Linear Algebra: An Intuitionist Approach

Matt Pettis

2021-11-10

Contents

Chapter 1

Preface

> And you may ask yourself, "What is that beautiful house?"
> And you may ask yourself, "Where does that highway go to?"
> And you may ask yourself, "Am I right? Am I wrong?"
> And you may say to yourself, "My God! What have I done?"
> – Talking Heads, "Once in a Lifetime"

> What one fool can do, another can.
> – Ancient Simian Proverb
> – Sylvanus Thompson, "Calculus Made Easy"

> "Everything is the way it is because it got that way."
> – D'Arcy Wentworth Thompson

I'd like to say I'm writing this "for the democratization of science and math," but really, for my kids and their friends so that they don't get snookered into thinking this stuff is beyond them and therefore not for them. It is for you. It is everybody's birthright.

One of the best things about science, but one I've found least talked about in the classes that I took in high school and college, is the part that explains "why do we think things work this way?" Why do we believe things are made up of atoms? Why did people believe that without the ability to *see* atoms? What is it about the technology we've built that confirms that things are made of atoms? We believe things like this because we concocted hypotheses and made experimental tests that ruthlessly and cumulatively. We make assumptions that are verifiably true, and then we reach a little further with logic and some more

subtle observations, and extend the things we get to conclude, and what we have to throw away. That is a powerful process.

Somewhere along the way, math got divorced from that process. It wasn't helped by great mathematicians like Carl Gauss who called Number Theory "the Queen of Mathematics" mostly because it didn't have much in the way of application, and that was a good thing. Or the eminent mathematician G. H. Hardy, who said,

> "I have never done anything"useful". No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world."

The perspective was, and often is, that math is a thing more akin to art, like poetry, and though it is sometimes useful, its main value is in that it is beautiful and fun. Ironically, his favorite subject, number theory, is the foundation of our ability to transmit secrets safely on the internet, and does, in fact, probably cause more good and ill than he was comfortable with.

The downside of such a perspective is that it makes it seem like learning the discipline of mathematics is inscrutable. When you encounter mathematical definitions, such as, "What makes a thing a vector space?", or "What makes a thing a group?", or "What makes a set measureable?", what you read are a bunch of seemingly awkward little statements that seem either indecipherable, or unknowable, or so stupidly dead-simple as to make you wonder why one would even need to say such a thing. For instance, when we get to the definition of a vector space, you'll see this as a defining characteristic that your, uh, we'll call a thingy for now, needs to have to be called a vector space:

$$(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$$

For those familiar with how numbers work, this seems like something Captain Obvious would say about math. It could also make you wonder "what's the point of saying such a thing?"

These definitions don't come in an inspiration, like Athena springing fully formed from the forehead of Zeus. When you study the history of mathematics, you'll see that when trying to come up with descriptions like this, mathematicians will often argue, and even disagree violently. You'll often see different characterizations like this depending on where you look, because originators disagreed on what was *fundamental* about what was going on. For systems to be compatible, though, the fundamental assumptions of one camp need to be at least derivable from the other, and vice versa.

This monograph is intended give you the motivations of why linear algebra is the way it is. It will address:

- Where do those rules about a vector space come from?
- What's the big deal about linear independence?
- What would make you come up with an idea like an inner product?
- What's helpful about orthogonality?
- Eigenvectors: how do they even work?

I'll approach this as science would ideally approach this. What do we observe? What sort of simplifications can we make to help us understand what is going on? What sort of models help us understand the important parts?

Chapter 2

Well, How Did I Get Here?

> It takes a very unusual mind to undertake the analysis of the obvious.
> – Alfred North Whitehead
>
> Point of view is worth 80 IQ points
> – Alan Kay

2.1 Intuition about $\mathbb{R}^n$

At this point, if you are reading this as opposed to an intro to linear algebra book, I assume the one thing you have is a good familiarity with are the vector spaces of $\mathbb{R}^2$ and $\mathbb{R}^3$. These are the vector spaces used extensively in physics and engineering to model things like position, velocity, and forces.

You know that you can add two vectors by putting the tail of one vector at the head of the other and connecting the base of one to the tip of the other vector.

And we know how to stretch, shrink, and flip a vector, which we do by multiplying a vector by a real number. If you multiply a vector by 1/2, the length of the vector shrinks to one-half its original length (but the direction doesn't change). If you multiply a vector by $3\sqrt{2}$ the magnitude of the result is stretched by that amount.

