1 Introduction:

This set of notes reviews the basic mathematical techniques that are necessary to understand the material in an upper level undergraduate physical chemistry course. The notes are not intended to be a complete course in mathematical methods, but rather they are designed to give you the essential tools that will be required to solve the assigned problems. Physical chemistry is a fascinating subject since it deals with a fundamental understanding of the world around us, but it does require a modest level of skill in using mathematics. Mathematics is the language that people use to describe and interpret the **physical** world, so just as it would be necessary to learn a new language to function in another country, one must learn the language of the physical sciences to be able to communicate an understanding of physics and chemistry. So consider this a Berlitz Guide to your first physical chemistry course, and use it to get functioning with the at least partially new language. For some it will be just a review.

2 Calculus:

The two essential operations in calculus are differentiation and integration. We will first consider functions of one variable such as $y = f(x)$. We will assume the functions are **well behaved** meaning the derivatives exist and the functions are single valued for a given value of $x$.

2.1 Derivatives:

The derivative of a function is the slope of a function at a point. It is the limit of the ratio:

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.$$  \hspace{1cm} (1)

Some common derivatives we will need are:
\[
\frac{de^x}{dx} = e^x \tag{2}
\]
\[
\frac{de^{ax}}{dx} = ae^x \tag{3}
\]
\[
\frac{d \cos x}{dx} = -\sin x \tag{4}
\]
\[
\frac{d \sin x}{dx} = \cos x \tag{5}
\]
\[
\frac{dx^n}{dx} = nx^{n-1} \tag{6}
\]
\[
\frac{d \log x}{dx} = \frac{1}{x} \tag{7}
\]

The chain rule says ‘take the derivative of the outside first and then work inward…’. Mathematically that means
\[
\frac{df(g(x))}{dx} = \frac{df}{dg} \frac{dg}{dx}. \tag{8}
\]
Equation 3 above is an example of this rule. See the CRC Handbook for lots of derivatives and integrals, etc.

2.2 Partial derivatives:

Say we have a function of several variables: \( f(x, y, z) = 0 \). It may seem strange to make it equal to zero, but let’s say there was a constant on the rhs, then we could always just subtract it from both sides to get zero on the right and a slightly more complicated expression on the left. The concept of partial derivative comes from taking the derivative with respect to one of the variables while holding the others constant. We can write a small change in \( f(x, y, z) \), i.e. \( df \) as:
\[
df = \left( \frac{\partial f}{\partial x} \right)_{y,z} dx + \left( \frac{\partial f}{\partial y} \right)_{x,z} dy + \left( \frac{\partial f}{\partial z} \right)_{x,y} dz \tag{9}
\]
where the subscripted variables mean they are both held constant while the other one is varied.

For example, look at the equation for the surface of a sphere:
\[
x^2 + y^2 + z^2 = r^2 \tag{10}
\]
Then:

$$\left( \frac{\partial f}{\partial x} \right)_{y,z} = 2x.$$  \hspace{1cm} (11)

Or for the function \( f(x, y, z) = x^2 y^3 z^4 \) we have:

$$\left( \frac{\partial f}{\partial z} \right)_{x,y} = 4x^2 y^3 z^3.$$  \hspace{1cm} (12)

If we want to consider second derivatives, an example is (from the previous function):

$$\left( \frac{\partial^2 f}{\partial z^2} \right)_{x,y} = 12x^2 y^3 z^2.$$  \hspace{1cm} (13)

For a mixed derivative, we first take the derivative with respect to one variable, holding the other two constant, and then do the same for the next variable.

$$\left( \frac{\partial^2 f}{\partial x \partial y} \right) = 6xy^2 z^4.$$  \hspace{1cm} (14)

The variables that are held constant are omitted since they change for the two variables. For the cases that we will consider, the order of the differentiation does not matter.

So it’s pretty simple.

