Miscellaneous Mathematical Background

Blake C. Stacey

December 6, 2023

One reason that a physics course can be more painful than strictly necessary is that the student might have seen the mathematical prerequisites without having grown comfortable with them. It is very easy for the professor to think, "Well, they've taken three semesters of calculus by now, so obviously they know how to do such-and-such." Consequently, the professor will make a step from one equation to the next that the student might not follow, perhaps just because their memory was rusty.

It is also often the case that the mathematical background we assume for a physics course might be scattered across three or four different math textbooks, with significant fractions of each textbook irrelevant for physics purposes. Accordingly, in this chapter we will review as much as feasible of the mathematics that we will end up using. Parts of this review may be very "elementary" compared to the reader's current standing, in which case feel proud of how far you have come!

0.1 Combinatorics

To start, let's update an example from a handy little classic, which I first found an old physics book on a colleague's "miscellaneous" shelf: *University of Chicago Graduate Problems in Physics,* by Cronin, Greenberg and Telegdi (Addison–Wesley, 1967). It looked like fun, so I started working through some of it. The opening problem in the "statistical physics" chapter turns out to be a good place to begin our review. Editing slightly to adjust for shifts in the cultural milieu, the problem runs as follows:

> Asami is meeting Korra for lunch downtown. Korra is E blocks east and N blocks north of Asami, on the rectangular street grid of downtown Republic City. Because Asami is eager to meet Korra, her path never doubles back. That is, each move Asami takes must bring her closer to Korra on the street grid. How many different routes can Asami take to meet Korra?

Asami must traverse a total of $E + N$ block edges. At each intersection, she must choose whether to move east or to move north, unless she is on the same street as Korra, at which point she keeps going in a straight line along that street. Her path can be represented as a string of ns and es, of total length $E + N$ characters. The question is how many different ways such a string can be written. That is, given a total of $E+N$ slots, how many ways can we pick a subset of E of them to fill with es?

Phrasing the problem this way, we see that it's a job for *binomial coefficients.* The number of ways to pick K things out of a set of M items is

$$\binom{M}{K} = \frac{M!}{K!(M-K)!}. \tag{0.1}$$

This is one of the more enthusiastic formulas that one encounters in everyday mathematics. The factorial $M!$ of a positive integer M is the product of M with $M-1$ and $M-2$ and so on, down to 1:

$$M! = M(M-1)\cdots 1. \tag{0.2}$$

If we ever find ourselves needing to take the factorial of 0, we use the special rule that $0! = 1$. The factorial operation counts the number of ways to arrange M items in a line. The special rule $0! = 1$ amounts to saying that if we have *no* items, there's only one possible way to order them, *i.e.,* do nothing, because it's all we can do!

Here, we need the number of ways to pick E items out of a set of size $E + N$, which is

$$\binom{E+N}{E} = \frac{(E+N)!}{E!\,N!}. \tag{0.3}$$

This is what the problem asked us to find. Note that we would have gotten the same answer if we decided to pick N items out of $E + N$.

It's good practice to check the "limiting cases" of a solution. If we derive a complicated formula which applies across a wide swath of scenarios, and if we have an answer we trust for a special case, then we should check that our complicated calculation gives the right answer for that special case. Here, we know that if Korra and Asami start off on the same street, there's only one path which meets the "never doubling back" criterion. In such a situation, either $N = 0$ or $E = 0$. Let's plug that into our boxed equation:

$$\frac{(E+0)!}{E!\cdot 0!} = \frac{E!}{E!\cdot 1} = 1. \tag{0.4}$$

Simple enough!

Likewise, if the initial configuration has Korra one block north and one block east of Asami, then it's pretty plain that Asami has two possible paths that she can take: she can go north-and-then-east, or she can travel east-and-then-north. Here, $E = N = 1$, and our formula says

$$\frac{(E+N)!}{E!\,N!} = \frac{2!}{1!\,1!} = 2. \tag{0.5}$$

The figure in the Chicago book gives for specific numbers $E = 3$ and $N = 4$. Plugging these in, we find

$$\frac{(E+N)!}{E!\,N!} = \frac{7!}{3!\,4!} = \frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2}{3\cdot 2\cdot 4\cdot 3\cdot 2} = 7\cdot 5 = 35. \tag{0.6}$$

0.2 Trigonometry

Let's start with the most useful formula I learned later than I should have:

$$e^{i\theta} = \cos\theta + i\sin\theta. \tag{0.7}$$

This equation, due to Leonhard Euler, relates two ways of labelling points in the complex plane. The left-hand side is in the spirit of polar coordinates: it locates a point on the unit circle by turning an angle θ from the horizontal. The right-hand side is rectangular in character: it tells us to reach a destination by taking a step of length $\cos\theta$ in the horizontal direction and then one of length $\sin\theta$ in the vertical direction.

First trick: take the derivative of Eq. (0.7) with respect to θ.

$$\frac{d}{d\theta}e^{i\theta} = ie^{i\theta} \tag{0.8}$$

$$= i(\cos\theta + i\sin\theta) \tag{0.9}$$

$$= -\sin\theta + i\cos\theta. \tag{0.10}$$

The real part of Eq. (0.7), which was a cosine, has become minus a sine; the imaginary part, which was a sine, has become a cosine. This meshes with our basic calculus lessons.

Now, multiply two numbers of the form $e^{i\theta}$:

$$e^{iA}e^{iB} = e^{i(A+B)}, \tag{0.11}$$

by the rules of exponentials. But using Eq. (0.7),

$$e^{i(A+B)} = \cos(A+B) + i\sin(A+B). \tag{0.12}$$

This must be equal to the product

$$e^{iA}e^{iB} = (\cos A + i\sin A)(\cos B + i\sin B) \tag{0.13}$$

$$= \cos A\cos B - \sin A\sin B + i\sin A\cos B + i\sin B\cos A. \tag{0.14}$$

Matching the real parts of the two expressions, we find the addition formula for cosines:

$$\cos(A+B) = \cos A\cos B - \sin A\sin B. \tag{0.15}$$

Likewise, by matching the imaginary parts, we find the addition formula for sines:

$$\sin(A+B) = \sin A\cos B + \sin B\cos A. \tag{0.16}$$

Setting $A = B$ in these expressions yields the double-angle formulae,

$$\cos(2A) = \cos^2 A - \sin^2 A, \tag{0.17}$$

$$\sin(2A) = 2\sin A\cos A. \tag{0.18}$$

Combining the double-angle formula for cosines with the Pythagorean identity,

$$\sin^2 A + \cos^2 A = 1, \tag{0.19}$$

we can equivalently say that

$$\cos(2A) = 2\cos^2 A - 1 = 1 - 2\sin^2 A. \tag{0.20}$$

It can also be useful to know that

$$\tan(2A) = \frac{\sin(2A)}{\cos(2A)} \tag{0.21}$$

$$= \frac{2\sin A\cos A}{\cos^2 A - \sin^2 A} \tag{0.22}$$

$$= \frac{2\sin A\cos A}{\cos^2 A - \sin^2 A}\cdot\frac{1/\cos^2 A}{1/\cos^2 A} \tag{0.23}$$

$$= \frac{2\tan A}{1-\tan^2 A}. \tag{0.24}$$

If we flip the sign of the angle in Eq. (0.7), we get the same thing as if we took the complex conjugate:

$$e^{-i\theta} = \cos(-\theta) + i\sin(-\theta) = \cos\theta - i\sin\theta. \tag{0.25}$$

Adding this to Eq. (0.7), the imaginary parts cancel out:

$$e^{i\theta} + e^{-i\theta} = 2\cos\theta. \tag{0.26}$$

We can, therefore, write the cosine function as the sum of two complex exponentials:

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}. \tag{0.27}$$

This can be handy when one has to manipulate expressions involving combinations of exponential and trigonometric functions. We also have that

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}. \tag{0.28}$$

Well, that's high-school trigonometry sorted.