These are the prototypical vector spaces, one I'd argue 99 times out of 100 people imagine if they know what a vector space is already. And if this is the only space you have to deal with, this intuition is pretty much all you need. It will take you far. But if you study quantum state space vectors, with their complex entries, and the notion of $\mathbb{R}^n$ vector spaces can help some of your intuition, and for other parts, you are totally at sea as to how the hell you are supposed to interpret things.

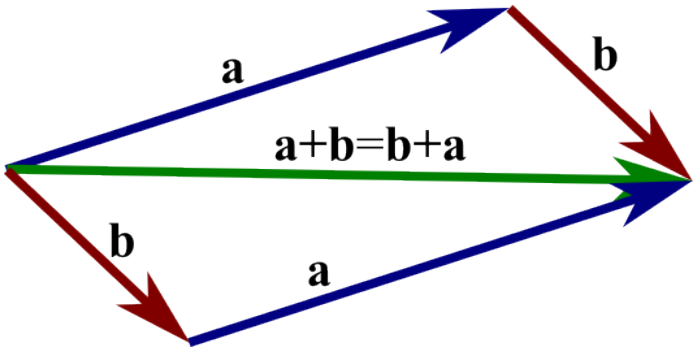

Figure 2.1: Parallelogram Law of Addition

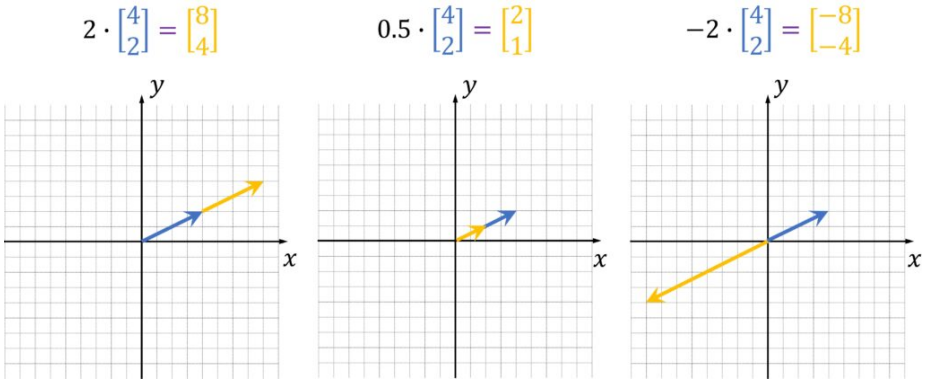

Figure 2.2: Vector Scaling

So it may help to go back to $\mathbb{R}^n$ and do some careful syntax-to-interpretation explanations. And take some of the things we take for granted as obvious and cast them in new, abstract lights.

We just discussed how addition and scalar multiplication are performed. But we should note the choices we made to do what seems so natural to us.

First, addition. We did it using the parallelogram law. How come we didn't choose to make the resulting vector have a length equal to the sum of the length of the two starting vectors? You can see from the picture the length of our resulting vector is usually shorter, but never longer, than the length of the two vectors you are adding. You should counter, "well, that's because things like forces don't work that way. If I have two equal and opposite forces acting on an object, I want the result to be no net force. If the magnitudes were added, which direction would the result point? Doesn't make sense, so we can't use that because it doesn't model the situation."

Good answer, reader!

We want our model to reflect what is going on. So that sort of addition is out. It turns out head-to-tail addition does model things like forces. What is also intuitive is that for any arrow, or force, or velocity, you can "counteract" it with another arrow/force/velocity. You could not do this if the magnitues themselves always add. You can never add two positive things to get a net zero, which would represent two things counteracting each other.

Second, just what do we mean by multiplying a vector by a scalar? That's a little trickier. We know what we mean: we shrink and grow the vector, in the same direction, by the scalar's sign and magnitude. As above, multiplying a vector by 1/2 shrinks its length to 1/2 the original length, but the direction stays the same.

But if you are an engineer or scientist, your unit-analysis Spidey-sense should start to tingle… if scalars and vectors are different things, and I multiply them, whatever happens, I shouldn't get a vector out, like we are seeing.