### 2.3 Integral:

The integral is the limit of the sum of small areas under a curve:

$$I = \lim_{h \to 0} \sum_{i}^{N} f(x_i)h$$  \hspace{1cm} (15)

where \( f(x_i) \) is the height and \( h \) is the width of the small area centered at the point \( x_i \). In the limit of very small \( h \), this equation goes over to the integral:

$$I = \int f(x)dx$$  \hspace{1cm} (16)
The integral above is an indefinite integral since there are no limits. The definite integral has limits, say \( a \) and \( b \):

\[
I = \int_{a}^{b} f(x) \, dx \quad (17)
\]

Integration is the ‘inverse operation’ of differentiation, so if you know a derivative you can work backwards to get an integral. A couple examples of indefinite integrals are:

\[
I = \int e^{x} \, dx = e^{x} + c \quad (18)
\]
\[
I = \int \cos x \, dx = \sin x + c \quad (19)
\]

If we change an indefinite integral to a definite integral then we take the difference of the function on the rhs at the two limits of integration, and since the constant is the same on both sides, it drops out:

\[
I = \int_{a}^{b} \cos x \, dx = \sin x \big|_{a}^{b} = \sin b - \sin a. \quad (20)
\]

An important method for solving integrals is integration by parts. This comes from the identity:

\[
d(uv) = udv + vdu \quad (21)
\]

If we move one term to the other side and then integrate, we get:

\[
\int udv = uv - \int vdu. \quad (22)
\]

### 2.4 Multiple integrals:

The concept of integral can be extended to multiple dimensions. In fact some integrals in Quantum Mechanics are formally infinite dimensional! Don’t worry, we won’t have to do those by hand. In words, you do one integral first, holding the other variable constant, then do the second and so on. You write a two dimensional integral like this:
\[ I = \int_a^b \int_c^d f(x,y) \, dx \, dy. \]  

(23)

If the function \( f(x,y) \) happens to be of the form \( f(x,y) = g(x)h(y) \) (separable) then the integral breaks down into a product of two one dimensional integrals:

\[ \int_a^b \int_c^d f(x,y) \, dx \, dy = \int_a^b g(x) \, dx \int_c^d h(y) \, dy. \]  

(24)

If you can make the integral separable, that’s the easiest way to solve it, but most often that’s not so easy.

2.5 Integration change of variables:

Sometimes we need to change the variables for an integral, say from \( x \) and \( y \) to \( r \) and \( \theta \), polar coordinates. Remember \( x = r \cos \theta \) and \( y = r \sin \theta \). The coordinate ranges are: \( 0 \leq r \leq \infty; 0 \leq \theta \leq 2\pi \). We need to express the function \( f(x,y) \) in terms of \( r \) and \( \theta \) and then change the ‘volume element’ (here area element) from \( dx \, dy \) to \( rdrd\theta \), where \( J \) is called the Jacobian of the coordinate transformation. The Jacobian is the determinant given below:

\[ J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \]  

(25)

The 2-d determinant is given by:

\[ \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \]  

(26)

So if you do the four partial derivatives and plug them in, you get

\[ J = r(\cos^2 \theta + \sin^2 \theta) \]  

(27)

But we know that \( \cos^2 \theta + \sin^2 \theta = 1 \) so \( J = r \) for the two dimensional change of variables. Then the ‘volume element’ (here area element) change is \( dx \, dy \rightarrow rdrd\theta \). Note the units stay the same (as they should), i.e. area for both. Remember angles don’t have units. For the 3d case, the variable change is: \( x = r \sin \theta \cos \phi; y = r \sin \theta \sin \phi; z = r \cos \theta \). The coordinate ranges are: \( 0 \leq r \leq \infty; 0 \leq \theta \leq \pi; 0 \leq \phi \leq 2\pi \). The determinant is a little more complicated (you can look up the 3d determinant in a calculus text), and for the change of volume elements we get \( dx \, dy \, dz \rightarrow r^2 dr \sin \theta d\theta d\phi \).
2.6 Complex numbers:

In Quantum Mechanics, we need to use a generalization of real numbers called complex numbers. These numbers originate from solving quadratic or other algebraic equations where we end up needing to take the square root of a negative number. This sounds meaningless but it turns out to be very important in Quantum Mechanics. So we define a new number $i = \sqrt{-1}$. Then, in general a complex number can be written as $z = x + iy$, where $x$ and $y$ are both the usual real numbers. The ‘complex conjugate’ of $z$ is given by $z^* = x - iy$. Some useful relations are given below:

$$e^{ix} = \cos x + i \sin x$$ (28)

$$e^{-ix} = \cos x - i \sin x$$ (29)