Seriously! Life would have been so much easier then — well, that part of life *inside* the classroom, anyway — if I'd only been properly introduced to the idea that a complex number stands for a stretch and a twist. (See http://math.ucr.edu/home/baez/trig.html for a concurring opinion.)

0.3 Rotation Matrices

My group theory teacher, Prof. Daniel Freedman, had some interesting professorial habits. When invoking some bit of background knowledge with which we were all supposed to have been familiar, he would say, "As you learned in high school..." Typically, this would make a lecture sound a bit like the following:

"To finish the proof, note that we're taking the trace of a product of matrices. As you learned in high school, the trace is invariant under cyclic permutations..."

Prof. Freedman also said "seventeen" for "zero" from time to time. After working out a long series of mathematical expressions on the blackboard, showing that this and that cancel so that the overall result should be nothing, with the students alternating their glances between the board and their notes, he would complete the equation and proclaim, "Equals seventeen!" At which point, all the students look up and wonder, momentarily, what they just missed.

"Here, we're summing over the indices of an antisymmetric tensor, so by exchanging i and j here and relabeling there, we can show that the quantity has to equal the negative of itself. The contraction of the tensor is therefore, as you learned in high school — seventeen!"

One day, I managed to best his line. I realized that the formula currently on the board had to work out to *one,* not zero, so when he wrote the equals sign, paused

and turned to the class with an inquiring eye, I quickly raised my hand and said, "Eighteen!"

Incidentally, when it came to the *truly* simple topics like Euler's formula and trigonometric identities, we were supposed to have learned that in middle or elementary school.

In this section, we'll talk about one of the things Prof. Freedman said we should have covered in high school: the *rotation matrices* for two- and three-dimensional rotations. This will give us the quantitative, symbolic tools necessary to talk about *commutativity.*

We start with a point in the (x, y) plane, and we break out the trig functions to relate Cartesian and polar coordinates:

$$x = r\cos\theta,\ y = r\sin\theta. \tag{0.29}$$

Now, what happens when we take the point (x, y) and rotate it around the origin by some fixed amount? The *distance* to the origin doesn't change, but the *angle* certainly does: the angle shifts by the amount we turn. If the original angle was θ and we call the size of the turning ϕ, then the *new* point will be located at the position labeled by r and $\theta + \phi$.

Where is this in Cartesian coordinates? Well, adopting the common practice of using apostrophes or "primes" to denote transformed variables, then the previous formula gives us the values of "x prime" and "y prime":

$$x' = r\cos(\theta + \phi),\ y' = r\sin(\theta + \phi). \tag{0.30}$$

We can easily work out what these expressions are in terms of the sines and cosines of θ and ϕ, just using the addition formulas.

$$x' = x\cos\phi - y\sin\phi, \tag{0.31}$$

$$y' = x\sin\phi + y\cos\phi. \tag{0.32}$$

There's a certain pleasing regularity to the appearance of the cosines and sines. We can get at this more neatly if we *separate* the coordinate variables x and y from the functions which transform them. First, let's consider x and y joined together: they're both labels for marking the same point in space, so we should by rights consider them something of a "married couple." We'll write this combined quantity as a vertical arrangement enclosed in parentheses, thusly:

$$\begin{pmatrix} x \\ y \end{pmatrix}. \tag{0.33}$$

This is one common way of writing a *vector,* a quantity with both magnitude and direction. (In our case, we denoted the magnitude by r and the direction by θ.) "Seven kilometers" is not a vector, but "seven kilometers due south" is. After a transformation like a rotation, our vector will have primes on its components:

$$\begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{0.34}$$

We'll pass over any moral implications in the fact that a rotation "mixes up" one member of this "marriage" with another.

The vector before the transformation and the new vector produced by the transformation are related, and in writing the vectors in this fashion, we can "pull out" the part of the equations which is the transformation proper, separating it from the variables on which the transformation acts:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{0.35}$$

The square array of expressions is called a *matrix.* Here, we are *multiplying* a vector by a matrix to get a new vector. In fact, a vector is just a one-column matrix. (If we wanted our vectors to be written short and wide instead of tall and skinny, we could use *row vectors* instead of *column vectors,* which would just involve shuffling entries around on the page.) Matrices can be multiplied together just like numbers can, using a rule which you might be able to figure out from this example, remembering what x' and y' are:

$$\begin{pmatrix} x\cos\phi - y\sin\phi \\ x\sin\phi + y\cos\phi \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{0.36}$$

To get the entry in the first row of the product vector, we multiply the first row of the *rotation matrix* by the first (and only) column of the original vector, item-by-item, and add them up. To get the second row of the result, we multiply the *second* row of the matrix by the column entries, item-by-item, and add *them* up. Here, we're multiplying a 2×2 matrix by a 2×1 matrix to get another 2×1 matrix, but we could multiply bigger matrices. If we were multiplying two 40×40 matrices, the entry in "cellblock 11, 38" would be the product of the entries in the 11^{th} row of the first times those in the 38^{th} column of the second, added together.

Matrix multiplication is one of those things which takes practice to get right. Fortunately, the definition for *matrix addition* is considerably simpler: just add the corresponding entries!

One special case of matrix multiplication is worth calling attention to. If we multiply a matrix into a column vector that is 1 in a single spot and 0 otherwise, something interesting happens:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}, \tag{0.37}$$

and likewise,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}. \tag{0.38}$$

Multiplying into a column vector of this form reads out the corresponding column of the matrix. We could in fact build up matrix multiplication from this as a starting point. First, observe that

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{0.39}$$

Now, we use the reading-out-a-column property:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \left[x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] \tag{0.40}$$

$$= x \begin{pmatrix} a \\ c \end{pmatrix} + y \begin{pmatrix} b \\ d \end{pmatrix} \tag{0.41}$$

$$= \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} . \tag{0.42}$$

0.4 Matrix Manipulations

If matrix arithmetic is to be like ordinary numerical arithmetic, matrix multiplication should include *inverses.* If we take a number a which is not 0, then we know there is a number a^{-1} which when multiplied by a gives 1. Do matrices have inverses? Sometimes, they do.

