

Linear Algebra: An Intuitionist Approach

Matt Pettis

2021-11-28

Contents

1	Preface	5
2	Well, How Did I Get Here?	9
2.1	\mathbb{R}^2 , the Pythagorean Theorem, and the Law of Cosines	9
2.2	The Inner Product, or Dot Product	11
3	What is This Beautiful House?	15
3.1	Starting with what we know	15
3.2	Inner products	18
4	How do I work this?	21
4.1	Lay of the Land	22
4.2	What things do we need to model this?	23

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 \mathbb{R}^2 , the Pythagorean Theorem, and the Law of Cosines

Let's start with what we know pretty solidly – the Pythagorean Theorem. We know that, if we have a triangle $\triangle ABC$, with C being a right angle, then:

$$(\text{Pythagorean Theorem}) \quad c^2 = a^2 + b^2$$

The thing to note here is that the relationship of the three sides in a right triangle has a very tidy little relationship according to the theorem. c^2 is totally determined by a 's and b 's value in a right triangle.

But what happens if we keep \vec{a} and \vec{b} the same length, but start to wiggle the right angle at the top, making it smaller or larger?

You'll note that the length of \vec{c} will shrink or grow from what it is in a right triangle with the same leg lengths. Can we make an equation like the Pythagorean Theorem that accounts for this deviation from what you get if you had a right triangle? Turns out, yes you can, and that *adjusted* formula is known as **The Law of Cosines**:

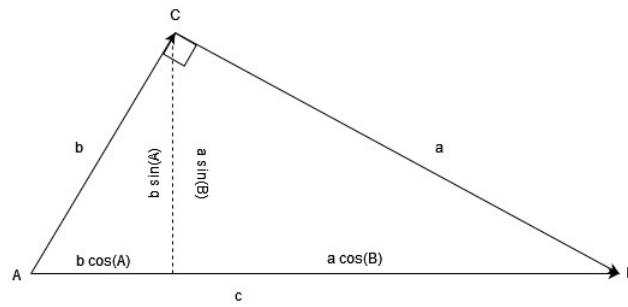


Figure 2.1: Pythagorean Theorem

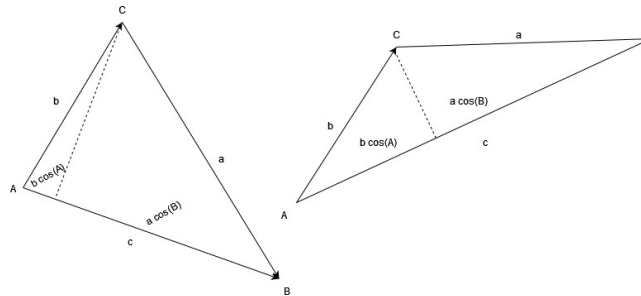


Figure 2.2: Acute and obtuse triangles

$$\text{(Law of Cosines)} \quad c^2 = a^2 + b^2 - 2ab \cos(C)$$

This should make some rough sense as a formula. As C shrinks below 90 degrees, $\cos(C)$ becomes slightly positive, and $-2ab \cos(C)$ will ultimately reduce the value of the right side of the equation, and c^2 will be a little smaller than what it would be if C were a right triangle. In the other direction, as C grows above 90 degrees, $\cos(C)$ becomes slightly negative, and so $-2ab \cos(C)$ becomes slightly positive, and c^2 the right hand side makes c^2 a little larger, so length c is a little larger.

I won't derive it here, as it is easy to find proofs. We should ask why this formula is important. Note that if you look at the diagrams, you will notice that:

$$(1) \quad c = b \cos(A) + a \cos(B)$$

in all cases. That's pretty easy to see. What the Law of Cosines does, through some trigonometric identity substitutions, is re-express that equation, going from the angles that \vec{c} makes with \vec{a} and \vec{b} , to using the angle that \vec{a} and \vec{b} make with each other instead. So the Law of cosines expresses \vec{c} totally in terms of the lengths of \vec{a} and \vec{b} and the angle they make with each other, never relying on information about c .

It turns out that that this concept, of expressing vectors in terms of other vectors, is crucial in linear algebra, and we'll build up some concepts which may seem weird or abstract at the beginning, but turns out to be super useful in general, as it helps us lift off of reasoning about vector spaces in Euclidian space to general vector spaces.