A dirty little detail is usually glossed over. Because it is short-circuiting to your intuition about stretching or shrinking, people don't talk about what $\alpha\vec{x}$ is. Your brain probably assumes it is a multiplication, and what the scalar is multiplying is the "magnitude" property of the vector. You can think about it that way for $\mathbb{R}^n$ vectors. But that's not really a scalar-vector multiplication. What it really is is the scalar acting as a function that maps a vector to another vector. Like so:

$$f_\alpha : V \to V, \alpha \in \mathbb{R}$$

So, $\alpha\vec{x}$ is really shorthand for $f_\alpha(\vec{x})$.

For $\mathbb{R}^n$, that function is the function that stretches or shrinks the magnitude of the vector by α. Again, this may seem like nit-picky pedantry, but we have to start leveraging this fact in the near future to keep some other things clear…

2.2 The Axioms

So if vector spaces were just stretching and head-to-tail vector additions, mathematicians probably wouldn't make the seemingly awkward definition that are always presented in the first three pages of any textbook:

From Wikipedia:

A vector space over a field F is a set V together with two operations that satisfy the eight axioms listed below. In the following, $V \times V$ denotes the Cartesian product of V with itself, and $\to$ denotes a mapping from one set to another.

- The first operation, called **vector addition** or simply **addition** $+ : V \times V \to V$, takes any two vectors $\mathbf{v}$ and $\mathbf{w}$ and assigns to them a third vector which is commonly written as $\mathbf{v} + \mathbf{w}$, and called the sum of these two vectors. (The resultant vector is also an element of the set V.)
- The second operation, called **scalar multiplication** $\cdot : F \times V \to V$, takes any scalar a and any vector $\mathbf{v}$ and gives another vector $a\mathbf{v}$. (Similarly, the vector $a\mathbf{v}$ is an element of the set V. Scalar multiplication is not to be confused with the scalar product, also called *inner product* or *dot product*, which is an additional structure present on some specific, but not all vector spaces. Scalar multiplication is a multiplication of a vector *by* a scalar; the other is a multiplication of two vectors *producing* a scalar.)

Elements of V are commonly called *vectors*. Elements of F are commonly called *scalars*. Common symbols for denoting vector spaces include U, V, and W.

In the two examples above, the field is the field of the real numbers, and the set of the vectors consists of the planar arrows with a fixed starting point and pairs of real numbers, respectively.

To qualify as a vector space, the set V and the operations of vector addition and scalar multiplication must adhere to a number of requirements called axioms.[1] These are listed in the table below, where $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ denote arbitrary vectors in V, and a and b denote scalars in F.[2][3]

Axiom	Meaning
Associativity of vector addition	$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
Commutativity of vector addition	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
Identity element of vector addition	There exists an element $\mathbf{0} \in V$, called the *zero vector*, such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
Inverse elements of vector addition	For every $\mathbf{v} \in V$, there exists an element $-\mathbf{v} \in V$, called the *additive inverse* of $\mathbf{v}$, such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
Compatibility of scalar multiplication with field multiplication	$a(b\mathbf{v}) = (ab)\mathbf{v}$ [nb 2]
Identity element of scalar multiplication	$1\mathbf{v} = \mathbf{v}$, where 1 denotes the multiplicative identity in F.
Distributivity of scalar multiplication with respect to vector addition	$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
Distributivity of scalar multiplication with respect to field addition	$(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$

Figure 2.3: [Record scratch] Yep, that's me, Vector Space, spewing a lot of incomprehensible stuff. But I wasn't always like this. Let me tell you a story...

It is a bad way to start, because you really have no frame of reference for what any of the terms mean. It is frustrating and made me angry **We shouldn't really start here.**

This reminds me of the joke about a cab driver driving around the Seattle area in a fog, and he asks a guy coming out of a building if he can tell the cabbie where he is. The guy looks at him and says, "You're in a cab," and walks on. The cabbie says "Perfect, I know where I am." The fare asks him how he can figure out anything from what the guy outside said, and the cabbie say, "Well, he told me something that was completely true and completely useless. So this must be the Microsoft building."

There's so many questions that should be triggered. Like:

- Why so many simple axioms?
- What are vectors, really?
- What is a field, really?
- What does it mean to add vectors?

- What does it mean to multiply a vector by a scalar?
- How do I know what other vector $a \cdot \vec{u}$ becomes?
- It looks like the scalar field talks about how *scalar addition* and *scalar multiplication* works. But vectors only talk about *vector addition*. What's up with that?