We can understand the problem better by working through the idea explicitly for a 2×2 matrix. Let us deduce the inverse of the matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} . \tag{0.43}$$

We want to fill in the question marks in this equation:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} ? & ?? \\ ??? & ???? \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{0.44}$$

First, let's try to get the bottom left-hand entry to work out.

$$\begin{pmatrix} c & d \end{pmatrix} \begin{pmatrix} ? \\ ??? \end{pmatrix} = 0. \tag{0.45}$$

We can satisfy this if we choose the column vector to have d in the top slot and $-c$ in the bottom:

$$\begin{pmatrix} c & d \end{pmatrix} \begin{pmatrix} d \\ -c \end{pmatrix} = 0. \tag{0.46}$$

Now, we try for the top right corner:

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} ?? \\ ???? \end{pmatrix} = 0. \tag{0.47}$$

We fill this in with

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} -b \\ a \end{pmatrix} = 0. \tag{0.48}$$

Combining these two column vectors into a 2×2 matrix, we get

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & -bc + ad \end{pmatrix} . \tag{0.49}$$

We're almost there! We need to rescale our inverse matrix by the quantity $1/(ad-bc)$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \left[\frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \right] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{0.50}$$

We called the original matrix M, so we can call the inverse matrix M^{-1}. We have shown that

$$M^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{0.51}$$

To find the inverse of a 2×2 matrix, we flip the elements on the diagonal, take the negatives of the off-diagonal entries and then divide the whole matrix by the quantity $ad - bc$. If $ad - bc = 0$, then we cannot divide by it, so the matrix has no inverse.

This quantity is important enough that it deserves a name. We call it the *determinant* of the matrix:

$$\det(M) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc. \tag{0.52}$$

Note that the determinant of the identity matrix is 1, and the determinant of M^{-1} is the inverse of the determinant of M. This suggests a general rule, which turns out to be true: for two matrices M_1 and M_2,

$$\det(M_1)\det(M_2) = \det(M_1 M_2). \tag{0.53}$$

It would be possible to meet the matrix determinant much earlier in one's education than is typical. High-school algebra touches upon matrices as a tool for solving systems of linear equations. For instance, we can write the pair of equations

$$ax + by = g\,, \tag{0.54}$$
$$cx + dy = h \tag{0.55}$$
$$\tag{0.56}$$

in matrix-vector form, thusly:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} g \\ h \end{pmatrix}. \tag{0.57}$$

The solution is then found by multiplying both sides by the inverse of the matrix:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} g \\ h \end{pmatrix}. \tag{0.58}$$

Thinking geometrically, we know that if the lines defined by our two simultaneous equations are parallel, there won't be any solution for x and y. For the lines to be parallel, their slopes must be the same. Solving both equations for y, we find the slopes in question:

$$y = -\frac{a}{b}x + \frac{g}{b}\,, \tag{0.59}$$
$$y = -\frac{c}{d}x + \frac{h}{d}\,. \tag{0.60}$$

The slopes are equal if $a/b = c/d$; or, rearranging, $ad = bc$. So, the lines are parallel (or identical) just when the determinant is zero, which is the condition that the matrix has no inverse.

Consider the matrix

$$M = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}, \tag{0.61}$$

and what it does to the unit vectors in the x and y directions:

$$M \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \; M \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}. \tag{0.62}$$

The tips of these vectors are two corners of a square with area 1. Multiplying by M gives us a rectangle with area 6, which is the determinant of M. Next, consider the matrix

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{0.63}$$

which interchanges the x and y directions. The matrix T has a determinant of -1. Its action is to leave the area of the unit square unchanged, but to flip it about its diagonal.

These two examples illustrate a general property of the determinant. Its value tells us the area of the parallelogram that the matrix transforms the unit square into; and if there is a flip involved in the transformation, the determinant picks up a minus sign. This is not too difficult to establish. We just follow what matrix multiplication does to the corners of the unit square, find the area of the big rectangle that encloses the resulting parallelogram, and subtract away the triangles surrounding it.

The determinant is one of the two most important mappings which turn matrices into numbers. The other is the *trace,* which is the sum of the diagonal elements:

$$\mathrm{tr} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a + d. \tag{0.64}$$

The trace operation also interacts with matrix multiplication. An important property is that if we have three matrices M_1, M_2 and M_3, then

$$\mathrm{tr}(M_1 M_2 M_3) = \mathrm{tr}(M_2 M_3 M_1) = \mathrm{tr}(M_3 M_1 M_2). \tag{0.65}$$

If we wrote the labels M_1, M_2 and M_3 clockwise around a circle, we could get these three orderings by picking a starting label and reading around, clockwise. Such rearrangements are called *cyclic permutations.*

So far, we have only defined the determinant and the trace for 2×2 matrices, but we can also define them for bigger matrices, in such a way that their nice properties still hold true.

0.5 3D Rotations

Whew! We've built up a fair amount of mathematical machinery, and it would be a shame if we didn't *use* it. What can we do with our rotation matrices?

While rotations in two dimensions commute, rotations about *different axes in three dimensions* don't: in 3D, the order of rotation operations matters. Rotation matrices provide a way of teasing out the essence of this odd behavior and presenting the geometrical phenomenon in a useful form.

First, we need to bulk up our machinery from two dimensions to three. This just means making the matrices bigger.

In 3D, we can rotate around the axis of our choice. To get started, we'll pick three axes such that x denotes the distance to the right, y denotes the distance up and z denotes the distance forward, "out of the chalkboard." Negative numbers imply a placement in the opposite direction.

Up until now, we've been discussing rotations which mix together x and y. In this picture, we can see that such an operation is a turn around the z-axis. We'll denote it by $R_z(\phi)$, where ϕ is the angle through which we turn. In matrix form,

$$R_z(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{0.66}$$

We've just added an extra row and column to go from 2D to 3D; this extra realm is filled with zeros except for the lower right corner, which is 1. This is the matrix way of saying that a rotation around the z-axis leaves z unchanged.

We can also write a matrix for rotation around the x-axis. It will *also* have sines and cosines of the angle ϕ, and a row and column containing zeros except for a solitary 1. Why? Because a rotation around the x-axis leaves the value of x unchanged. We can figure out what goes in the other spaces by observing that a rotation of $\pi/2$ radians (90 degrees) around the x-axis will take a point on the y-axis onto the z-axis. This means that the entry in the second column, third row must be the sine of ϕ (remember that the sine of $\pi/2$ is 1). The same rotation will take a point on the z-axis onto the negative y-axis, so the entry in the third column, second row must be $-\sin\phi$.

$$R_x(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix}. \tag{0.67}$$

We can figure out a matrix for rotations about the y-axis in much the same way. Skipping ahead to the punchline,

$$R_y(\phi) = \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix}. \tag{0.68}$$

Note what happens to each of these matrices when we rotate by *zero* radians. Because the sine of zero is zero and the cosine of zero is one, each of the rotation matrices reduce to the same thing, a 3×3 matrix whose entries are all zero, except along the *diagonal,* where they are all identically 1. This "do nothing" operation is called the

identity, and the *identity matrix* is

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{0.69}$$

Multiplying any matrix by the identity gives back the original matrix; it's the equivalent of the number 1 on the familiar number line. (Identity matrices exist for all sizes $n \times n$ but here we'll only need the 3×3 version.)

