mathbb{R}$, we expect that $\|cw\| = |c|\ \|w\|$, which is the content of Assumption A.

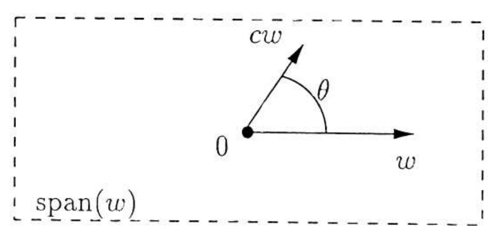

Figure 3.

Likewise, Assumption B still makes sense as long as we agree that a nonzero vector should never be orthogonal to any nonzero multiple of itself. This leads to what is, at first glance, a somewhat less intuitive phenomenon: Assumption B implies that

w and iw are *not* othogonal even though they may "look" orthogonal in span(w). This is actually reasonable. Intuition suggests that a nonzero scalar multiple of a vector w should not be orthogonal to w. Thus, in asserting that Assumption B holds for complex vector spaces, we are abstracting to $\mathbb{C}^n$ this "obvious" fact about $\mathbb{R}^n$.

Again, we define the inner product $\langle\ ,\ \rangle$ by Definition 1. The only difference is that now $\langle v, w\rangle$ can be a complex number. Given $v, w \in V$, $E_w(v)$ is in the one-dimensional space $\text{span}(w) = \text{span}(w/\|w\|^2)$. The inner product $\langle v, w\rangle$ is the scalar (possibly a complex number) that $w/\|w\|^2$ must be multiplied by to equal $E_w(v)$. This is illustrated in Figure 4. (Compare Figures 1 and 4.)

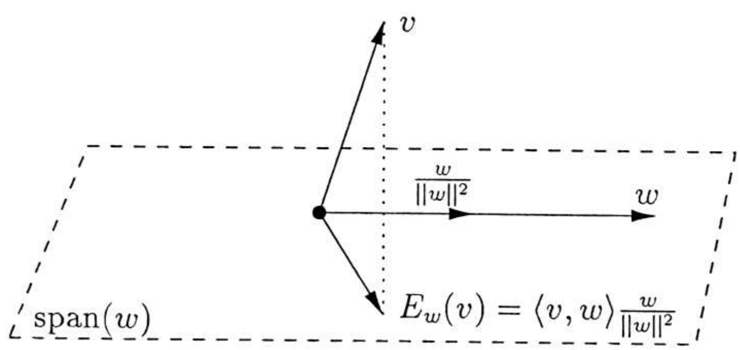

Figure 4.

Properties 1, 2, and 3 follow exactly as in the real case, but it takes a little more work to justify property 4. To do this, we first derive property 6.

Property 6 is clearly true if either w or c is zero. Suppose then that neither c nor w is zero. Notice that the projections E_w and E_{cw} are equal, for they are both projections onto the space $\text{span}(w) = \text{span}(cw)$, and both have null space $\text{span}(w)^\perp = \text{span}(cw)^\perp$. Thus, $E_w(v) = E_{cw}(v)$, which by Definition 1 means

$$\begin{aligned}\langle v, w\rangle \frac{w}{\|w\|^2} &= \langle v, cw\rangle \frac{cw}{\|cw\|^2} = \langle v, cw\rangle \frac{cw}{|c|^2\|w\|^2} \\ &= \langle v, cw\rangle \frac{cw}{c\overline{c}\|w\|^2} = \frac{\langle v, cw\rangle}{\overline{c}} \frac{w}{\|w\|^2}\end{aligned}.$$

(Assumption B was used in the second equality.) Looking at the first and last terms in this expression, we see $\langle v, cw\rangle = \overline{c}\langle v, w\rangle$, which is property 6.

Finally, we turn to property 4. First, we are going to show first that property 4 is true when $\langle v, w\rangle$ is real, in which case we want to show $\langle v, w\rangle = \langle w, v\rangle$. Certainly this is true when $v \perp w$, for then $\langle v, w\rangle = 0 = \langle w, v\rangle$. So suppose v and w are not orthogonal, that is neither $\langle v, w\rangle$ nor $\langle w, v\rangle$ is zero. Now, since $\langle v, w\rangle$ is a nonzero real number, we have a situation that is illustrated in Figure 2, and as in the case $\mathbb{F} = \mathbb{R}$, this gives $\langle v, w\rangle = \langle w, v\rangle$.

In general, $\langle v, w\rangle$ will be a nonreal number, and we want to verify $\langle v, w\rangle = \overline{\langle w, v\rangle}$. Since

$$\left\langle \overline{\langle w, v\rangle} w, v \right\rangle = \overline{\langle w, v\rangle}\langle w, v\rangle \in \mathbb{R},$$

what was said in the previous paragraph implies

$$\left\langle \overline{\langle w, v\rangle} w, v \right\rangle = \left\langle v, \overline{\langle w, v\rangle} w \right\rangle.$$

Combining this observation with property 6 gives

$$\overline{\langle w, v\rangle}\langle w, v\rangle = \left\langle \overline{\langle w, v\rangle} w, v \right\rangle = \left\langle v, \overline{\langle w, v\rangle} w \right\rangle = \langle w, v\rangle\langle v, w\rangle.$$

Comparing the first and last expressions of this equality, we get the desired result $\langle v, w\rangle = \overline{\langle w, v\rangle}$.

We have now confirmed that Assumptions A, B, and C lead to a function $\langle\ ,\ \rangle$ satisfying properties 1–6. On the other hand, given a function $\langle\ ,\ \rangle$ satisfying properties 1–6, and defining $\|v\| = \sqrt{\langle v, v\rangle}$, and $v \perp w$ if $\langle v, w\rangle = 0$, one easily verifies Assumptions A, B and C. Consequently, these assumptions are equivalent to the existence of an inner product $\langle\ ,\ \rangle$.

I believe this geometric presentation of inner products gives students a firmer conceptual ground on which to build than does the standard axiomatic presentation. I can introduce inner products from this perspective in a 50-minute class, and still have some time left over to talk about some related topics, such as the Cauchy-Schwartz inequality. I have had success with this method, and hope that other instructors find it useful.

Mathematics Subject Classification (2000): 15-01

Reference

1. K. Hoffman and R. Kunze, *Linear Algebra*, second edition, Prentice Hall, 1971.

Richard Hammack
Department of Mathematics
Randolph-Macon College
P. O. Box 5005
Ashland, Virginia 23005
email: rhammack.rmc.edu