With $\mathbb{R}^n$, we can look over the axioms and confirm that yep, they meet the axioms, because we have a strong intuition about how addtion of vectors work, how scalars stretch the vectors, and how the addition and multiplication of scalar (read: real) numbers work.

We can boil these axioms down as follows:

- It won't matter what order you do your operations in, you are guaranteed to get the same answer. All of the axioms that deal with associativity, commutativity, compatibility, and distributivity are telling us this.
- Your vector space has an element that acts as a $\vec{0}$ element, and every vector has an opposite that, when you add them together, gives you $\vec{0}$.
- Your scalar field, which has a 1 in it, when combined with a vector, won't change that vector. Or, $1 \cdot \vec{x} = \vec{x}$

But that doesn't answer why we say what we want in a vector space that way. Even more, we don't have a good idea of what other structures my meet these axioms and be a vector space. Or how to interpret that.

2.3 Why are the axioms this way?

In elementary school, you learned your 1-digit by 1-digit times tables by rote. And then you learned how you could leverage that to multiply numbers bigger than 1-digit together by using the 1-digit table entries, putting the results in certain columns, learning how to carry, write down partial sums, and add them together. You need only know the 1-digit times tables and a procedure with special addition rules to handle the multiplication of any two numbers of any length.

That reduction to a small set of memorized multiplication facts and easy rules to combine the results to get us any arbitrary computation is what we are striving for in vector spaces. We don't want to have to memorize the sum of every two-vector combination in the space, because by my last count, I'm pretty sure there are more than 57 vectors in $\mathbb{R}^2$ alone. What we want is a nice way to get a fixed set of facts we have to know about, and find simple procedures to compute the facts only as we need them.

So here's the game plan, to try and tackle complex things like $\mathbb{R}^n$ and other things that will qualify as vector spaces. We are going to reference some topics

that will be familiar to you if you've dealt with linear algebras before, but haven't talked about yet here:

Task	Vector space concept
Get a fixed number of facts (vectors $\vec{x}_i$) we need to deal with, like the times table.	Pick linearly independent vectors $\vec{x}_i$.
Figure out how to express them as simple combinations of known facts	Linear combinations with scalars: $\vec{x} = \alpha_1\vec{x}_1 + \alpha_1\vec{x}_1 + ... + \alpha_n\vec{x}_n$. A minimal set of linearly independent vectors that can do this is called a *basis*.
Discover that for vector spaces (finite ones), any set of vectors that work to do this *always* has the same number of vectors	This is the dimensionality N of the space.
Figure out how to calculate α_i for a given set of vectors.	To do this, we'll have to create the concept of the *inner product*.
Find out that lots of choices of $\vec{x}_i$ will work to express any vector, but *certain* ones are easier to compute the α_i's for.	These will be vectors that are orthogonal and of a certain size (magnitude = 1). The word that captures both of these concepts is *orthonormal*. There are still lots of orthonormal sets.
Figure out how to take an arbirary linear independent set and construct an equivalent orthonormal set out of them, because they are easier to deal with.	This is the Gram-Schmidt theorem.

This will be enough for now. Once we get into functions *between* vector spaces, and pick out the special functions (the *linear* transforms), we'll talk about what *eigenvectors* and *eigenvalues* are. But not yet.

The insight here is that for a given vector space, rather than work with the raw vectors themselves, it is easier to break them down into computations on known vectors (the basis), and do scalar multiplication (really, mappings that work like the underlying arithmetic on the scalars) on those basis vectors. This is because the operations on the scalar field, and the way they map (or stretch) the basis vectors is ultimately an easier accounting scheme. So we go through all of this work to find linear independent sets, bases, orthonormal bases, an inner product function, etc., so that we can use the inner product to easily get scalar coefficients and go on our merry way doing calculations on the vector space elements themselves in terms of their fewer and simpler basis elements.

Chapter 3

What is This Beautiful House?

As with most useful finds in life, you grope around blindly in the darkness for a long time, stumbling and swearing, messing up, giving up, coming back, and sometimes tripping over something useful. If what you trip over is really useful, you take it to the garage, clean and polish it, put it on display, and tell lies about how you followed a primrose path, stemming from your genius, to its discovery, making people in awe of you.

We're going to follow that a bit. It will have some convenient lies that will have enough grains of truth in it to be helpful. I don't pretend to know the history of how linear algebra came into its present form, but I have made a satisfying hero's origin story narrative that has helped me keep important notions straight. I'll tell that lie here.

