Inner product and orthogonality

Far and Near in Vector Spaces

by Javier Rodríguez Laguna, Univ. Carlos III de Madrid

Last Update: Feb 28, 2008

Scalar Product and Distance

If we want to do geometry, we need the notion of distance. So, we need no excuse to study this notion. But we will generalize it so much in this chapter, that you may come to think why do we take such a disturbance... In a more abstract setting, sometimes we want to approximate a complex object by a simpler representative. In that case, we should choose the representative whose distance to the original object is minimum.

But let's start out with our old and dear $\mathbf{R}^n$. The **scalar product** (also called *dot product*) of two vectors u and v is defined as

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

This is the obvious generalization of the dot product in $\mathbf{R}^2$ and $\mathbf{R}^3$. It is rather obvious that $u \cdot v = u \cdot u$, right? And also some other things, like $u \cdot u \geq 0$ (see why?). This is rather important, because the **length** (or **norm**) of a vector v in $\mathbf{R}^n$ is defined by

$$\|v\|^2 = v \cdot v = \sum_{i=1}^{n} v_i^2$$

Why that? In $\mathbf{R}^2$ and $\mathbf{R}^3$ it came natural, through Pythagoras theorem. So it's just the natural generalization, right?

A **unit vector** is one with unit length. Any vector can be **normalized**, dividing by its norm. See that for a unit vector only the *direction* is important.

E1. (Rather trivial one, just to check out) Pick up a vector of $\mathbf{R}^4$. For example, $(2, 1, 0, -2)$. Now, compute its length and normalize it.

E2. Prove that the dot product between two vectors can be written like $u \cdot v = u^T v$.

If u and v are the vectors representing two points in any vector space, then $v - u$ is the vector which takes from u to v. That's clear, no? Then, the **distance** between u and v is the length of the difference vector:

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

If you expand it, it means in $\mathbf{R}^3$:

$$d(u, v) = \sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2 + (v_3 - u_3)^2}$$

Orthogonality and Angles

Normal people would say *perpendicular*, but mathematicians are like this, you know, and will prefer *orthogonal*. It's just the same.

I give you two vectors u and v. In symbols, we'll say $u \perp v$ when they are orthogonal. But, how to know when they are? Take a look at figure 3. We have drawn u and v, orthogonal. But we have also depicted $-v$. It's just geometry: if u and v are orthogonal, then $d(u, v) = d(u, -v)$. Take a look and convince yourself...

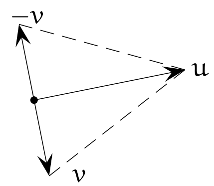

Figure 3. When $u \perp v$, $d(u, v) = d(u, -v)$.

Now, a little bit of algebraic manipulation, and we'll find the condition for orthogonality. This idea of $d(u, v) = d(u, -v)$ is the key!

$$d(u, v)^2 = \|u - v\|^2 = (u - v) \cdot (u - v) =$$

Using the distributive property we see that:

$$= u \cdot u + v \cdot v - 2u \cdot v = \|u\|^2 + \|v\|^2 - 2u \cdot v$$

OK, no more simplification can be done. Now, let's check $d(u, -v)$:

$$d(u,-v)^2 = \|u-(-v)\|^2 = (u+v)\cdot(u+v) = \|u\|^2 + \|v\|^2 + 2u\cdot v$$

Take a look at both expressions. They are the same if and only if $u\cdot v = 0$!! So, we conclude that two vectors are orthogonal when their inner product is zero. Moreover, we have a very intrincate proof of Pythagoras theorem! If $u \perp v$, $d(u,v) = \|u\|^2 + \|v\|^2$...

E3. Find a vector which is orthogonal to $(1,2,-2,0)$.

For sure you remember from $\mathbf{R}^2$ and $\mathbf{R}^3$ the old formula:

$$u\cdot v = \|u\|\cdot\|v\|\cdot\cos(\alpha)$$

where α is the angle between u and v. This formula can be extended to any vector space with inner product, thus defining angles in a very general setting. Notice that, if two vectors are orthogonal, then $\alpha = \pi/2$, thus $\cos(\alpha) = 0$ and, therefore, $u\cdot v = 0$, as we said before. Angles will not play a very important role here, but at least you should be able to find them...