2.2 The Inner Product, or Dot Product

Let's look at that adjustment factor to the Pythagorean Theorem that allows us to calculate any third side knowing the other two. Taking out the factor of -2 , we are left with:

$$(2) \quad ab \cos(C)$$

That particular computation, in Euclidian Space, is known as the **inner product of \vec{a} and \vec{b}** . It shows up in many contexts, and in general, the inner product will have different notations:

$$\vec{a} \cdot \vec{b}$$

$$\langle \vec{a}, \vec{b} \rangle$$

$$\langle \vec{a} | \vec{b} \rangle$$

How do we interpret this? There are a few components that go into building this up.

First is $\cos(C)$, which is the angle between \vec{a} and \vec{b} . The cosine can be thought of as a measure of how much two lines, or vectors, point in the same direction. A value of 1 means they point completely in the same direction, -1 means they point in opposite directions, and 0 means that they are perpendicular to each other. We discussed above with the Law of Cosines that “point-in-the-same-direction” measure is baked into the adjustment to the Pythagorean Theorem to calculate the length of a third side of a triangle given the other two.

Second, how do we interpret multiplying the cosine by the lengths of a and b ? Here we will make a stunning insight. Let’s look back at equation (1). If we multiply both sides of the equation by c , we get:

$$c^2 = cb \cos(A) + ca \cos(B)$$

or,

$$c^2 = \vec{c} \cdot \vec{b} + \vec{c} \cdot \vec{a}$$

by the definition of inner product we saw above. So if you squint, this kinda sorta looks like the Pythagorean Theorem. Furthermore, You can think of c^2 as the inner product of a vector with itself, as the angle between a vector and itself can reasonably be thought of as 0, and $\cos(0) = 1$. So we can further rewrite the above as:

$$\vec{c} \cdot \vec{c} = \vec{c} \cdot \vec{b} + \vec{c} \cdot \vec{a}$$

Again, squinting, that is starting to look a lot like the Pythagorean Theorem, at least on the surface.

So, for Euclidian space, I like to think of the inner product, or dot product, as something you can calculate that accounts for how much in the same direction two vectors point (the cosine part), and add factors that make it dimensionally correct so that you can write a Pythagorean Theorem-like formula relating the lengths of sides of any triangle.

It turns out that the inner product has a lot of super-useful properties, mostly due to the high degree of symmetry it has.

What do I mean by “symmetry”? Notice in the Pythagorean Theorem we distinguish between the hypotenuse and its legs. The formula only works when the square of the hypotenuse is on one side of the equation, and the squares of the legs are on the other side.

But this is not true of the Law of Cosines. The formula works no matter what two sides of the triangle you pick to derive the length of the third side from. That is, these formulas are all valid, and they have the same “form”:

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

$$a^2 = c^2 + b^2 - 2cb \cos(A)$$

$$b^2 = a^2 + c^2 - 2ac \cos(B)$$

or,

$$\vec{c} \cdot \vec{c} = \vec{c} \cdot \vec{b} + \vec{c} \cdot \vec{a}$$

$$\vec{a} \cdot \vec{a} = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$$

$$\vec{b} \cdot \vec{b} = \vec{b} \cdot \vec{a} + \vec{b} \cdot \vec{c}$$

That’s what I mean by symmetry, and that the Law of Cosines has it, but the Pythagorean Theorem does not.

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

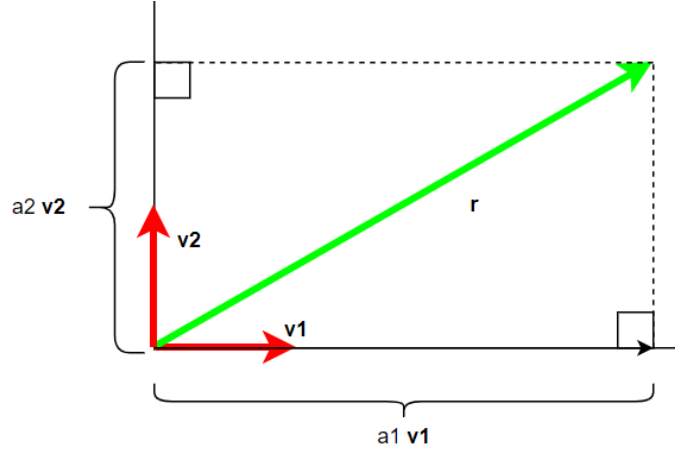


Figure 3.1: How most of us understand vector decomposition

This is picking vectors so that you have effectively the Cartesian coordinate plane that you are used to – \vec{v}_1 and \vec{v}_2 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{v}_1 and \vec{v}_2 , 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{v}_1 and \vec{v}_2 by dropping perpendiculars to the lines determined by \vec{v}_1 and \vec{v}_2 , 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 a_1 and a_2 respectively.
- You can then write \vec{r} as follows:

$$\vec{r} = a_1 \cdot \vec{v}_1 + a_2 \cdot \vec{v}_2$$

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 a_1 and a_2 .