3.1 Starting with what we know

The approach I'll take here is to start with $\mathbb{R}^2$, and work up to more general abstractions. We'll start with the most commonly understood stuff, and then start to wiggle stuff around to see what happens and how we can use it. We will rely on things we know intuitively about Euclidian space, like being able to construct perpendicular and parallel lines, measure lengths and angles, and the like. Then we will start to erase and replace things with more abstract concepts, until we get to a level of some abstract generality. With this established, we will hop over to the vector space of quantum states, and we'll be able to pull over the general abstractions, as some of the concrete intuitions don't work in quantum vector spaces.

3.1.1 Scenario 1: Orthonormal vectors

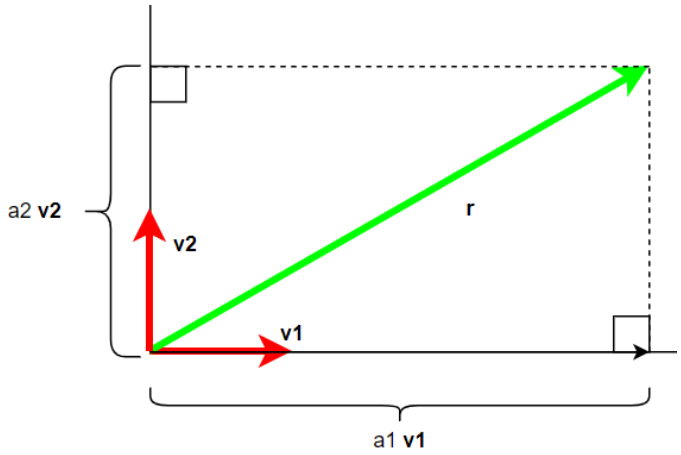

Figure 3.1: How most of us understand vector decomposition

This is picking vectors so that you have effectively the Cartesian coordinate plane that you are used to – $v1$ and $v2$ act like the x- and y-axis of your high school years.

To recap the highlights:

- You are given two vectors that are perpendicular to each other (*ortho-*), $\vec{v1}$ and $\vec{v2}$, and are of unit length(*-normal*).
- You are given the vector $\vec{r}$
- You figure out the way to write $\vec{r}$ in terms of $\vec{v1}$ and $\vec{v2}$ by dropping perpendiculars to the lines determined by $\vec{v1}$ and $\vec{v2}$, and then figuring out the ratio of the length of those intersections to the origin to the length of the individual unit vectors. We call those lengths $a1$ and $a2$ respectively.
- You can then write $\vec{r}$ as follows:

$$\vec{r} = a1 \cdot \vec{v1} + a2 \cdot \vec{v2}$$

The big idea here is that, because we can draw perpendiculars and measure lengths (which we are allowed to do because we can do all of the good things the geometry we learned allows us to do), it's easy to compute the scalars $a1$ and $a2$.

I contend that it is pretty self-evident here that the following two statements are true:

- You can figure out the scalars (also known as *components*) for $a1$ and $a2$ for *any* vector $\vec{r}$, no matter the size or direction. You can do this with geometric tools (marking of unit lengths, dropping perpendiculars or drawing parallels – anything you can do with a compass and straight-edge).
- If you have $\vec{r}$ in hand, the values you compute for $a1$ and $a2$ are unique – you can't use some other way of constructing parallels and measuring lengths, and legitimately come up with different numbers.

This seems like a boring observation, but, as I've said before, this will be important.

3.1.2 Scenario 2: Getting a little more artsy: any two linearly independent vectors

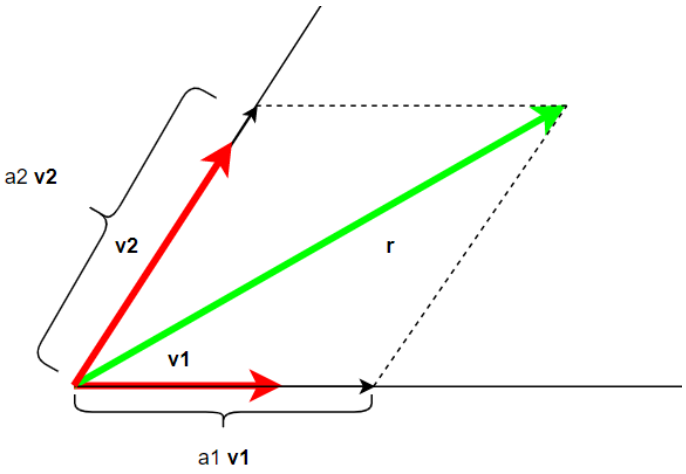

Figure 3.2: Scalars decomposition using general vectors.