You can prove these using the Taylor’s expansions for $e^{ix}$, $\cos$, and $\sin$ (below). Then we can represent a complex number as:

$$z = re^{i\theta} = r(\cos \theta + i \sin \theta)$$ (30)

Also

$$|z|^2 = z^*z = (x - iy)(x + iy) = x^2 - (i^2)y^2 = x^2 + y^2.$$ (31)

2.7 Taylor series expansion:

We can approximate a function near a certain point if we know its value at that point and the values of its derivatives there. The expansion is called a Taylor expansion of the function about the point $x_0$.

$$f(x_0 + h) = f(x_0) + \left(\frac{df(x)}{dx}\right)_{x_0} h + \frac{1}{2} \left(\frac{d^2f(x)}{dx^2}\right)_{x_0} h^2 + \ldots$$ (32)

This can be rewritten in a shorter form as:

$$f(x_0 + h) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{f^{(n)}(x)}{x_0}\right) h^n.$$ (33)

Here the term $\left(\frac{f^{(n)}(x)}{x_0}\right)$ means take the $n$th derivative of $f$ and then evaluate it at the point $x_0$. The quantity $n!$ (factorial) means take the product of $n$ times
all the succeeding numbers below $n$. For example, $3! = 3 \cdot 2 \cdot 1$. The Taylor expansion is important when we want to approximate functions.

## 3 Differential Equations

Many of the equations of classical mechanics and quantum mechanics are written as differential equations. What does this mean? It means that we have an equation which involves derivatives of a function and we need to solve for that function. A second requirement is that we need the boundary conditions, which means we need the value of the function and maybe its first derivative at some points or points in time or space. We'll see examples as we go along.

### 3.1 Ordinary differential equations.

An ordinary differential equation is one where there are only derivatives with respect to one variable. A good example is one of Newton’s laws, that is $F = ma$, in one dimension. This equation says that the force is the mass times the acceleration. Now the force is given by the derivative of a potential energy,

$$F = \frac{\partial V(x)}{\partial x},$$

where $V(x)$ is the potential energy, and the acceleration is the second derivative of the coordinate with respect to time:

$$a = \frac{d^2 x}{dt^2}.$$

Since typically we know the functional form of the potential $V(x)$, we can take that derivative and get:

$$m \frac{d^2 x}{dt^2} = F(x).$$

A very important system in both classical and quantum mechanics is the harmonic oscillator. Classically, it is a model for the oscillations of a collection of springs, and quantum mechanically it is a model for the vibrations of molecules. The harmonic oscillator potential energy is given by
The variable $\omega$ is the frequency of the oscillator. Then using the formula given above, the force is

$$F(x) = -m\omega^2 x$$  \hspace{1cm} (38)$$

So the force acts 'in opposition' to the particle moving up either side of the quadratic potential energy surface. It is called a restoring force.

Then the differential equation to be solved ($F = ma$) is:

$$m\frac{d^2x}{dt^2} = -m\omega^2 x$$  \hspace{1cm} (39)$$

which is the same as

$$\frac{d^2x}{dt^2} + \omega^2 x = 0.$$  \hspace{1cm} (40)$$

This is a second order (second derivative is the highest one) ordinary differential equation (and homogeneous since it has zero of the rhs; if it is inhomogeneous it will have some function of $t$ there). It is linear since only the first power of $x$ appears. Solving this equation means finding the coordinate $x$ as a function of $t$. We also need to know the initial value of $x$, that is $x(0)$ and the initial velocity which is the first derivative of $x$ wrt $t$. We call the initial velocity $v(0)$. Two initial conditions ($x(0)$ and $v(0)$) are needed since it is a second order differential equation. If third derivatives appeared we would need three and so on.