Now that we've graduated high school, it's time to go to college!

We're going to take "baby steps" and consider *infinitesimal* rotations. That is, we're going to look at the form these matrices take when the angle ϕ is very small indeed. Recall that for small angles, the sine of the angle is just the angle (when you're working in radians). Cosines of such small angles are approximately 1. This means that each of the rotation matrices becomes a matrix containing only ones, zeros and two instances of the small angle. Furthermore, each entry on the *diagonal* of a rotation matrix is either 1 or a cosine, but in the small-angle regime, the cosines become 1, so the diagonal is all ones, just like the identity matrix.

We're going to take advantage of this fact and write the rotation matrices for small angles as the identity matrix *plus* another matrix. We can be even sneakier if we realize that the entries in this other matrix are all either 0 or the small angle, which by convention we'll call epsilon, ϵ. We can *factor out* the number ϵ to give a matrix which contains just zeros and ones. (Multiplying a number times a matrix just multiplies each element in the matrix by that number.) Thus we say,

$$R_j(\epsilon) = I + \epsilon J_j. \tag{0.70}$$

Here, the letter j denotes any one of the axes x, y or z — we mean that this equation is true for all of them, and we don't want to write the same thing three times. I get the following for the matrix J_x:

$$J_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{0.71}$$

We can turn around and substitute this into the previous equation: multiply each entry by ϵ, add 1 to each diagonal entry, and we get back the small-angle version of R_x.

We can play the same game with the y-axis rotation matrix to get J_y:

$$J_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \tag{0.72}$$

Again, multiply by ϵ and add the identity matrix to check. Finally,

$$J_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{0.73}$$

The three matrices J_j are called the *generators* of the rotation transformations.

Remember, we're trying to study what happens when we perform successive rotations around different axes. So, let's multiply the generators of x- and y-axis rotations. The matrix multiplication isn't so bad, because everything is either 0 or 1, and the result is the following:

$$J_x J_y = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{0.74}$$

But, oh la, what happens when we multiply the same matrices *in the other order*? As it turns out, the procedure for matrix multiplication I sketched earlier means that *matrix multiplication is not commutative.*

$$J_y J_x = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{0.75}$$

The two products, taken in opposite orders, are *different!* By how much do they differ? Well, we can just subtract one from the other:

$$J_x J_y - J_y J_x = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = J_z. \tag{0.76}$$

Surprise! The difference is just the *third* generator, J_z.

That's an odd enough coincidence that we should be tempted to press further. What about the generators of x- and z-axis rotations? First, try one order:

$$J_x J_z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{0.77}$$

Then, try the other:

$$J_z J_x = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{0.78}$$

Again, the answers differ, and again, we subtract to find how much they differ:

$$J_z J_x - J_x J_z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} = J_y. \tag{0.79}$$

Double whammy! The "mismatch" between the z and x generators is just the y generator.

Well, now there's nothing stopping us from trying the next pair:

$$J_y J_z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{0.80}$$

Again, we check the multiplication in the other order:

$$J_z J_y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{0.81}$$

The matrices differ by the amount...

$$J_y J_z - J_z J_y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = J_x. \tag{0.82}$$

All of the generators are tied up, somehow, such that combinations of two give back the third.

Here, we introduce a more compact notation, since we're going to be dealing with a good many quantities which fail to commute. For any two thingamajigs A and B,

$$[A, B] = AB - BA. \tag{0.83}$$

In this notation, the results we worked out a moment ago take on an interesting form:

$$[J_x, J_y] = J_z, \ [J_z, J_x] = J_y, \ [J_y, J_z] = J_x. \tag{0.84}$$

Notice the pattern of the subscripts. Each equation lists the generators in the order xyz, but sometimes "wrapped around" with a cyclic permutation.

This is, almost, the form of the rotation operations used in quantum mechanics. The next step, which should be a brief one to cover, will be to roll these three equations into one, taking advantage of their cyclic character.

0.6 Einstein Summation and Levi–Civita Symbols

Earlier, we took a look at rotations and found a way to summarize their behavior using commutator relations. Recall that the commutator of A and B is defined to be

$$[A, B] = AB - BA. \tag{0.85}$$

For real or complex numbers, the commutator vanishes, but as we saw, the commutators of *matrices* can be non-zero and quite interesting. We recognized that this would have to be the case, since we used matrices to describe rotations in three-dimensional space, and rotations about different axes in 3D do not commute. Looking at very small rotations, we also found that the commutators of *rotation generators* were tied up together in a way which involved cyclic permutations. Today, we'll express this discovery more neatly, using the *Einstein summation convention* and a mathematical object called the *Levi–Civita tensor.*

First, let's ground ourselves by noting a few basic properties of commutators. We observe that the commutator of anything with itself vanishes:

$$[A, A] = 0, \tag{0.86}$$

and that swapping the order of the bracketed symbols introduces a minus sign:

$$[A, B] = -[B, A]. \tag{0.87}$$

In the previous section, we looked at *infinitesimal* rotations, where we twisted through a very small angle ϵ, and we found that the matrix for rotating around axis j took the following form:

$$R_j(\epsilon) = I + \epsilon J_j. \tag{0.88}$$

The matrices J_j were comprised of ones and zeros, and they satisfied the following commutator relationships:

$$[J_x, J_y] = J_z,\ [J_z, J_x] = J_y,\ [J_y, J_z] = J_x. \tag{0.89}$$

Alternatively, we saw, these relations could be written this way:

$$[J_1, J_2] = J_3,\ [J_3, J_1] = J_2,\ [J_2, J_3] = J_1. \tag{0.90}$$

Now, we'd like to combine these three equations into one "package."

Let's think for a moment about *vectors,* which we used before as a way of combining several numbers into a "marriage" (with possibly many more than two spouses). We said that a vector has both a magnitude (say, 10 kilometers per hour) and a direction (*e.g.,* 35 degrees west of due north). We can write a vector as a list of numbers, organized if you'd like into a one-column or one-row matrix; the first number might give an easterly speed in kph, the second a northerly speed and the third a vertical speed, for example. Going between the magnitude-direction description and the list of "components" just requires a little trigonometry, which for the 2D case we explored earlier. (Well, you *could* go off and invent yourself a non-orthonormal basis in some abstract vector space, but patience! We're not quite there yet.)