You might find yourself in a situation where you have two reference vectors that are not perpendicular to each other, and not of unit lenghts. That is the situation above. It is not neat and familiar like the first orthonormal case. But it will work. Any vector $\vec{r}$ can be written by stretching or shrinking both $\vec{v1}$ and $\vec{v2}$ by just the right amounts so that you can add them, parallelogram-wise, to get to the same spot as $\vec{r}$.

The one caveat is that in this case, instead of dropping perpendiculars to lines, we constructed lines parallel to each of the different vectors through the end of $\vec{r}$. We could have constructed parallels in the orthonormal case as well, and gotten the same results. The interesting thing to keep in your back pocket is, if you

ever end up studying tensors, these alternate ways of finding the *components* of the vectors are both used, depending on if you want what tensors refer to as *contravariant* or the *covariant* components. It's not exacly like we show here, but very, very close in spirit. So it's nice to have an example. See: http://www.danfleisch.com/sgvt/ .

3.1.3 Important observations

We should have a pretty good intuition that the amount we have to stretch vectors to have it represent $\vec{r}$ are going to totally depend on which vectors $\vec{v1}$ and $\vec{v2}$ you use as your reference set. To be clear, for $\mathbb{R}^2$, the two independent vectors you choose to represent any other arbitrary vector $\vec{r}$ is called the *basis*.

We are able to use our intuition on $\mathbb{R}^2$ because we understand how geometry works. We know how to draw perpendiculars and parallels, and measure distances. When we got to more abstract spaces, or when we use complex numbers as a scalar field, we will have to leave our intuition behind, as we won't really have perpendicular and parallel lines to deal with. We will have to replace "perpendicular" with "orthogonal", which we will be able to determine by having the inner product being 0.

The scaling factors $a1$ and $a2$ are called the *components*, to repeat myself. Components are important because they are the numbers that actually get thrown around, showing up as the entries in any matrix and vector representation once a basis is chosen.

Note that the vector $\vec{r}$ is it's own, eternal, platonic thing. It exists no matter what two basis vectors you choose to decompose it in.

Every vector can be decomposed into a set of basis vectors and appropriate scalar factors. We pretty much have to decompose an arbitrary vector into it's basis representation to do any meaningful computations on the vector.

3.2 Inner products

Inner products, in their most general form, don't seem to lend themselves to a good intuition. Parts of inner products have useful interpretations, in certain circumstances, but it will be helpful to build up to that.

So let's start with a place where the inner product shows up naturally – The Law of Cosines:

$$c^2 = a^2 + b^2 - 2ab\ cos(C)$$

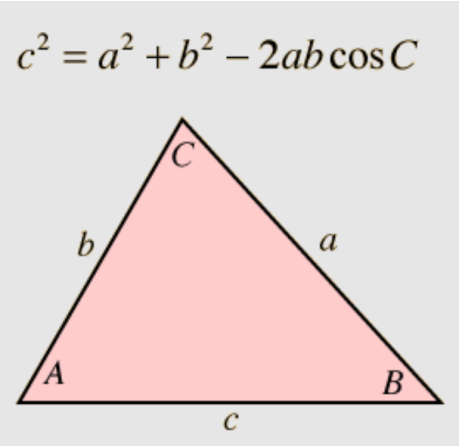

Figure 3.3: Law of Cosines

Before we get to formal definitions, we will take for given the fact that the inner product in Euclidian space is defined as:

$$< \vec{x}, \vec{y} >= |\vec{x}| \cdot |\vec{y}| \cdot \cos(\theta)$$

That is, the inner product between two vectors (lengths, in this case), is the product of the lengths of the two vectors multiplied by the cosine of the angle between the vectors/lines.

We will take for granted that we can write the vector lengths with just plain Roman letters.

