Inner Product Spaces and Adjoints

Aditya Sengupta

December 19, 2019

If you're reading this for Frenkel's Math 110, a lot of this is from the textbook (Friedberg-Insel-Spence), so please also read that!

1 Motivation

When we're talking about $\mathbb{R}^n$, we have the component-wise-vector view, i.e. that $a \in \mathbb{R}^n$ consists of a column with elements $a_1, a_2, \ldots, a_n \in \mathbb{R}$ that uniquely specify it, and we've done a good job making that more abstract so far. But we don't just have to think about it in terms of a collection of numbers; we also think about them geometrically. They're objects in some space and they have lengths and orientations. This is a useful way of thinking about vector spaces, mostly because it'll give us the ideas of orthogonality, norms, unitary transformations (ones satisfying $U^\intercal = U^{-1}$), and so on, all of which will make it way easier to apply linear algebra in different settings.

In particular, one motivating example is quantum mechanics, in which observable quantities are eigenvalues of operators: what you measure as momentum is an eigenvalue of the momentum operator, what you measure as energy is an eigenvalue of the energy operator, and so on. However, since quantum mechanics is a probabilistic theory, a "state" is a linear combination of different eigenstates, and according to the coefficients on those eigenstates, you may measure any of the possible eigenvalues with some probability. Suppose we had some state described by a wavefunction (the absolute value squared of a wavefunction is a probability density for a particle),

$$\psi(x) = c_1\psi_1(x) + c_2\psi_2(x) + \ldots$$

. Let's say we're trying to find the average energy, where each state has energy $E_1, E_2, \ldots$. (For simplicity, let's say everything is real.) This is done by computing the following integral:

$$\begin{aligned}
\langle E \rangle = \int_{-\infty}^{\infty} \psi(x)\hat{H}\psi(x)dx &= \int_{-\infty}^{\infty} (c_1\psi_1(x) + c_2\psi_2(x) + \ldots)(E_1c_1\psi_1(x) + E_2c_2\psi_2(x) + \ldots)\, dx \\
&= \int_{-\infty}^{\infty} \left(E_1c_1^2\psi_1^2(x) + (E_1+E_2)c_1c_2\psi_1(x)\psi_2(x) + E_2c_2^2\psi_2^2(x) + \ldots\right) dx \\
&= E_1c_1^2 + \left(\int_{-\infty}^{\infty} (E_1+E_2)c_1c_2\psi_1(x)\psi(x)dx\right) + E_2c_2^2 + \ldots
\end{aligned}$$

Intuitively, the answer you want is $E_1c_1^2 + E_2c_2^2 + \ldots$: a sum of all the possible energies, weighted by the probability of getting each one. The $\int \psi_j^2(x)dx$ terms behave nicely: they're probability densities, so if you take them out over all space they go to 1. But you also have the cross terms like $(E_1 + E_2)c_1c_2\psi_1(x)\psi_2(x)$. There's no chance you'll measure $E_1 + E_2$ (unless that happens to be another energy eigenvalue, in which case it'll have its own term later anyway), so you want that to go to 0.

It'd be really nice if the theory could guarantee that $\int_{-\infty}^{\infty} \psi_i(x)\psi_j(x)dx = 0$ whenever $i \neq j$, because then we could almost just eyeball things like expected energy or momentum just by looking at a wavefunction. Actually, you can do this, because the basis $\psi_i(x)$ is orthonormal: any two unequal elements integrate to 0, and any two equal elements integrate to 1 (because it's a probability density.)

This example might have been a lot, but it's a good illustration of how we can build up a nice mathematical theory to make an application much easier. Note that this was an example where we considered orthogonality over a vector space (of square-integrable functions $\mathbb{R} \to \mathbb{R}$) that wasn't $\mathbb{R}^n$: it helped us to think about this inherently geometric idea over objects that aren't as naturally thought of as geometric. When you think about a function, you don't usually picture it as a vector - in the sense of a line with a magnitude and direction - in a space. But thinking about it that way is really helpful! So what we'd like to build is a general way to think about elements of a vector space geometrically: how long are they, what are they orthogonal to, and so on.