A common operation done on vectors is the *dot product* or *inner product,* which means taking two vectors and multiplying their components, one by one, and adding the total up. The dot product of two velocity vectors in the 2D plane would be, for example, the product of vector 1's easterly component and vector 2's easterly component, *plus* vector 1's northerly component times vector 2's northerly component. That's enough of a mouthful to warrant expressing in symbols instead:

$$\vec{v} \cdot \vec{u} = v_1 u_1 + v_2 u_2 + \cdots + v_N u_N. \tag{0.91}$$

Here, we recognize that a vector could in principle have an arbitrarily large number of components. We can squeeze still harder by introducing a *summation* symbol, a big version of the Greek letter Σ:

$$\vec{v} \cdot \vec{u} = \sum_{i=1}^{N} v_i u_i. \tag{0.92}$$

This says exactly what we said before, but more stylishly. If we trust ourselves to remember where the "index" i starts and stops, we might omit the 1 and N for brevity:

$$\vec{v} \cdot \vec{u} = \sum_{i} v_i u_i. \tag{0.93}$$

The final step in streamlining the notation is due to Einstein, who realized that most of the time when you see indices repeated across variables — like the i we have on two variables here — you're going to be summing over them. So, why keep the summation sign around at all? Einstein found that he didn't have to, and in his honor, we call the idea that *summation is implied whenever you have repeated indices* the "Einstein summation convention." In our example, this means we can write the dot product like this:

$$\vec{v} \cdot \vec{u} = v_i u_i. \tag{0.94}$$

At last, our equations are starting to look like *physics!*

OK, it's time to stare down the commutators. Here is, again, what we're trying to study:

$$[J_1, J_2] = J_3,\ [J_3, J_1] = J_2,\ [J_2, J_3] = J_1. \tag{0.95}$$

Notice that each of these three equations has the form of a commutator between two generators yielding the third generator. We might try to summarize them by saying that in general, the commutator of generator i with generator j is the third generator, k. If we're more careful, we'll note that if i equals j, for example, the result is *zero.* It's also possible to get *minus* a generator on the right-hand side (how?), so we're really talking about getting *something times* that third generator:

$$[J_i, J_j] = (??)J_k. \tag{0.96}$$

Let's give that question mark a name. Whatever it is, its value will depend upon i, j and k. By convention, this object is identified using the Greek letter ϵ — it's not an infinitesimal angle; in fact, it's not an infinitesimal anything, but that's the letter everybody uses. (The Greek alphabet is only so big.) To make sure we don't forget which epsilon we're talking about, and to indicate what it depends upon, we attach i, j and k as subscripts:

$$[J_i, J_j] = \epsilon_{ijk} J_k. \tag{0.97}$$

Aha! We've caught ourselves repeating indices. Since each of i, j and k can be 1, 2 or 3, this equation really means

$$[J_i, J_j] = \epsilon_{ij1} J_1 + \epsilon_{ij2} J_2 + \epsilon_{ij3} J_3. \tag{0.98}$$

Hmmm. Have we made life any easier? That is, we've managed to squeeze three commutator equations into one, but we have to find out what this "epsilon thing" — or *Levi–Civita tensor* — is all about before we can claim success. To begin with, let's look at the commutator of J_1 with J_2. By our formula,

$$[J_1, J_2] = \epsilon_{12k} J_k, \tag{0.99}$$

which by the Einstein summation convention expands out to

$$[J_1, J_2] = \epsilon_{121} J_1 + \epsilon_{122} J_2 + \epsilon_{123} J_3, \tag{0.100}$$

but *since we already know* that J_1 with J_2 gives J_3, we can say that

$$\epsilon_{123} = 1, \tag{0.101}$$

and furthermore that

$$\epsilon_{121} = \epsilon_{122} = 0. \tag{0.102}$$

Following the same procedure with the other two commutators tells us that

$$\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1. \tag{0.103}$$

Here we see the cyclic nature of the commutators making itself manifest: If the subscripts are a *cyclic permutation* of the sequence 123, the value of the Levi–Civita symbol is 1.

Next, we note what happens when we *swap* the variables i and j. We know that in general, swapping the things inside the commutator brackets gives a minus sign:

$$[J_j, J_i] = -\epsilon_{ijk} J_k, \tag{0.104}$$

but looking at the formula where we introduced the Levi–Civita tensor, we see that

$$[J_j, J_i] = \epsilon_{jik} J_k. \tag{0.105}$$

Comparing these two, we observe a neat property:

$$\epsilon_{ijk} = -\epsilon_{jik}. \tag{0.106}$$

The Levi–Civita symbol is what we call *antisymmetric.* A "symmetry" means that we can transform an object and have it look the same as it did before; rotating a symmetrical vase around its axis, for example, yields a configuration which looks the same as the vase before the transformation. Here, we perform an operation — swapping the indices — and we find that the result is almost, but not quite the same as the initial quantity: it's got an extra minus sign.

As we mentioned earlier, for any value of i,

$$[J_i, J_i] = 0, \tag{0.107}$$

which implies that

$$\epsilon_{iik} = 0. \tag{0.108}$$

Note that this is consistent with the antisymmetry: If we *swap* the two instances of the index i, we get that

$$\epsilon_{iik} = -\epsilon_{iik}, \tag{0.109}$$

and the only way for a thing to equal *minus itself* is for that thing to be *zero.* Furthermore, looking at the original three commutators, we see that when a particular index is on the left-hand side, it *can't be on the right.* If J_1 is used on the left, J_1 won't be on the other side. This lets us write that

$$\epsilon_{iji} = \epsilon_{jii} = 0, \tag{0.110}$$

which is just the cyclic permutation of the statement we had two equations ago.

The fact that swapping indices introduces a minus sign means that, for example,

$$\epsilon_{213} = -\epsilon_{123} = -1. \tag{0.111}$$

We can equally well say that

$$\epsilon_{132} = -\epsilon_{312} = -1, \tag{0.112}$$

or that

$$\epsilon_{321} = -\epsilon_{231} = -1. \tag{0.113}$$

Summarizing what we've found,

$$\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1, \tag{0.114}$$

$$\epsilon_{132} = \epsilon_{213} = \epsilon_{321} = -1. \tag{0.115}$$

This exhausts the six possible ways of writing the numbers 1, 2 and 3 in a chosen order without repeating a digit. We observe that the cyclic permutation property holds whether the result is $+1$ (when the subscripts are "the right way around") or the result is -1 (when two indices have been swapped). Also, if any two indices have the same value, the result is always 0.

Our new toy, the Levi–Civita tensor (also known as the Levi–Civita symbol and the permutation tensor) is a *completely antisymmetric* object. We could define the analogous contraption for any number of dimensions, but we'll work with three most of the time. One finds this object in calculations of matrix determinants, vector cross products and other places, and while all of that will probably come up sooner or later, the use for which we will employ it now is in commutator relations:

$$[J_i, J_j] = \epsilon_{ijk} J_k. \tag{0.116}$$

Incidentally, physicists typically define the rotation generators with a factor of i, the square root of -1, worked in:

$$R_j(\theta) = I - i\theta J_j \text{ (for } \theta \text{ small)}. \tag{0.117}$$

This brings a factor of i into the commutator relation:

$$[J_i, J_j] = i\epsilon_{ijk} J_k. \tag{0.118}$$

Somehow, people live with themselves even after they use the letter i to represent two different things in the same equation.

The last boxed equation is all the information about 3D spatial rotations we'll need, distilled to its operatorial essence.

0.7 Eigenvalues and Eigenvectors

We characterized our rotation matrices by the *angle* of rotation they implemented and, in three dimensions, by the *axis* about which the rotation is effected. A rotation

about a given axis leaves a vector which points along that axis unaffected, but changes other vectors. We can turn this idea around and use it to define what a "characteristic vector" would be for any transformation represented in matrix form. A characteristic vector, or *eigenvector,* is a vector which a transformation leaves unchanged, or changes only in a simple way. Specifically, if M is a matrix, and $\vec{v}$ is an eigenvector of M, then

$$M\vec{v} = \lambda\vec{v}, \tag{0.119}$$

for some number λ. This number is called the *eigenvalue* of the matrix M associated with the eigenvector $\vec{v}$.