Orthogonal Complement

For a vector v, its **orthogonal complement** is the set of vectors which are orthogonal to it. It happens to be a subspace of the full vector space, of course. For example, let's find the orthogonal complement of vector $(1,0,-1)$ in $\mathbf{R}^3$. Let us consider an arbitrary vector (x,y,z). The condition for orthogonality is

$$0 = (1,0,-1)\cdot(x,y,z) = x - z$$

So we have a plane, $x-z=0$. In general, if W is a subspace, its orthogonal complement $W^\perp$ is the set of vectors which are orthogonal to all vectors in W. Uff... this requires an example, at least!

For example, find the orthogonal complement of the plane π, given by $x-z=0$ in $\mathbf{R}^3$. This is the set of all vectors which are perpendicular to the plane. A plane is determined by two directions. Therefore, in $\mathbf{R}^3$, there is a single direction left. This means that the orthogonal complement of a plane must be a line. Which line? OK, the equation $x-z=0$ is a way of saying that all vectors of the plane are perpendicular to $(1,0,-1)$. Therefore, $\pi^\perp = \mathrm{Span}(1,0,-1)$.

Q1. If you're working on a vector space of dimension n and W is a subspace of dimension d, which is the dimension of $W^\perp$?

E4. Find in $\mathbf{R}^3$ the orthogonal complement of the subspace given by $\mathrm{Span}(1,2,-3)$.

Now, what to do for more general subspaces? For example, a plane π spanned by vectors a_1 and a_2, $\pi = \mathrm{Span}(a_1,a_2)$. Then, we are asked for the set of vectors X which are orthogonal to all vectors of the plane. An equivalent condition is to ask for the set of vectors which are orthogonal both to a_1 and a_2:

$$\begin{cases} a_1\cdot X = 0 \\ a_2\cdot X = 0 \end{cases}$$

As usual, we call A the matrix which contains a_1 and a_2 as columns. So, A^T is the matrix whose rows are a_1 and a_2, right? Then, the previous set of equations is equivalent to

$$A^T X = 0$$

(check it!) So, we only have to solve that homogeneous system in order to obtain the orthogonal complement of any subspace!

E5. Find the orthogonal complement of the subspace spanned by the vectors $(1,0,0)$ and $(1,1,0)$. Is it reasonable?

Generalizing the Scalar Product

This notion of distance can be very useful also outside geometry. For example, imagine that we're working on a vector space of functions. We have a complicated function, like $\cos(\sqrt{e^x+1})$ or something like that. We wish we had a polynomial, because polynomials are nice both for analytical and for numerical work, instead of that monster. The monster *is not* a polynomial, no matter how much you look at it. *But* you can find the polynomial, of a given degree, which is *closest* to that function. That's where we are going to, ok?

This means that we should find out what is the *essence* of distance, in order to generalize it. Experience has shown us that the most appropriate concept to generlize was that of dot product, which we promote to something we will call **inner product**. We'll give now a rather formal definition of that thing. Let V be any vector space,

• A binary operation "$\langle u, v\rangle$" mapping $V \times V$ into $\mathbf{R}$ on a vector space is an **inner product** if it fulfills the following conditions:

1. Symmetry: $\langle u, v\rangle = \langle v, u\rangle$.
2. Distributive property: $\langle (u+v), w\rangle = \langle u, w\rangle + \langle v, w\rangle$.
3. Product with numbers: $\langle cu, v\rangle = \langle u, cv\rangle = c\,\langle u, v\rangle$.
4. Positivity: $\langle u, u\rangle \geq 0$, and $\langle u, u\rangle = 0$ if and only if $u = 0$.

For example, we can define a new inner product on $\mathbf{R}^2$:

$$\langle u, v\rangle \equiv 4u_1v_1 + 5u_2v_2$$

Who says that's not a fine inner product? It is, you can check all four properties. But it is also hard to imagine how come it can be useful for anything. Believe me: it can. We will reserve the dot, $u \cdot v$ for the old and venerable scalar product in $\mathbf{R}^n$.

Inner Product in Functional Spaces

OK, then, let us play a little bit with a vector space of real functions, defined on a certain interval $[a, b]$. Remember a little bit of naming conventions. We call $C[a, b]$ the vector space of continuous real functions on $[a, b]$, ok?

