Electron Spin and Linear Algebra

Jesse Thaler

April 12, 2001

1 Mathematics and Physics

At the beginning of the semester, Prof. Banchoff would often divide the white board into three sections: Algebra, Geometry, and Applications. Because he is a geometer, his favorite subject was center stage. In this lecture, Theoretical Physics will be center stage. Theoretical Physics requires us to know something about Experimental Physics (right) and Mathematics (left), too, so we will use results from these areas to develop a beautiful theory of *Electron Spin Quantum Mechanics*.

Electron Spin Quantum Mechanics (ESQM) is based almost entirely on Linear Algebra. Of course, there are many other areas of mathematics besides Linear Algebra, many of which have direct relevance to Physics (or were actually developed in order to solve problems in Physics). The most important area by far is Calculus, which allows us to study the time evolution of systems. Along the same lines comes Differential Equations. Probably the most famous differential equation is Newton's Second Law $F = ma$, which describes the time evolution of a system if we know the nature of the forces acting on the system.

The theory of Quantum Mechanics (on which ESQM is founded) was originally based on a differential equation developed by Schrödinger. Eventually, an algebraic formulation of Quantum Mechanics based on Linear Algebra (among other things) was developed, and though this formulation gives the same answers as Schrödinger's Equation, it has symmetries that make it far more elegant and understandable. Today, theoretical physicists are working on M-Theory (where M stands for Matrix), which in many ways is an extension of Algebraic Quantum Mechanics.

As a side note, a fourth area of Mathematics that has direct relevance to Physics is Differential Geometry (taught in some years by Prof. Banchoff). This branch of mathematics forms the foundation for Einstein's theory of General Relativity, a description of space-time curvature and gravity.

2 Electron Spin

In order to understand what electron spin is, we need to go over to Experimental Physics. One of the results from Electricity and Magnetism (with a bit of quantum intuition) is that an electron in a magnetic field will feel a force of magnitude

$$|F| = |S||B| \cos \theta$$

where S is the spin of the electron, B is the magnetic field, and θ is the angle between these vectors. (There is also a force due to the velocity of the electron but we can ignore that in this discussion. Also, there are constants in the above equation that I have omitted.) Because we know a lot about vectors, we can go one step further and write down a vector equation to describe not only the magnitude of the force, but also its direction.

$$F = \frac{S \cdot B}{\sqrt{B \cdot B}} B$$

You should recognize this as the projection of S onto the line along B (with an extra factor of $|B|$ to account for the fact that the magnitude of the magnetic field affects the magnitude of F).

So let's say we have a beam of electrons and we point it at a magnet. What do we expect to happen? Well, we assume that the electrons in the beam all have random spin orientations. That is, there is no preferred direction for S. If our magnetic field is vertical and an electron has a horizontal spin vector, then the electron will not deviate from its path, because the dot product between the vectors is zero, thus the force is zero. If, however, the spin is in the direction of the magnetic field, then the electron will feel a large force and be pushed upwards. Because the spin vectors are assumed to have random orientations, the deviations caused by the magnet should be roughly Gaussian.

In reality, this is *not* what we find at all! Instead, if we project the electrons onto a phosphorescent screen, we see two dots corresponding to the spin vectors pointing precisely up and down. So, you say, maybe all of the electron spin vectors were pointing up and down to begin with. But if we rotate our magnet to change the orientation of the magnetic field, then we still get two dots corresponding to the spin being parallel and anti-parallel to the magnetic field. Somehow, the magnet is "changing" the spin vector of the electron. The goal of ESQM is to explain this phenomenon.

Also, let me mention that the electron dots we find correspond to a measured value of $|S| = \hbar/2$, where $\hbar$ is a fundamental constant that shows up all of the time in Quantum Mechanics. Though there is a deep reason for this particular measured value of $|S|$, ESQM alone cannot explain it.

3 Postulates of Electron Spin Quantum Mechanics

Just as we began our study of Linear Algebra with the axioms of a vector space, I will start by introducing the postulates of ESQM. The first four postulates make a lot of sense (to a physicist at least). Unfortunately, the last postulate is not very aesthetic and is in some ways the Achilles' heel of Theoretical Physics. All five postulates generalize to other areas of Quantum Mechanics.