It is helpful to work out the eigenvalues of a 2×2 matrix, not only because the exercise illustrates the general idea, but also because it's a calculation which comes up over and over again in practice. We start by defining the matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{0.120}$$

Let $\vec{v}$ be an eigenvector of M, with λ its corresponding eigenvalue. Then

$$(M - \lambda I)\vec{v} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{0.121}$$

This will be satisfied if

$$\det(M - \lambda I) = 0. \tag{0.122}$$

The determinant is easy to work out:

$$\begin{aligned} \det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} &= (a-\lambda)(d-\lambda) - bc \\ &= ad - \lambda(a+d) + \lambda^2 - bc \\ &= \lambda^2 - \lambda \mathrm{tr} M + \det M. \end{aligned} \tag{0.123}$$

We solve the quadratic equation for the eigenvalue λ, obtaining the pair of solutions

$$\lambda_\pm = \frac{\mathrm{tr}M \pm \sqrt{(\mathrm{tr}M)^2 - 4\det M}}{2}. \tag{0.124}$$

You can check from this equation that

$$\lambda_+ + \lambda_- = \mathrm{tr}M, \tag{0.125}$$

and with just a little more work,

$$\lambda_+\lambda_- = \det M. \tag{0.126}$$

These properties will hold true for larger matrices: The sum of the eigenvalues will be the trace, and the product of the eigenvalues will be the determinant.

0.8 Dirac Notation

When we get to quantum mechanics, we will be using vectors in $\mathbb{C}^d$, and we will be writing them in *Dirac notation.* A "ket" is a column vector:

$$|\psi\rangle = \begin{pmatrix} z_1 \\ \vdots \\ z_d \end{pmatrix}, \tag{0.127}$$

and its corresponding "bra" is the row vector made by transposing and complex-conjugating a ket:

$$\langle\psi| = (z_1^* \ \cdots \ z_d^*). \tag{0.128}$$

Putting a bra and a ket together makes a bra[c]ket, which by the rules of matrix multiplication is just a number:

$$\langle\psi|\psi\rangle = |z_1|^2 + \cdots + |z_d|^2. \tag{0.129}$$

This implies that $\langle\psi|\psi\rangle \geq 0$, and it is only equal to 0 if the vector $|\psi\rangle$ is the zero vector itself. Moreover, for any pair of vectors $|\psi\rangle$ and $|\phi\rangle$,

$$(\langle\psi|\phi\rangle)^* = \langle\phi|\psi\rangle. \tag{0.130}$$

We will call

$$||\psi|| = \sqrt{\langle\psi|\psi\rangle} \tag{0.131}$$

the *norm* of the vector $|\psi\rangle$, and we'll say that a vector is *normalized* if its norm is 1. Any vector can be made normalized by dividing each element by the norm.

A *spanning set* for a vector space is a set of vectors such that any vector in the space can be written as a linear combination of those in the spanning set:

$$|\psi\rangle = \sum_{i=1}^{d} a_i |v_i\rangle. \tag{0.132}$$

For example, we can make a spanning set for $\mathbb{C}^2$ by taking the vectors

$$|v_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ |v_2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{0.133}$$

in which case

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1|v_1\rangle + a_2|v_2\rangle. \tag{0.134}$$

But we could equally well use the spanning set

$$|\tilde{v}_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ |\tilde{v}_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{0.135}$$

because we can say

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{a_1 + a_2}{\sqrt{2}}|v_1\rangle + \frac{a_1 - a_2}{\sqrt{2}}|v_2\rangle. \tag{0.136}$$

These are both examples of a kind of spanning set we will often use, the *orthonormal bases.* An orthonormal basis for a vector space is a spanning set $\{|v_i\rangle\}$ for that space that enjoys the property

$$\langle v_i|v_j\rangle = \delta_{ij}\,. \tag{0.137}$$

A set of vectors $\{|v_i\rangle : i = 1, \ldots, n\}$ is *linearly dependent* if there is a way to write one of them as a linear combination of the others. This is equivalent to saying that there is some set of coefficients, not all of which are zero, such that

$$\sum_{i=1}^{n} a_i|v_i\rangle = 0\,. \tag{0.138}$$

All linearly independent sets that span the same space must have the same size, a number that is called the *dimension* of the space.

Linear operators are transformations of a vector space back to itself that obey the rules

$$L(|v_1\rangle + |v_2\rangle) = L|v_1\rangle + L|v_2\rangle,\ L(\alpha|v_1\rangle) = \alpha L|v_1\rangle\,, \tag{0.139}$$

for arbitrary vectors $|v_1\rangle, |v_2\rangle$ and scalar α. This means that we can specify what a linear operator is in terms of its action upon a spanning set:

$$L\left(\sum_i a_i|v_i\rangle\right) = \sum_i a_i L|v_i\rangle\,. \tag{0.140}$$

Moreover, if $\{|v_i\rangle\}$ is an orthonormal basis, then $\langle v_i|L|v_j\rangle$ are the *elements* of L in that basis. If we have some linear operator L defined in a higher-level, more abstract way, we can "compile it down to machine language" by introducing an orthonormal basis and writing the operator as a matrix:

$$[L]_{ij} = \langle v_i|L|v_j\rangle\,. \tag{0.141}$$

The *adjoint* $L^\dagger$ of a matrix L is found by flipping it across its main diagonal and complex-conjugating all of the entries:

$$[L^\dagger]_{ij} = L^*_{ji}\,. \tag{0.142}$$

If a matrix is equal to its own adjoint, we call it *self-adjoint* or *Hermitian.* If the adjoint of a matrix is the *inverse* of that matrix, then we call that matrix *unitary.* The eigenvalues of a self-adjoint matrix are real, and those of a unitary matrix lie on the unit circle in the complex plane $\mathbb{C}$. A matrix L that commutes with its adjoint $L^\dagger$ is called *normal.* Both the self-adjoint and the unitary matrices are examples of normal matrices. Moreover, normal matrices have the property that they can be *diagonalized*: There exists some orthonormal basis $\{v_i\}$ such that

$$L = \sum_i \lambda_i|v_i\rangle\langle v_i|\,, \tag{0.143}$$

where $\{\lambda_i\}$ are the eigenvalues of L. The ket-bra combinations in this expression are an example of a construction that turns up all the time, wherein we turn a vector $|v\rangle$ into a special kind of linear operator we call a *projector*. A linear operator P is a projector if it satisfies $P^2 = P$. Projectors are self-adjoint by construction.