2 Inner Products

We give a vector space V/F a notion of geometry by equipping it with an inner product: a function $\langle \cdot, \cdot \rangle : V \times V \to F$ that takes in two vectors in V and returns a scalar in F. This scalar, roughly speaking, indicates how much one vector is along another. In particular, if the two vectors are the same, we want to get the squared length of the vector.

For $\mathbb{R}^n/\mathbb{R}$, the inner product is

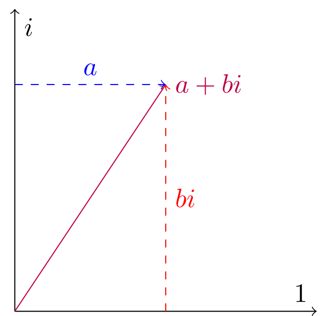
$$\langle x, y \rangle = \sum_{j=1}^{n} x_j y_j.$$

If we take $\langle x, x \rangle$, we just get the sum of squares of each component, which thanks to Pythagoras we know is what we want. Also, this inner product has a number of nice properties that we might want to carry over to more general inner products:

1. it satisfies scaling in both arguments: $\langle cx, y \rangle = \langle x, cy \rangle = c\langle x, y \rangle$.
2. it satisfies superposition in both arguments: $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ and $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.
3. it's commutative: $\langle x, y \rangle = \langle y, x \rangle$.
4. it gives non-negative norms (the result of taking an inner product with a vector and itself): $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \iff x = 0$.

Unfortunately, we can't perfectly carry these over if the field of our vector space isn't $\mathbb{R}$. Suppose our vector space was $\mathbb{C}$ over the field $\mathbb{C}$, and consider the norm of i: $\langle i, i \rangle = i^2 = -1$. This is saying that the i vector has squared length -1, which isn't geometrically reasonable. Let's think about how we can make this work.

By making a plot on the complex plane, we can figure out what we want the general norm of a complex number to be.



We see that the vector $a + bi$ is made up of two perpendicular components, one with length a and one with length b, so we want our length squared to be $a^2 + b^2$. Ideally, we'd like our new inner product to have something like the form of the inner product of $\mathbb{R}^n/\mathbb{R}$, so that it keeps most of the properties we listed above. Let's make that more concrete by saying that there is some other element of $\mathbb{C}$, say $c + di$, that we want to construct for each $a + bi$, such that

$$\langle a + bi, c + di \rangle = (a + bi)(c + di) = a^2 + b^2$$

where $c = c(a, b)$ and $d = d(a, b)$ implicitly. Note that this is still the sum-of-componentwise-products inner product, just that we have only one component. We get

$$(a + bi)(c + di) = (ac - bd) + i(ad + bc) = (a^2 + b^2) + i(0) \implies ac - bd = a^2 + b^2, ad + bc = 0$$

We've got two equations in two unknowns; you can either solve explicitly, or recognize that in order to match coefficients in the real-component equation we need $c = a$ and $d = -b$, and check that that gives an imaginary component of zero. What we've found is the definition of the complex conjugate:

$$(a + bi)^* = (a - bi)$$

So the form of our new norm on the complex numbers is $\langle x, x \rangle = \sum_{j=1}^{n} x_i x_i^*$. (Conventionally we put the conjugate second, but some definitions put it first). Extending this to the more general inner product, we get

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i^*$$

We've built an inner product that works for any vector space over $\mathbb{C}$ instead of just $\mathbb{R}$. A nice thing to notice is that by definition, real numbers are their own complex conjugate, so with a vector space over $\mathbb{R}$, this new definition is exactly what we had before!

The next question is: do we still satisfy all the nice properties from before?