Postulate 1 *A* spin state *of an electron corresponds to a normalized vector in* $\mathbf{C}^2$*.*

Postulate 2 *An* observable *is a* self-adjoint *element of* $\mathrm{M}_{2\times 2}(\mathbf{C}^2)$.

Postulate 3 *A* measurement *of an observable can only result in an* eigenvalue *of the observable.*

Postulate 4 *The* probability *of getting such an eigenvalue is given by an equation that we shall "derive" later on.*

Postulate 5 *After measuring an observable, the spin state magically changes to the eigenvector corresponding to the measured eigenvalue.*

These postulates contain a lot of definitions, so let's look more closely at their meanings.

4 Spin States

A *spin state* of an electron corresponds to a normalized vector in $\mathbf{C}^2$.

Normalization of spin states is important because is allows us to simplify parts of the theory (including the probability postulate). This is similar to the reason why it is easier to project a vector to the line along U when U is a unit vector. The normalization condition for complex vectors is nearly identical to the normalization condition for real vectors. For

$$X = \begin{pmatrix} a \\ b \end{pmatrix} \text{ with } a, b \in \mathbf{C}$$

the normalization condition is

$$|a|^2 + |b|^2 = 1 \text{ where } |a|^2 = \bar{a}a.$$

However, we also want this normalization condition to correspond to our previous result that $|X| = \sqrt{X \cdot X}$. This means that we will need to refine our definition of the dot product when working with complex numbers.

$$\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} c \\ d \end{pmatrix} = \bar{a}c + \bar{b}d$$

If a, b, c and d are real, then this is the same as the dot product we have seen before. Note, however, that the dot product does not commute with complex numbers.

$$A \cdot B = \overline{B \cdot A}$$

With the real numbers, we could rewrite the dot product as $A \cdot B = A^t B$. In complex numbers the dot product becomes

$$A \cdot B = A^* B$$

where the asterisk (*) means to take the adjoint of the matrix. The adjoint is the conjugate transpose of the matrix; that is, we take the transpose of the matrix, then complex conjugate all of the elements. So for X defined above

$$X^* = \begin{pmatrix} \bar{a} & \bar{b} \end{pmatrix}.$$

5 Observables

An *observable* is a *self-adjoint* element of $\mathrm{M}_{2\times 2}(\mathbf{C}^2)$. A *measurement* of an observable can only result in an *eigenvalue* of the observable.

The key feature of quantum mechanics is that *every* measurable quantity has an associated observable. There is an observable corresponding to position, to momentum, to angular momentum, to energy, and so on. In ESQM, our observables correspond to the magnitude of the electron spin in a given direction.

Clearly, the observables are linear transformations on the spin states, but a common mistake is to think that when an when an observable M acts on the spin state X we get a new spin state MX. This isn't the way it works at all. When we measure an observable, the spin state indeed changes, but by Postulate 5, it changes into an eigenvector of the observable and *not* into MX. As we shall see, this causes some theoretical problems, too, which have yet to be resolved.

Going back to the other postulates, by Postulate 3, a measurement of an observable can only result in an eigenvalue of the observable. This is the most powerful statement of ESQM. It means that if we make measurements in a lab, we can "backsolve" the characteristic equation to go from the eigenvalues to a possible observable. If we want our theory to make sense, then these eigenvalues better be real numbers, because we never measure, say, a distance of $(3 + 2i)$ meters.

We can actually show that in ESQM, the eigenvalues are *guaranteed* to be real. Postulate 2 tells us that the observables must be self-adjoint. That is, for an observable M, $M^* = M$. This fact leads to an important theorem.

Theorem 1 *A self-adjoint linear transformation has only real eigenvalues.*

We will do this in the two by two case by directly solving the characteristic equation. First let's find the form of a self-adjoint matrix.

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad M^* = \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix}$$

Solving for each coordinate, we find that

$$M = \begin{pmatrix} a & b \\ \bar{b} & d \end{pmatrix} \text{ where } a, d \in \mathbf{R} \text{ and } b \in \mathbf{C}$$

Looking at the characteristic equation, $t^2 - (\mathrm{tr} M)t + \det M = 0$, we know this will have two real solutions if the argument under the radical in the quadratic formula is non-negative. In other words, if

$$(\mathrm{tr} M)^2 - 4 \det M \geq 0$$

Solving this for our self-adjoint M we find

$$(a + d)^2 - 4(ad - b\bar{b}) = (a - d)^2 + |b|^2 \geq 0.$$

So indeed, the eigenvalues of our observables are real.