Suppose that P is the projector defined by the ket-bra combination (or "outer product") of $|v\rangle$ with itself:

$$P = |v\rangle\langle v| \,. \tag{0.144}$$

Now, let $|u\rangle$ be any other vector, and "sandwich" P on both sides:

$$\langle u|P|u\rangle = \langle u|v\rangle\langle v|u\rangle = \langle u|v\rangle(\langle u|v\rangle)^* = |\langle u|v\rangle|^2 \,. \tag{0.145}$$

This is necessarily a real number, and more than that, it must be zero or larger. Consequently, all projectors are what we call *positive semidefinite*. A linear operator is "p.s.d." if it satisfies $\langle u|L|u\rangle \geq 0$ for all vectors $|u\rangle$.

Let $\{|v_i\rangle : i = 1, \ldots, n\}$ be an orthonormal basis. Then

$$\sum_{i=1}^{n} |v_i\rangle\langle v_i| = I \,. \tag{0.146}$$

This is called "providing a *resolution of the identity*."

0.9 Tensor Products and Partial Traces

The inner product of two vectors yields a scalar. We can also define a product that takes in two vectors and produces another vector. One particularly useful way of doing so is the *tensor* or *Kronecker* product. Basically, we use one vector as a template for stamping out copies of the other. For example, start with the two vectors

$$\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix}, \ \vec{u} = \begin{pmatrix} c \\ d \end{pmatrix} \,. \tag{0.147}$$

Then the tensor product of these two vectors is

$$\vec{v} \otimes \vec{u} = \begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix} \,. \tag{0.148}$$

We can also take the tensor product of matrices. For example, let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \ B = \begin{pmatrix} e & f \\ g & h \end{pmatrix} \,. \tag{0.149}$$

The tensor product of these two matrices is

$$A \otimes B = \begin{pmatrix} ae & af & be & bf \\ ag & ah & bg & bh \\ ce & cf & de & df \\ cg & ch & dg & dh \end{pmatrix} \,. \tag{0.150}$$

So to speak, we slot copies of the second matrix into the shape made by the first.

The tensor product is noncommutative: Generally, $A \otimes B$ will not equal $B \otimes A$. However, it is associative:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C). \tag{0.151}$$

Moreover, the tensor product has a nice interplay with ordinary matrix multiplication:

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \tag{0.152}$$

The conceptual inverse of taking tensor products is taking *partial traces*. These are written by the trace operator tr with an extra subscript saying what part is being "traced over". We can define

$$\mathrm{tr}_2(A \otimes B) = A\mathrm{tr}B. \tag{0.153}$$

By requiring the action of tr_2 to be *linear* on its input, this actually establishes what tr_2 does on *all* matrices, even those that can't be written as a tensor product. To be more explicit about this, let's work out what the matrix $A\mathrm{tr}B$ is, in terms of the entries of A and B:

$$A\mathrm{tr}B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} (e + h) = \begin{pmatrix} ae + ah & be + bh \\ ce + ch & de + dh \end{pmatrix}. \tag{0.154}$$

Each term in these sums is an element of the matrix $A \otimes B$:

$$A\mathrm{tr}B = \begin{pmatrix} [A \otimes B]_{00} + [A \otimes B]_{11} & [A \otimes B]_{02} + [A \otimes B]_{13} \\ [A \otimes B]_{20} + [A \otimes B]_{31} & [A \otimes B]_{22} + [A \otimes B]_{33} \end{pmatrix}. \tag{0.155}$$

But we can slot any 4×4 matrix into this formula, whether or not it came from a tensor product! If M is an arbitrary 4×4 matrix, then

$$\mathrm{tr}_2 M = \begin{pmatrix} [M]_{00} + [M]_{11} & [M]_{02} + [M]_{13} \\ [M]_{20} + [M]_{31} & [M]_{22} + [M]_{33} \end{pmatrix}. \tag{0.156}$$

If we take the trace of $\mathrm{tr}_2 M$, we get the trace of M, which makes sense. We can define $\mathrm{tr}_1 M$ analogously, by working out and then extending the validity of $(\mathrm{tr}A)B$.

0.10 Differential Equations

Differential equations are expressions that characterize a function in terms of its derivatives. Finding a function that satisfies a given differential equation is known as "integrating" that differential equation. Sometimes, we can do this on sight. For example, if we are told that

$$\frac{dy}{dx} = m, \tag{0.157}$$

then we know that

$$y = mx + b, \tag{0.158}$$

because any function of the latter form will satisfy the former. As is almost ubiquitous in the land of differential equations, an extra constant has appeared, because differentiating any constant b by x yields zero. Asking that the slope of y is m says nothing about its intercept.

If we are told that

$$\frac{dy}{dx} = y, \tag{0.159}$$

then we can fumble our way through our memories of calculus and realize eventually that any function of exponential form will do:

$$y = be^x. \tag{0.160}$$

Here, the undetermined constant is a multiplicative factor. Very often in physics, we meet differential equations very much like this, only with an extra constant already included:

$$\frac{dy}{dx} = my. \tag{0.161}$$

Integrating this differential equation merely requires combining the previous solution with the chain rule:

$$y = be^{mx}. \tag{0.162}$$

This describes exponential growth if m is positive and exponential decay if m is negative. Often, the input variable is *time,* and we deduce the value of b from the value of y at time zero.

Another type of differential equation that is frequently encountered in physics involves a second derivative with respect to time. If we imagine a particle of mass M moving in one dimension while subject to an force, Newton's second law tells us that

$$M\frac{d^2x}{dt^2} = F(x) = -\frac{dV}{dx}, \tag{0.163}$$

where we have written the force as minus the derivative of the potential energy. A position x_0 is a point of *stable equilibrium* if the force exactly at x_0 vanishes, and if small steps away from x_0 in either direction cause the force to push back towards x_0. This is equivalent to saying that x_0 must be a *minimum* of the potential energy curve $V(x)$. Time and time again in physics, we study a system by finding its points of stable equilibrium and then seeing what happens when we push it slightly from those points. In the bulk of cases, this means approximating the potential energy $V(x)$ near x_0 by a Taylor series, i.e., drawing the parabola that best mimics $V(x)$ when x is close to x_0:

$$V(x) \approx V(x_0) + \frac{1}{2}\left.\frac{d^2V}{dx^2}\right|_{x_0}(x - x_0)^2. \tag{0.164}$$