1. scaling in both arguments?

$$\langle cx, y \rangle = \sum_{j=1}^{n} c x_j y_j^* = c \sum_{j=1}^{n} x_j y_j^* = c\langle x, y \rangle$$

So far so good, but what about the other one?

$$\langle x, cy \rangle = \sum_{j=1}^{n} x_j (c y_j)^* = \sum_{j=1}^{n} x_j c^* y_j^* = c^* \langle y, x \rangle$$

We don't have perfect linearity in both arguments, but we're close: we call this a <u>sesquilinear</u> form.

2. superposition in both arguments?

$$\langle x + y, z \rangle = \sum_{j=1}^{n} (x_j + y_j) z_j^* = \sum_{j=1}^{n} x_j z_j^* + \sum_{j=1}^{n} y_j z_j^* = \langle x, z \rangle + \langle y, z \rangle$$

and the other one,

$$\langle x, y + z \rangle = \sum_{j=1}^{n} x_j (y_j + z_j)^* = \sum_{j=1}^{n} x_j y_j^* + \sum_{j=1}^{n} x_j z_j^* = \langle x, y \rangle + \langle x, z \rangle$$

Superposition does work! This is because the complex conjugate satisfies superposition (as you can verify.)

3. commutativity?

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j y_j^* \neq \sum_{j=1}^{n} x_j^* y_j = \langle y, x \rangle$$

It seems like commutativity isn't satisfied, but we're close. We can use the properties that the complex conjugate satisfies superposition, that $(a^*)^* = a$, and that $(ab)^* = a^* b^*$ (again, all of which you can verify), to get

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j y_j^* = \sum_{j=1}^{n} (x_j^* y_j)^* = \left(\sum_{j=1}^{n} x_j^* y_j \right)^* = \langle y, x \rangle^*$$

I think there's a nice term for this like "conjugate commutativity", but I can't find the specific term at the moment.

4. non-negative norms?

   We built up this definition based on the fact that it would give us non-negative norms, so we can trust that that'll work.

We've successfully built the inner product for $\mathbb{C}^n/\mathbb{C}$, which has given us the general properties we'd like any vector space over $\mathbb{R}$ or $\mathbb{C}$ to have (we won't worry about fields that aren't those two.) This is how you get the definition of an inner product as given in Friedberg-Insel-Spence, or Axler, or any other common text. Note that those definitions will not specifically require distributivity in the second argument, because that comes out of distributivity in the first combined with "conjugate commutativity":

$$\langle x, y+z\rangle = \langle y+z, x\rangle^* = \langle y, x\rangle^* + \langle z, x\rangle^* = \langle x, y\rangle + \langle x, z\rangle$$

Some examples of valid inner products on other spaces are:

- over $C([0,1])$ (real-valued continuous functions over $[0,1]$): $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$.
- over $M_{n\times n}(F)$ ($n \times n$ matrices with elements from a field): $\langle A, B\rangle = \mathrm{tr}(B^*A)$; here B^* is the conjugate transpose, $(B^*)_{ij} = (B_{ji})^*$.

3 Orthogonality and Orthonormality

We say that two elements $\vec{v_1}, \vec{v_2}$ of an inner product space are orthogonal if $\langle \vec{v_1}, \vec{v_2}\rangle = 0$. Again, this extends the idea of being orthogonal in $\mathbb{R}^2$, where $\vec{a}\cdot\vec{b} = |a||b|\cos\theta$, so if $\theta = 90°$ the dot product is 0. This matches nicely the idea that an inner product reflects how much one vector is "along" another one. If two vectors have an inner product of 0, they have no component along one another - think back to the quantum mechanics example where we wanted ψ_1 to be completely separate from ψ_2 so that we wouldn't measure a "mixed" energy.