Another important property of self-adjoint transformations is that

$$A \cdot (MB) = (MA) \cdot B$$

This follows from the definition of the dot product and mirrors a calculation we did in the case of real numbers for symmetric matrices.

$$A \cdot (MB) = A^* MB = A^* M^* B = (MA)^* B = (MA) \cdot B$$

For certain vector spaces where there isn't an easily defined conjugate transpose, the dot product condition becomes the definition of a self-adjoint transformation.

6 The Spin Observables

Before we can discuss the meaning of the remaining postulates we should try to find the observables that correspond to our electron spin experiment. We can think of the magnet as acting as a measurement of S. As I mentioned above, the values of $|S|$ that resulted from the experiment were $\hbar/2$ (the up state) and $-\hbar/2$ (the down state). Thus, we are looking for observables with eigenvalues $\hbar/2$ and $-\hbar/2$. We can write down the characteristic equation:

$$\left(t - \frac{\hbar}{2}\right)\left(t + \frac{\hbar}{2}\right) = t^2 - \left(\frac{\hbar}{2}\right)^2 = 0$$

This tells us that the trace of this observable has to be zero and the determinant of the observable has to be $-(\hbar/2)^2$. One such matrix is

$$S_z = \left(\frac{\hbar}{2}\right)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The subscript z indicates that we are measuring spin in the vertical (i.e z) direction.

Let's find the eigenvectors of this observable. Even better, we can find an eigenspinor of the observable, which means an eigenvector that satisfies the spin state condition of normalization. It's not too difficult to show that:

$$E_{\hbar/2} = E_{\uparrow z} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad E_{-\hbar/2} = E_{\downarrow z} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The $\uparrow z$ and $\downarrow z$ notation is a big more convenient when working with spin. So convenient, in fact, that physicists often drop the E altogether and just write $|\uparrow z\rangle$! We will stick to to $E_{\uparrow z}$ notation, though.

Note that $E_{\uparrow z}$ and $E_{\downarrow z}$ are orthogonal. In fact, this is a completely general phenomenon relating to an important theorem.

Theorem 2 *Given any self-adjoint matrix M, you can find an* orthonormal *basis of the vector space consisting of eigenvectors of M.*

This theorem is the generalization to complex vector spaces of the result from symmetric matrices in real vector spaces. The fact that we can find such an orthonormal basis is quite important for probability theory.

Now that we know what happens in the z direction, what if we now look in the x or y direction? It's not entirely clear how we should go about find those spin observables. We want these observables to satisfy some sort of orthogonality condition and we want linear combinations of them to represent measuring spin in, say, the direction of 45° latitude and 30° longitude. It turns out that the best way to do so is to define S_x and S_y as follows:

$$S_x = \left(\tfrac{\hbar}{2}\right)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad S_y = \left(\tfrac{\hbar}{2}\right)\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

And the corresponding eigenspinors:

$$E_{\uparrow x} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \quad E_{\uparrow y} = \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix}$$
$$E_{\downarrow x} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \quad E_{\downarrow y} = \begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix}$$

Note that multiplying any of these eigenspinors by a complex number of magnitude 1 does not change the normalization condition (for example, multiplication by i, $-i$, $(1+i)/\sqrt{2}$, etc.). This is equivalent to the case in real numbers where if X is a unit eigenvector, $-X$ is also a unit eigenvector. As advertised, the eigenspinors of S_x form an orthonormal basis of $\mathbf{C}^2$ and the eigenspinors of S_y form an orthonormal basis of $\mathbf{C}^2$.

What about directions other than x, y and z? For spherical coordinates θ and ϕ, the spin observable in this direction is

$$S_{\theta\phi} = \cos\phi \sin\theta S_x + \sin\phi \sin\theta S_y + \cos\theta S_z.$$

Finally, to show how beautifully symmetric these operators are, note that S_x, S_y, S_z and I form a basis for $\mathrm{M}_{2\times 2}(\mathbf{C}^2)$ when the scalars are complex, and a basis for the self-adjoint subspace when the scalars are real.

7 Probability Theory

The *probability* of getting a particular eigenvalue is given by an equation that we shall "derive" right now.