The second derivative of V, which is also minus the first derivative of the force, measures the steepness of the potential well. We can call it K for brevity. Because x_0 is just a constant value,

$$M\frac{d^2}{dt^2}(x - x_0) = M\frac{d^2x}{dt^2} = -K(x - x_0). \tag{0.165}$$

Introducing the shifted variable $y = x - x_0$,

$$\frac{d^2y}{dt^2} = -\frac{K}{M}y. \tag{0.166}$$

Here, we are being told that the function $y(t)$ reproduces itself, up to a negative constant factor, when differentiated twice. This brings us into the realm of trigonometric functions (§0.2). Recalling that the second derivative of the sine is minus the sine, and likewise for the cosine, we discover

$$y = A\sin(\omega t) + B\cos(\omega t). \tag{0.167}$$

where we have defined $\omega := \sqrt{K/M}$. If we know the position y at time $t = 0$, then we can find B, and if we know the velocity at time zero, then we know the quantity A. We can also use the addition formulae from §0.2 to write this as

$$y = C\sin(\omega t + \phi). \tag{0.168}$$

These have all been examples of *ordinary* differential equations, because the unknown functions have each been functions of a single variable, and we have only taken ordinary derivatives with respect to that variable. A *partial* differential equation involves partial derivatives. Perhaps the first topic where a physics student encounters these, in the ordinary way of the curriculum, is in the topic of *wave motion.* Waves are disturbances that over time progress through space, and to a good first approximation they often maintain their shape while doing so. This leads us to consider functions of position and time that "slide along" without changing shape. How do we write a function with this property? Let us say that if we take a snapshot of a wave at a particular time, its contours follow a function $f(u)$. The middle of the wave pattern is at $u = 0$. As time flows on, the only thing that really changes is *where the center point of the pattern has moved.* If it started at position 0, then after t seconds, it will be at position vt. So, if we want to know what the wave is like at position x, then we just need to compare that value of x with the center position vt, because only the distance to the center point matters. So, we describe a wave pattern by a function $f(u)$, and to see what the wave is doing at position x and time t, we evaluate $f(x - vt)$.

If we differentiate $f(u)$ with respect to x while holding t constant, then the function f changes value only through how changing x changes u. But

$$\frac{\partial u}{\partial x} = \frac{\partial(x - vt)}{\partial x} = 1. \tag{0.169}$$

Likewise, if we hold x fixed and vary the time t, then $f(u)$ will change based on how much changing t changes u:

$$\frac{\partial u}{\partial t} = -v. \tag{0.170}$$

Consequently, the *second* derivative of f with respect to position is related to its second derivative with respect to time:

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}. \tag{0.171}$$

To check the units here, note that v is a measure of length per time, so $1/v^2$ has units of time squared per length squared. Differentiating twice by time brings in units of inverse time squared, so both sides have units of inverse length squared (times whatever the units of f might be).

Eq. (0.171) is the *wave equation*. Much of the theory of differential equations in undergraduate physics is the care and feeding of the wave equation, along with the equation for exponential change,

$$\frac{df}{dt} = \lambda f, \tag{0.172}$$

and the *simple harmonic oscillator* equation

$$\frac{d^2 f}{dt^2} = -\omega^2 f. \tag{0.173}$$

We will encounter others as we go.

What if we make the growth (or decay) rate of the exponential dependent upon t? Can we integrate the differential equation

$$\frac{df}{dt} = \lambda(t) f \ ? \tag{0.174}$$

It seems like the solution here should not be too dissimilar to what we had before. We ought to try an exponential again, but it needs to give us a $\lambda(t)$ when we differentiate. Thinking back to our calculus homework, we recall that differentiating e^u with respect to x pulls down, not u itself, but its derivative du/dx. So, we let $\Lambda(t)$ be an integral of $\lambda(t)$:

$$\Lambda(t) = \int^t dt' \, \lambda(t). \tag{0.175}$$

The derivative of this at t is just $\lambda(t)$, by the fundamental theorem of calculus. Now, we evaluate

$$\frac{d}{dt} e^{\Lambda(t)} = \lambda(t) e^{\Lambda(t)}. \tag{0.176}$$

So, any function of the form

$$f(t) = C \exp\left(\int^t dt' \, \lambda(t) \right) \tag{0.177}$$

will solve our differential equation. If $\lambda(t)$ is constant, we simply recover the special case that we had before. Note that we can pick any antiderivative of $\lambda(t)$: A constant shift can be absorbed into the choice of constant factor C out in front of the exponential.

It may so happen that our system of interest is being affected by an external influence whilst it is also doing its own thing. For example, if we charge up a capacitor and connect the two sides of it through a resistor, the voltage across the capacitor will decay exponentially. The process of periodically re-charging the capacitor would require an additional term in our differential equation, representing the extra input as a function of time.

So, let us consider differential equations of the form

$$\frac{df}{dt} - \lambda(t)f = \mu(t). \tag{0.178}$$

The left-hand side represents the system's own internal dynamics, and the right-hand side captures the outside influence. Let $\Lambda(t)$ be an antiderivative of $\lambda(t)$, as before, and consider the derivative of the product $e^{-\Lambda(t)}f$.

$$\begin{aligned}\frac{d}{dt}(e^{-\Lambda(t)}f) &= -\lambda(t)e^{-\Lambda(t)}f + e^{-\Lambda(t)}\frac{df}{dt} \\ &= e^{-\Lambda(t)}\left(\frac{df}{dt} - \lambda(t)f\right) \\ &= e^{-\Lambda(t)}\mu(t).\end{aligned} \tag{0.179}$$

We can integrate both sides:

$$e^{-\Lambda(t)}f = \int^t dt'\, e^{-\Lambda(t')}\mu(t'). \tag{0.180}$$

With one final turn of the algebra crank, we obtain

$$f = f_0 e^{\Lambda(t)} + e^{\Lambda(t)}\int_{t_0}^t dt'\, e^{-\Lambda(t')}\mu(t'). \tag{0.181}$$

As usual, we have some undetermined constants, which we have here written f_0 and t_0. A special case worth understanding is if $\lambda(t)$ and $\mu(t)$ are both constant. Then $\Lambda(t)$ can be taken to be λt, and

$$f = f_0 e^{\lambda t} + e^{\lambda t}\int_{t_0}^t dt'\, \mu e^{-\lambda t'}. \tag{0.182}$$

Pulling the constant μ out of the integral and evaluating, we get

$$f = f_0 e^{\lambda t} + \left(-\frac{\mu}{\lambda}\right) e^{\lambda t}\, e^{-\lambda t}\Big|_{t_0}^t = f_0 e^{\lambda t} - \frac{\mu}{\lambda}(1 - e^{\lambda(t-t_0)}). \tag{0.183}$$

If λ is negative, then the first term is an exponential decay, and the second term in parentheses will be so as well. After an interval of time has passed, then f will look like $-\mu/\lambda$. This is an example of a differential equation displaying *transient* behavior that dies out over time, leaving the long-term behavior governed by the external influence.

There is a certain "rabbit out of a hat" aspect to how we solved Eq. (0.178). Where exactly did the notion come from to take the derivative of that particular product of functions? It works, but without a way to see how one might guess it, the approach feels unsatisfying. This is unhappily emblematic of many lessons in differential equations: The motivation for a technique is not clear, and one is left wondering whether each new method being taught is a trick that only works for the one problem where it was introduced.

The mathematician Gian-Carlo Rota wrote a textbook about differential equations. Later [1], he reflected on the experience:

One of several unpleasant consequences of writing such a textbook is my being called upon to teach the sophomore differential equations course at MIT. This course is justly viewed as the most unpleasant undergraduate course in mathematics, by both teachers and students. Some of my colleagues have publicly announced that they would rather resign from MIT than lecture in sophomore differential equations. No such threat is available to me, since I am incorrectly labeled as the one member of the department who is supposed to have some expertise in the subject, guilty of writing an elementary textbook still in print.

Bibliography

[1] G.-C. Rota, "Ten lessons I wish I had learned before I started teaching differential equations," MAA invited address, Simmons College, 1997.