We're practical people, so we can imagine the functions to be "discretized". This means that we only consider the values at some points, $x_i = a$, equally spaced on $[a, b]$. Let's assume that there are n of them. So the function is completely described by giving the values at those n points: $\{f(x_i)\}$, ok? Look at figure 1 for a glimpse. We can imagine all those n values displayed on a very long vector:

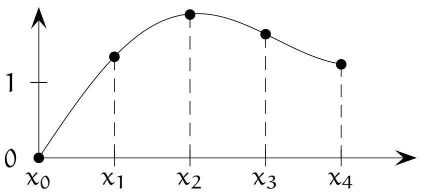

Figure 1. A discretized function.

$$f = (f(x_1), f(x_2), f(x_3), \cdots, f(x_n))$$

The vector is equivalent to the function! So, now let's consider that we have two functions, f and g. We can start by defining their inner product employing just the normal $\mathbf{R}^n$ dot product:

$$\langle f, g\rangle = \sum_{i=1}^{n} f(x_i) \cdot g(x_i)$$

For the same price, we can multiply the product into Δx, and get another inner product which is equivalent:

$$\langle f, g\rangle = \sum_{i=1}^{n} f(x_i) \cdot g(x_i) \Delta x$$

This last inner product has a nice advantage. When you take the limit $n \to \infty$, to reach the continuum, then the sum becomes an integral!

$$\langle f, g\rangle \to \int_a^b f(x)g(x)dx$$

This way we can define the "length" of a function:

$$\|f\|^2 = \int_a^b [f(x)]^2 dx$$

which is, obviously, greater than zero.

E6. Calculate the inner product of $f(x) = x$ and $g(x) = x - 1$ as functions of $C[0, 1]$.

I would love to show you applications of this fast fast, but let's take it easy. Consider the vector space $C[0, 1]$. In them, we select the function $f(x) = \cos(x)$, and functions $g(x) = \sin(x)$ and $h(x) = 1 - x^2/2$. Which one is closer to f? If you had to substitute f for one of them, which one to pick? Just imagine that they're points in $\mathbf{R}^2$. Then you would find out how much is the distance from f to g, $\|f - g\|$, and compare it to $\|f - h\|$. Let's do it!

$$\|f - g\|^2 = \int_0^1 (\cos(x) - \sin(x))^2 dx \approx 0.3$$

(do the integral, or just believe me for now). On the other hand,

$$\|f - h\|^2 = \int_0^1 (\cos(x) - 1 + x^2/2)^2 dx \approx 2 \cdot 10^{-4}$$

So, it is apparent that functions $\cos(x)$ and $1 - x^2/2$ are really close in $[0, 1]$. You may try to repeat the experiment on $[0, 2\pi]$ and it does not come

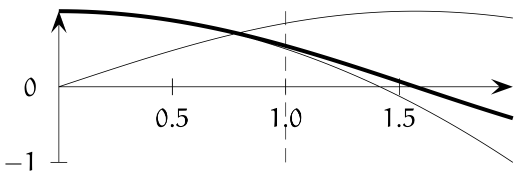

FIGURE 2. The functions $f(x) = \cos(x)$ (thicker line), $g(x) = \sin(x)$ and $h(x) = 1 - x^2/2$, in the interval $[0, 2]$. Notice that, in the interval $[0, 1]$, f and h are almost indistinguishable.

so nicely. Why? Simple! $1 - x^2/2$ is a Taylor approximation of $\cos(x)$ around $x = 0$, so it will only work for small intervals around that point... In figure 2 you may see all three functions and understand better...

Ah, it may seem to you that integration makes the process difficult, but this is not completely true. With computers, integration is a pretty trivial task, believe me. It is hard only when you work with pencil and paper. So, for many practical applications, when you reduce it to an integral... ok, you're done!

E7. Find, on $C[0, 1]$, the distance between $f(x) = 1$ and $g(x) = 1 + x$.

E8. Find the straight line passing through the origin which is closest to $\exp(x)$ on $[0, 1]$ Compare it with the Taylor approximation to first order. What do you deduce? This is important!!

Functions can also be orthogonal, why not? :) Any couple of objects can, if only you have defined an inner product beforehand...

E9. Find a function in $C[0, 1]$ which is orthogonal to $f(x) = 1$. Caution: orthogonality of function

E10. Find the angle formed by $\sin(x)$ and $\cos(x)$ on $C[0, \pi/2]$ and on $C[0, 2\pi]$.