You can think of the law of cosines as a generalized Pythagorean theorem. You can figure out the length of a third side c if you know the other two sides of the triange a and b, but *adjusting for how much in the same direction a and b are.*

Let's imagine some limiting cases. Consider a and b as fixed lengths. As angle C widens to be close to 180 degrees, the triangle height shrinks to 0, and c becomes equal to $a + b$. The limit in the law of cosines agrees with this – as $C \to 180$, then $cos(C) \to -1$, and so the law of cosines becomes:

$$c^2 = a^2 + b^2 - 2ab(-1) = a^2 + b^2 + 2ab = (a + b)^2$$

or, $c = a + b$.

At the other extreme, as C shrinks to 0, the triangle again collapses, and a and b are again in the same line, but but this time c approaches the *difference* of a and b.

$$c^2 = a^2 + b^2 - 2ab(1) = a^2 + b^2 - 2ab = (a - b)^2$$

or $c = a - b$, or $b - a$, whichever doesn't give you a negative number.

In between, at some point, C becomes a right angle. In that case, $cos(C) = 0$, and the entire last term goes to 0. Then we are just left with the formula fom the Pythagorean Theorem.

One way to think about this is that the last term in the law of cosines is an adjustment factor for *how collinear a and b are.* When $C = 180$, a and b are in opposite directions, and the Pythagorean theorem underestimates how long C is, and we have to add that factor to the computation to get the right length. At the extreme of $C = 0$, the Pythagorean theorem *overestimates* c, so we have to subtract some to get the right amount.

In between, when the lines are perpendicular, no adjustment to the Pythagorean theorem calculation is needed.

So, when the angle is small, the lines point pretty much in the same direction, and the inner product is more positive. At the other end, when the angle is wide, the lines are pointing away from each other, and the inner product is more negative.

So you can get the idea that the inner product measures some notion of two lines pointing in the same, opposite, or perpendicular direction, based on the size of the inner product value.

As an aside, you may recall that:

$$\vec{x} \cdot \vec{y} = |\vec{x}| \cdot |\vec{y}| \cdot \cos(\theta)$$

and

$$|\vec{x} \times \vec{y}| = |\vec{x}| \cdot |\vec{y}| \cdot \sin(\theta)$$

There is a bit of a parallel going on here.

cos	sin
Gives a sense of how collinear two lines are	Gives a sense of how much planar area (the paralleleogram they define) the lines subtend

Chapter 4

How do I work this?

My uncle tells the story of how they help transition mechanical engineers fresh out of school to working in industry. A senior engineer tasks them with a simple task: a near complete machine needs a piece – it is missing a cog. The tell the new engineer to get them one. The new engineer studies the plans, designs the cog, painstakingly laying out the CAD designs for it, and comes back to the senior engineer after a week with his work to review how to machine that part, and how much it will cost. The senior engineer looks it over, says it is the right piece, and pulls out a catalog, and shows him how much it will cost to buy, usually at a fraction of the cost. The lesson they learn: don't create from scratch stuff you can buy cheaply from a catalog.

We might think of physics and mathematics having a similar relationship. Often, as you painstakingly observe a physical system, you figure out what are the important bits you need to capture with a model, and how those bits behave in the system. Then, you go shopping to your local math department, describing what you are doing, and hope that a mathematician can recommend a system they've already investigated that you can map your problem into, and take advantage of the notions, the notations, and theorems about the system behavior that they've already come up with.

So, let's proceed by looking at a simple spin-1/2 system, of electrons that can have either a spin up or down measurement, depending on how the spin-detector is oriented. Many newer books start their quantum discussion this way, and two I can recommend that do this are *Quantum Mechanics: The Theoretical Minimum* by Len Susskind and *Quantum Mechanics: A Paradigms Approach* by David McIntyre. I won't repeat their treatment here, and will assume you know the basics of measuring spins in those systems. I will talk about how you start to pick your mathematics from the catalog to model this phenomenon.

4.1 Lay of the Land

Here is a picture that captures some of the important information about a quantum experiment we run:

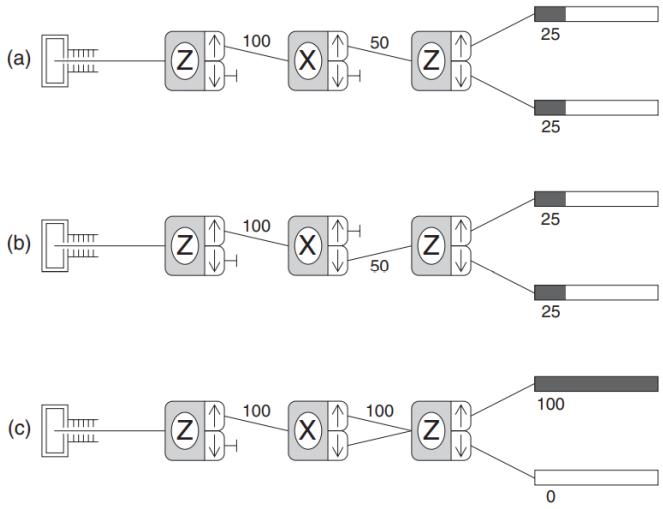

FIGURE 1.6 Experiment 4 measures the spin component three times in succession and uses (a and b) one or (c) two beams from the second analyzer.

Figure 4.1: Two-spin experiment, McIntire

Some notes on the picture:

- The apparatus on the left emits electrons, which have a spin we want to measure.
- The detectors are labeled with a 'Z' or 'X' to tell you which direction in space we are orienting the detector. We pick the directions arbitrarily, but 'Z' and 'X' are orthogonal directions.
- The arrows on the detectors indicate which spin orientation an electron comes out of. If it is measured with a spin up, relative to the how the detector is oriented, it exits from the top port, with the up arrow. If down, it exits from the lower port.
- The labels `|+>` and `|->` are also labels for the spins. The indicate absolute directions of the spin (spin up or down in the Z direction), rather than the up and down arrows on the port, that indicate if the spin is up or down *relative to the how the port is oriented*, in the 'Z' or 'X' direction.
- The labels `|+>_x` and `|->_x` represent the absolute spin up/down orientations, but for the X direction. Again, the detector can be reoriented, but if the electron has one of these labels, it is independent of how the detector is oriented.

- The numbers and shaded bars represent a percentage of electrons that end up in that bucket or state over a large number of electrons entering the system. Like a histogram.

Here are some things you, the experimenter, observe about the system:

- In (a), if you measure the Z spin, then the X spin, then the second time you measure the Z spin, it will be 50-50 up/down.
- In (b), it just tells you that the same thing happens, no matter if the secon measurement for X is up or down, like in experiment (a).
- In (c), somthing really weird happens: If you put the X spin detector in the middle, but carefully recombine both beams, as if you didn't measure the X spin, then the Z spin will be as if you didn't measure X spin at all, and stays in the spin state you measured it in in the first detector.

That can stand as the first of many weirdnesses you encounter in quantum. As McIntyre says: going from (a) or (b) to (c), it's as if you are in a half-lit room, throw open a window shade, and the whole room goes dark. In a classical model, (c) would still have a 50-50 split of spin measurements, but that doesn't happen in the quantum world.

4.2 What things do we need to model this?

So, we make the following observations.

- If we measure a spin as up in the Z direction, and keep measuring the spin with detectors all pointed along the same Z axis, we will always get the same spin up measurement. That is true as long as we don't orient the detector in a different direction.
- Focusing on Z measurements, we have two states: spin up (`|+>`) and spin down (`|->`). We get that with the detector registering a +1 or -1.
- If we measure a Z spin with a detector, and it registers a +1, we call it spin up, and take a second, identical detector, and flip it upside, it will register a -1. This scenario is not pictured.
- If we take a Z detector and measure a spin up electron, and then take a second detector, and start to slightly tilt the detector away from a straight Z orientation, we still only measure +1 and -1 readings. For a slight tilt, almost all electrons will measure spin up, and a few spin down. As it rotates to be perpendicular to the Z direction, electrons measure +1 and -1 with a 50% probability. As we get closer to the tilt making the detector upside down, then most of the electrons will measure -1, until it is exactly upside down, and the detetcor will consistently measure -1.

This is kind of weird. Note that if this were a classical spin, you would expect that if you measured a spin of +1, and tilted the detector a little bit, you would measure a value a little less than +1. But quantum doesn't work that way. Instead of reducing the spin a little bit, what happens is that the probability of a +1 starts to decrease as you tilt the detector. The *average* of multiple measurements approaches the value of what you would expect a single measured spin to decrease by (if it were classical).

So, whatever math we come up, it can't be the same classical math that gives a single spin measurement decreasing continuously from +1. It has to be something that accounts for:

- A spin will always either be +1 or -1
- The *probability* of detecting +1 and -1 change as you tilt the detector.

So we go talk to the math department…