Orthogonality is useful because it allows us to use distributivity and kill most of the terms. Suppose we have a basis of V/F in which all the terms are pairwise orthogonal, i.e. $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$. Take an arbitrary element v, and expand it in this basis.

$$v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

To get the component along v_i, we take their inner product:

$$\langle v, v_i\rangle = c_1\langle v_1, v_i\rangle + c_2\langle v_2, v_i\rangle + \cdots + c_i\langle v_i, v_i\rangle + \cdots + c_n\langle v_n, v_i\rangle$$

and because of orthogonality, this just reduces to

$$\langle v, v_i\rangle = c_i\langle v_i, v_i\rangle \implies c_i = \frac{\langle v, v_i\rangle}{\langle v_i, v_i\rangle} = \frac{\langle v, v_i\rangle}{\|v_i\|^2}$$

So based on this, the decomposition in this basis is

$$v = \frac{\langle v, v_1\rangle}{\|v_1\|^2}v_1 + \frac{\langle v, v_2\rangle}{\|v_2\|^2} + \cdots + \frac{\langle v, v_n\rangle}{\|v_{\|}^2}$$

To make this even simpler, we can require that every basis element has a norm of 1. This doesn't change any linear dependence, because all we're doing is rescaling. An orthogonal basis in which every element has unit norm is called an <u>orthonormal</u> basis. The decomposition then becomes

$$v = \sum_{j=1}^{n} \langle v, v_j\rangle v_j$$

Given any basis, we can convert it to an orthonormal basis by the Gram-Schmidt process, which is more than adequately covered elsewhere.

4 Linear Operators over Orthogonal Bases

Recall that a linear transformation is completely determined by how it acts on a basis. Since we now know how to use inner products to easily decompose vectors into an orthonormal basis, we can decompose the action of any linear transformation similarly. Suppose we have $T : V \to V$, and think of $T(v)$ as just another element of V we can decompose:

$$T(v) = \sum_{i=1}^{n} \langle T(v), v_i\rangle v_i$$

This also gives us an easy way of computing the matrix of a linear transform with respect to the orthonormal basis $\{v_i\}$. We just need to transform each basis element:

$$T(v_j) = \sum_{i=1}^{n} \langle T(v_j), v_i\rangle v_i = \sum_{i=1}^{n} A_{ij} v_i \implies A_{ij} = \langle T(v_j), v_i\rangle$$

and so, we can make the matrix $A = [T]$:

$$A = \begin{bmatrix} \langle T(v_1), v_1\rangle & \langle T(v_2), v_1\rangle & \dots & \langle T(v_n), v_1\rangle \\ \langle T(v_1), v_2\rangle & \langle T(v_2), v_2\rangle & \dots & \langle T(v_n), v_2\rangle \\ & & \ddots & \\ \langle T(v_1), v_n\rangle & \langle T(v_2), v_n\rangle & \dots & \langle T(v_n), v_n\rangle \end{bmatrix}$$

5 Inner Product Spaces as Dual Spaces

The inner product is a function $V \times V \to F$, and we've previously seen something similar where vectors are mapped to scalars, in the dual space. If we fix the second vector (so that we have linearity instead of conjugate-linearity), then the function $f(x) = \langle x, y\rangle$ is a linear functional $f : V \to F$. For every vector $y \in V$, we've now got a linear functional associated with it and vice versa. Let's prove this!

Theorem. *(F-I-S theorem 6.8) Let V/F be a finite-dimensional vector space, and let $g : V \to F$ be a linear transformation. Then there exists a unique $y \in V$ such that $g(x) = \langle x, y \rangle$ for all $x \in V$.*

Proof. g is linear, so its action is completely determined by how it acts on each element of a basis for V. For ease of working with inner products, let's require that this basis is orthonormal. From there, we'll construct an element of V such that its inner product with each element of a basis matches g, and we'll be done!

Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Then, consider the action of g on each of these basis elements, $g(v_1), \ldots, g(v_n)$. In our construction of y, we want $\langle v_1, y \rangle = g(v_1), \langle v_2, y \rangle = g(v_2)$, and so on, by the definition of y. But since β is an orthonormal basis, we can decompose y as follows:

$$\begin{aligned} y &= \sum_{i=1}^{n} \langle y, v_i \rangle v_i = \sum_{i=1}^{n} \langle v_i, y \rangle^* v_i \\ &= \sum_{i=1}^{n} g(v_i)^* v_i. \end{aligned}$$

Let's verify that this construction works: for any $x \in V$, we can decompose it as

$$x = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n,$$

and since g is linear, we can apply it as follows,

$$\begin{aligned} g(x) &= g(a_1 v_1 + a_2 v_2 + \cdots + a_n v_n) = a_1 g(v_1) + a_2 g(v_2) + \cdots + a_n g(v_n) \\ &= \sum_{i=1}^{n} a_i g(v_i). \end{aligned}$$

Next, let's compute the inner product with y!

$$\begin{aligned} \langle x, y \rangle &= \langle \sum_{i=1}^{n} a_i v_i, y \rangle && \text{decomposing } x \text{ to an orthonormal basis} \\ &= \sum_{i=1}^{n} a_i \langle v_i, y \rangle && \text{linearity and distributive properties of inner product} \\ &= \sum_{i=1}^{n} a_i \langle v_i, \sum_{j=1}^{n} g(v_j)^* v_j \rangle && \text{decomposing } y \text{ to an orthonormal basis} \\ &= \sum_{i=1}^{n} a_i \langle v_i, g(v_i)^* v_i \rangle && \text{orthonormality means } \langle v_i, v_j \rangle = \delta_{ij} \\ &= \sum_{i=1}^{n} a_i g(v_i) \langle v_i, v_i \rangle && \text{second-argument conjugate scaling} \\ &= \sum_{i=1}^{n} a_i g(v_i). \end{aligned}$$

The two results match, so we're done! $\square$

We know that V is isomorphic to V^*, and we can now write down an explicit form for that isomorphism - a one-to-one and onto mapping between each element of V and each linear functional from V!

Recall the following property of dual spaces: define $T : V \to W, T^\intercal : W^* \to V^*, g \in W^*$ and $T^\intercal(g) = g \circ T$. If we consider $W = V$ and say $g(x) = \langle x, y\rangle$ for some fixed $y \in F$, then this gives us

$$T^\intercal(\langle \cdot, y\rangle) = \langle T(\cdot), y\rangle$$

The notation in this is a little strange: I left out the x from the inner products to indicate they're functions of x that we aren't yet evaluating.

To proceed with this, let's think about what $T^\intercal$ represents in this context: it's some map $V^* \to V^*$. Here, it's taking in the operation "take the inner product with y" and returning a different operation. We know this is a linear functional, so there has to be some other fixed element of V, say $T^*(y)$, such that this new linear functional can be written as "take the inner product with $T^*(y)$". Effectively, we can think of $T^\intercal$ as only acting on y, so let's put it in as some other operation T^* that we'll have act on y.

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

We refer to T^* as the adjoint of T. It's a transformation $V \to V$, just like T is. Therefore, to find its matrix representation, we can put in each element of the orthonormal basis over which T is defined. (We're also going to start denoting complex conjugation by $\bar{x}$, so that we don't confuse that with the adjoint.)

We previously saw that $[T]_{ij}$ over an orthonormal basis was equal to $\langle T(v_j), v_i\rangle$, so let's try and write $[T^*]_{ij}$ in terms of that.

$$[T^*]_{ij} = \langle T^*(v_j), v_i\rangle = \overline{\langle v_i, T^*(v_j)\rangle} = \overline{\langle T(v_i), v_j\rangle} = \overline{[T]_{ji}}$$

Therefore, to get the matrix of T^*, we take its conjugate transpose, which I'll denote with a dagger †: we take the complex conjugate of each element, and transpose the result.