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 a_1 and a_2 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 a_1 and a_2 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

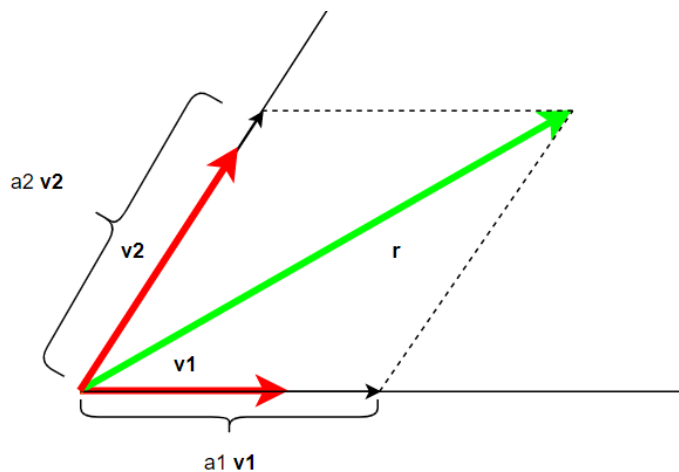


Figure 3.2: Scalars decomposition using general vectors.

You might find yourself in a situation where you have two reference vectors that are not perpendicular to each other, and not of unit lengths. 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{v}_1 and \vec{v}_2 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 exactly 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{v}_1 and \vec{v}_2 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 a_1 and a_2 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:

$$(\text{Law of Cosines}) \quad c^2 = a^2 + b^2 - 2ab \cos(C)$$

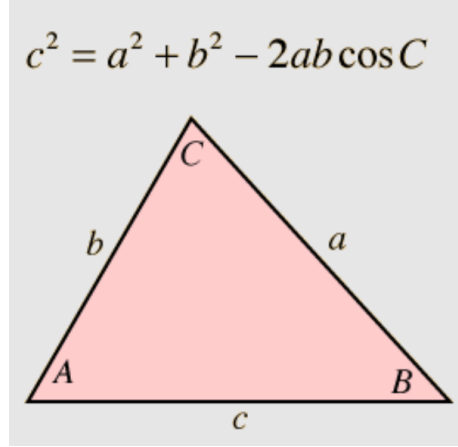


Figure 3.3: Law of Cosines

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

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

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 triangle 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 $C \rightarrow 180$, then $\cos(C) \rightarrow -1$, 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 from 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, there's some cool analogs between *cos* and *sin*. 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 parallelogram 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:

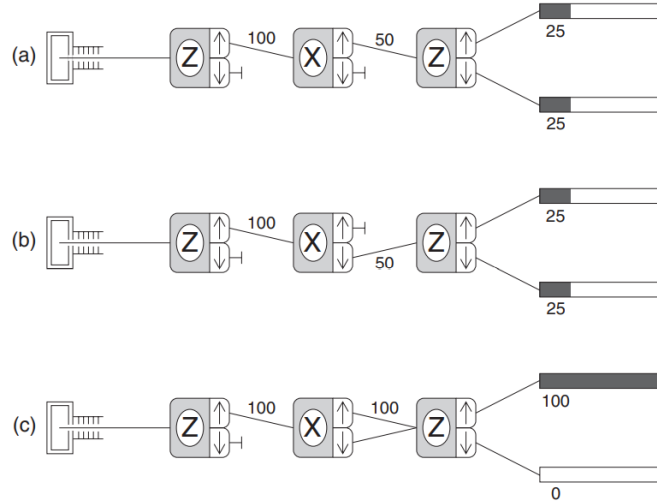


FIGURE 1.6 Experiment 4 measures the spin component three times in succession and uses (a and b) one or (c) two beams from the second analyzer.

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 $|+\rangle$ and $|-\rangle$ are also labels for the spins. They 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 $|+\rangle_x$ and $|-\rangle_x$ 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 second measurement for X is up or down, like in experiment (a).
- In (c), something 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 ($|+\rangle$) and spin down ($|-\rangle$). 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 detector 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...