INNER PRODUCT IN COMPLEX VECTOR SPACES

When complex numbers appear, i.e.: $\mathbf{C}^n$ spaces, you need to take special care... The reason is that lengths should be real and positive numbers, right? In the case of $\mathbf{R}^n$, it's immediate, because the usual scalar product is the sum of square numbers. But in $\mathbf{C}^n$, the square of a number is not necessarily a real positive number, right? For example, $i^2 = -1$. What to do?

It's not difficult. The modulus of a complex number is always a real

positive number. So, $|i|^2 = 1$, that's ok. But, how to take that modulus thing to the scalar product? This is little bit more tricky.

There is nice formula you should take into account. For any complex number z, $z^*z = |z|^2$, where z^* is the complex conjugate of z. I remind you that the complex conjugate is obtained by substituting i by $-i$. For example, if $z = 3 + 4i$, $z^* = 3 - 4i$, that's all. Now, $|z|^2 = 3^2 + 4^2 = 25$, so $|z| = 5$. Why don't you check that you get the same result with the new formula? Anyway, we prove it in general. It's easy. If $z = a + ib$,

$$z^*z = (a - ib)(a + ib) = a^2 + a \cdot ib - ib \cdot a - ib \cdot ib = a^2 - i^2b^2 = a^2 + b^2 = |z|^2$$

So, now we state the rule for the inner product in $\mathbf{C}^n$: you take the complex conjugate of the first vector, and do as always! In other terms, if u and v are vectors of $\mathbf{C}^n$,

$$\langle u, v\rangle \equiv \sum_{i=1}^{n} u_i^* v_i$$

But you might say, then $\langle u, v\rangle \neq \langle v, u\rangle$, since we're taking the complex conjugate of only the first vector. And you would be fully right. It happens that $\langle u, v\rangle^* = \langle v, u\rangle$. Check it! So the properties of an inner product change a little bit when we have a complex vector space...

Let's have some practice... Let's compute the scalar product of $u = (1, i)$ and $v = (i, -1)$...

$$\langle u, v\rangle = (1)^* \cdot i + (i)^* \cdot (-1) = i + (-i) \cdot (-1) = 2i$$

So, they are *not* orthogonal!

E11. Given $u = (1 + i, 1 - i)$ and $v = (0, i)$, compute $\langle u, v\rangle$ and $\langle v, u\rangle$. What is the relation between them?

E12. Find a vector which is orthogonal to $v = (0, i)$.

ADJOINT OPERATOR

We defined the transpose of a matrix $(A^T)_{ij} = A_{ji}$, just by reflecting the elements on the diagonal. But there is a deeper meaning of the transpose. A deeper meaning deserves a catchier name, so we will call it *adjoint*, and it is related to the inner product.

For any operator A we define $A^\dagger$, the adjoint operator, such that

$$\langle A^\dagger u, v\rangle = \langle u, Av\rangle$$

If the inner product is just the usual dot product in $\mathbf{R}^n$, then this role is played by the transpose. You can check it as an exercise.

E13. Check that, in $\mathbf{R}^n$ with the usual dot product, the adjoint of an operator coincides with its transpose.

E14. Find the expression of the adjoint of an operator in $\mathbf{R}^2$, given by a matrix A_{ij}, with the inner product given by $\langle u, v\rangle = 2u_1v_1 + 3u_2v_2$.

If a matrix is symmetric, then $A^T = A$: $A_{ij} = A_{ji}$. The obvious generalization takes place when an operator coincides with its adjoint. Again, a new name is given, and we say the operator is **hermitean** or **self-adjoint**: $A^\dagger = A$.

E15. Consider the complex vector space $\mathbf{C}^2$. Find the hermitean conjugate of a general 2×2 matrix with complex entries. Do you think you can generalize?

What can be the interest of the adjoint operator? Same as the transpose in the old "$\mathbf{R}^n$+ dot product" case! And what is the interest of the transpose? Take a look back at the orthogonal complement section. We deduced that, for any operator (matrix) A, the solution of the equation $A^TX = 0$ is the set of vectors X which are orthogonal to the subspace spanned by the columns of A. In other words:

$$A^TX = 0 \qquad \Leftrightarrow \qquad X \perp \mathrm{Col}\,(A)$$