Both the conjugate-transpose representation and the construction by dual spaces can tell us that adjoints are involutory, that is, they satisfy $(T^*)^* = T$. If you take an adjoint twice, you're left with the original transformation.

6 Normal and Self-Adjoint Operators

Suppose T is diagonalizable, and therefore that there exists a basis in which $[T]$ is a diagonal matrix. In this case, the conjugate-transpose $[T^*]$ is also diagonal, and so the two commute. We call such operators normal.

$$T^*T = TT^*.$$

This gives us an easy way to check if an operator is diagonalizable: if it commutes with its conjugate transpose, that's because they're both diagonal under the same basis. This is an easier condition to satisfy than it might seem, because we're guaranteed that whenever T is diagonalizable, its diagonal basis is the same as that of T^*.

The fact that they're diagonalized by the same basis is a bit of a leap of faith: how do we know that they're not diagonalized by different bases? To answer this, we'll first have to build up a bit of machinery.

As a simpler case than the eventual result we want to reach, let's consider unitary operators, i.e. U such that $U^*U = I$. Since we're only considering transformations from a vector space to itself, this automatically means they commute, because inverses are unique:

$$U^*U = I \implies U^* = U^{-1} \implies UU^* = UU^{-1} = I.$$

Further, if we consider matrix representations, we can see that over the reals, $\det(U) = \pm 1$, because $\det(U^\dagger) = \det(U)^*$ (transposes have the same determinant as the original matrix, so we just have to take the conjugate of everything). Therefore, $\det(U^\dagger U) = \det(U^\dagger)\det(U) = \det(U)^2 = 1$. If we allow complex elements, we have the more general statement that $\|\det(U)\|^2 = 1$, so $\det(U) = e^{i\varphi}$ for some φ.

So what's the point of looking at operators like this? Unitary operators represent transformations that preserve the inner product. Since their determinant has magnitude 1, they don't do any scaling, just some sort of complex rotation.

$$U \text{ unitary} \implies \langle Ux, Uy \rangle = \langle x, y \rangle.$$

We can show this using the definition of a unitary operator and properties of the adjoint:

$$\langle Ux, Uy \rangle = \langle x, U^*Uy \rangle = \langle x, Iy \rangle = \langle x, y \rangle.$$

Unitary operators allow us to build up to the result of "being normal is equivalent to being diagonal", because it makes our change of basis easier. If we have an orthonormal basis that is not an eigenbasis, and we want to make it an orthonormal eigenbasis, we can show that the transformation between these two is unitary. It's possible to do this rigorously, but in a nutshell: since both bases are orthonormal, any basis element has unit norm both before and after the transformation, so the transformation has to have a unit scaling factor. This in conjunction with the definition of an inner product proves it's unitary.

Now, we can prove being diagonal implies being normal fairly easily by looking at the matrix forms. Suppose Q is the change-of-basis transformation to an orthonormal eigenbasis β, and D is the diagonal form. Since the bases we're translating between are orthonormal, we can say Q is unitary.

$$[T] = QDQ^{-1} \implies [T^*] = [T]^\dagger = (QDQ^{-1})^\dagger = (Q^{-1})^\dagger D^\dagger Q^\dagger = QD^\dagger Q^{-1}.$$

Therefore, to show they commute, we use these matrix representations:

$$
\begin{aligned}
T^*T &= Q[T^*]_\beta Q^{-1} Q[T]_\beta Q^{-1} = Q[T^*]_\beta [T]_\beta Q^{-1} \\
&= Q[T]_\beta [T^*]_\beta Q^{-1} = Q[T]_\beta Q^{-1} Q[T^*]_\beta Q \\
&= TT^*.
\end{aligned}
$$

The last special case of operators we'll look at under inner products is self-adjoint operators, or Hermitian operators. These are cases in which $T^* = T$. We automatically get that all such operators are normal, because everything commutes with itself. For further results on what this means for diagonalizability, see Friedberg-Insel-Spence; I think I've exhausted what I can explain as effectively than them.