Now that we know what the spin observables are, we can look at the probabilistic interpretation of the spin states. If the spin state is already an eigenspinor, then we might intuitively expect that there is a 100% chance of getting the eigenvalue corresponding to that eigenspinor. What if the spin state is in some linear combination of the eigenspinors? We know that the eigenspinors of any spin observable form an orthonormal basis for $\mathbf{C}^2$, so we can write an arbitrary spin state as

$$X = aE_\uparrow + bE_\downarrow$$

where $a, b \in \mathbf{C}$ and $|a|^2 + |b|^2 = 1$. The normalization condition on a and b is necessary in order for $|X| = 1$, and there is no direction subscript on E because this works in general. Let's take the dot product of X with $E_\uparrow$ and $E_\downarrow$.

$$E_\uparrow \cdot X = a(E_\uparrow \cdot E_\uparrow) + b(E_\uparrow \cdot E_\downarrow) = a$$

$$E_\downarrow \cdot X = a(E_\downarrow \cdot E_\uparrow) + b(E_\downarrow \cdot E_\downarrow) = b$$

Note that

$$|E_\uparrow \cdot X|^2 + |E_\downarrow \cdot X|^2 = 1.$$

We also know that the probability of getting some value for a measurement is 1. Because the only eigenvalues are $\hbar/2$ (up state) and $-\hbar/2$ (down state), it follows that:

$$P(\uparrow) + P(\downarrow) = 1$$

"Equating" this formula with the one above it, it makes some intuitive sense that

$$P(\uparrow) = |E_\uparrow \cdot X|^2$$

$$P(\downarrow) = |E_\downarrow \cdot X|^2$$

Indeed, these are the postulated probability equations.

Let's do a concrete example to show how probability theory works. First, we need set up a spin state for our electron. The easiest way to do so is to use Postulate 5. That is, we measure the spin in, say, the z direction and then isolate all of the electrons that yielded $|S_z| = \hbar/2$. The spin state of these electrons will then be:

$$X = E_{\uparrow z} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Now we want to know what the probability of measuring the $|\uparrow x\rangle$ and $|\downarrow x\rangle$ states are. By the probability equations:

$$P(\uparrow x) = |E_{\uparrow x} \cdot X|^2 = \left(\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)^2 = \frac{1}{2}$$

$$P(\downarrow x) = |E_{\downarrow x} \cdot X|^2 = \left(\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)^2 = \frac{1}{2}$$

So there is a fifty-fifty chance of getting the $|\uparrow x\rangle$ and $|\downarrow x\rangle$ states. What if we go back and measure S_z? Because of Postulate 5, we have to consider the fact that the measurement of S_x changed the spin state to either $E_{\uparrow x}$ and $E_{\downarrow x}$. It turns out that the probability of getting the up and down states in the z direction are now fifty-fifty, even though we started with $E_{\uparrow z}$ at the very beginning!

8 Collapse of Spin States (and Theoretical Physics!)

After measuring an observable, the spin state magically changes (i.e. collapses) to the eigenvector corresponding to the measured eigenvalue.

We've been using Postulate 5 a lot in this discussion, but we haven't quite examined why it is necessary. It turns out that without Postulate 5, Quantum Mechanics would not correspond to experimental phenomena. Here is an experiment that you can do to prove it to yourself.

The polarization of light has mathematics that are quite similar to ESQM. A polarization filter acts kind of like the magnet in electron spin, except it not only measures the polarization state of the light but also blocks all light that does not have the correct polarization angle. Light from the sun is assumed to have random polarization directions. If we hold a vertical polarization filter to the sun, then the light becomes vertically polarized. If we then put a horizontal polarization filter behind the vertical one, then all of the light is blocked, because the vertical and horizontal polarization states are orthogonal.

What would happen if we put a third filter in between the other two and aligned it at 45°? We might naively think that if two filters block all of the light, then three filters would block all of the light, too. Well, as you might expect, this is not what happens at all. After the vertical filter, the light is polarized vertically. The 45° state is not orthogonal to the vertical state, though, so we can use probability theory to determine the amount of light that passes through the 45° filter. By Postulate 5, the light is now in the 45° state, and this state is not orthogonal to the horizontal state either, so we use probability theory again to determine the amount of light that passed through the horizontal filter.

It turns out that in this example, each filter blocks 1/2 of the light it receives, so that total transmission rate is 1/8. So by inserting a third filter, we increase the amount of light from 0 to 1/8. Strange quantum effects indeed!

So Postulate 5 is necessary in order to explain this and other experimental phenomena, but there is one major problem with it: it doesn't explain *how* the states collapse. In all other areas of physics, we are accustomed to studying the time evolution of systems. So what is the time evolution of the spin state under a measurement? Well, right now, it seems to be magic.

There are actually a few theories that try to explain state collapse, but most of them have magical names themselves (i.e. The Many Worlds Theory). State collapse was one of the primary reasons why Einstein never believed in Quantum Mechanics. Hopefully, further studies in particle physics will be able to explain Postulate 5. Or maybe, like the Parallel Line Postulate, it will turn out that state collapse is not a general phenomena but depends on the nature of the system.

9 Additional Algebraic Properties of Spin

If the postulates of ESQM only applied to electron spin then they wouldn't be very interesting. As I mentioned above, the postulates are actually completely general. For any system, we can describe its state by an element of a (possibly infinite) vector space, and we can define observables as linear transformations on this vector space. In order to do so, we need to have a vector space with a well-defined dot product (the self-adjoint condition can be derived from the dot product). These vector spaces are called Hilbert Spaces and are at the heart of Algebraic Quantum Mechanics.

So why study spin at all if it is just part of a larger theory? Well, the larger the theory, the harder it is to work with. We find that the algebraic properties of spin are very similar to the algebraic properties of angular momenta in Quantum Mechanics. Similarly, the algebraic properties of angular momenta in Quantum Mechanics may have relevance for the study of particle properties such as charge, color, and mass. So if we understand spin, then we can understand more complex structures in physics. In fact, this summer I will be an UTRA fellow with Antal Jevicki in the Physics Department to study these kind of algebraic structures in M-Theory.

To give you a sense of some of the algebraic issues that physicists worry about, we can look at the concept of simultaneous diagonalization in Linear Algebra.

Theorem 3 *Matrices A and B are simultaneously diagonalizable if and only if $AB = BA$.*

The proof of this theorem can be found in FIS, exercises 5.2.17 (p. 272) and 5.4.25 (p. 312).

Because of spin state collapse, physicists are very interested in observables that have the same eigenvectors, because alternating measurements of these observables will yield consistent values. Such observables are called *compatible* observables. To measure the "degree" of compatibility between observables A and B, we can look at the *commutator* of A and B.

$$[A, B] = AB - BA$$

From Theorem 3, it follows that if $[A, B] = 0$, then A and B are simultaneously diagonalizable, which means that A and B are compatible observables.

Are the spin observables compatible?

$$\begin{aligned}
[S_z, S_x] = S_zS_x - S_xS_z &= \left(\tfrac{\hbar}{2}\right)^2 \left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) \\
&= \left(\tfrac{\hbar}{2}\right)^2 \left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right) \\
&= \hbar \left(\tfrac{\hbar}{2}\right) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\
&= i\hbar \left(\tfrac{\hbar}{2}\right) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\
&= i\hbar S_y
\end{aligned}$$

Not only are S_z and S_x incompatible observables, but the commutation relation has produced a scalar times one of the other spin observables. What about other pairs?

$$[S_x, S_y] = i\hbar S_z \quad [S_y, S_z] = i\hbar S_x$$

It is easy to show that any operator commutes with itself (i.e. $[A, A] = 0$) and that commutation is anti-commutative (i.e. $[A, B] = -[B, A]$), so we can described all of the commutation relations of the basic spin operators.

$[A,B]/i\hbar$	S_x	S_y	S_z
S_x	0	$-S_z$	S_y
S_y	S_z	0	$-S_x$
S_z	$-S_y$	S_x	0

This kind of "closure" of the commutation relation has important consequences in Theoretical Physics, and is at the heart of studying the symmetries of physical systems.

As a final note, we can define an operator

$$S^2 = S_x^2 + S_y^2 + S_z^2$$

and show that

$$[S^2, S_z] = 0 = [S^2, S_x] = [S^2, S_y] = [S^2, S_{\theta\phi}]$$

This means we can simultaneously measure the spin in a given direction as well as the observable S^2. If this seems a little fishy given what I've said about all of the basic spin operators being incompatible, note that S^2 has another name:

$$S^2 = (3\hbar^2/4)I$$

where I is the standard identity matrix. So what does it mean to "measure" the identity observable? In a real sense, it means to measure nothing at all.