There are many tricks for solving differential equations. I will give you a few of these as we go along. For our case, we postulate a solution of the form $x(t) = e^{\lambda t}$, and plug this into the differential equation. What do we get?

$$\lambda^2 e^{\lambda t} + \omega^2 e^{\lambda t} = 0$$  \hspace{1cm} (41)$$

and we can cancel the exponentials so that it's just

$$\lambda^2 + \omega^2 = 0.$$  \hspace{1cm} (42)$$

The solution (for $\lambda$) is simply $\lambda = \pm i\omega$. Now the solution of the differential equation is the linear combination of the two possible solutions $e^{\lambda_1 t}$ and $e^{\lambda_2 t}$ with each multiplied by a constant:
\[ x(t) = Ae^{i\omega t} + Be^{-i\omega t}. \]  

Using the formula for \( e^{ix} \), you can show that:

\[ x(t) = (A + B) \cos \omega t + i(A - B) \sin \omega t. \]  

Let’s call the two new prefactors to the sin and cos \( A' \) and \( B' \). Then we have:

\[ x(t) = A' \cos \omega t + B' \sin \omega t. \]

This is just a superposition of cos and sin waves.

In classical mechanics, we need to consider the initial position and velocity in order to determine the particle’s future path. These are the initial conditions. We could also specify the initial location and final location; these would be boundary conditions. That is, when the baseball is hit, we need to know where it started, and how fast and in what direction it is going in order to predict where it will land. How do we get these? Set \( t = 0 \); then \( x(0) = A' \) by looking at the cos and sin of zero. So we have one initial condition. What about the velocity? Well, the velocity is the first derivative of the coordinate wrt time. So, its formula is:

\[ x'(t) = v(t) = -A' \omega \sin \omega t + B' \omega \cos \omega t. \]

The symbol \( x'(t) \) means the first derivative of \( x \) wrt time. Then we set the time equal to zero and see that \( v(0) = B' \omega \) so in the end:

\[ x(t) = x(0) \cos \omega t + \frac{v(0)}{\omega} \sin \omega t. \]

This is the general solution for the one dimensional classical harmonic oscillator. Notice that the units are right, that is length. The velocity has units of length/time and the frequency has units of 1/time so the net of the second term is length as it should be.

In quantum mechanics we will be more interested in boundary conditions than initial conditions. We will spend most of our time solving the time independent Schrödinger equation. In one dimension, this equation is an ordinary differential equation and looks like this:

\[ -\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x). \]
The variable $E$ is just a number, that is not a function of $x$, and it is called the ‘eigenvalue.’ The function $\psi(x)$ is called the wavefunction. The function $V(x)$ is the potential energy just like in classical mechanics. Typically we specify the value of $\psi(x)$ on the boundary which could be the value as $x \to \pm \infty$. Usually $\psi \to 0$ as $x \to \pm \infty$. Then it turns out only certain values of the number $E$ will be allowed, and that’s what gives rise to the ‘quantum levels’ that you’ve all heard about (‘quantum leap’ and so on). Those energy levels are what leads to peaks in the spectra that one takes in the laboratory, like IR, UV, etc. We’ll hear a lot about the Schrödinger equation as time goes on.

3.2 Partial Differential Equations

Partial differential equations are a generalization of ordinary differential equations, where you have derivatives with respect to more than one variable. For us, this will be derivatives wrt more than one spacial variable (like $x$, $y$, and $z$) or derivatives with respect to coordinate and time in the same equation (time dependent Schrödinger equation). As an example, I show the 3d time independent Schrödinger equation:

$$-\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi(x,y,z) + V(x,y,z) \psi(x,y,z) = E \psi(x,y,z). \quad (49)$$

This equation has (partial) derivatives with respect to three different variables. We will solve equations like this but will often change to a different coordinate system like spherical polar coordinates, where the problem becomes separable.

Notice that I have lumped the three partial second derivatives into one collection before letting those derivatives ‘act on’ the function $\psi(x,y,z)$. This introduces the concept of operator. Those derivatives can be considered mathematical entities by themselves, although they really only become meaningful when they act on the wavefunction. The potential $V(x,y,z)$ is also an operator, as is the constant $E$. They are simpler than the differential operators. Think of an operator as a ‘pac-man’ like creature that goes along and operates on the wavefunction $\psi$, giving some new function when it is done. The whole operator on the lhs above can be written as:

$$H = -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(x,y,z), \quad (50)$$

where $H$ is called the ‘Hamiltonian operator.’ This is the operator that corresponds to the energy in quantum mechanics. So now we can simply rewrite the Schrödinger equation as:
\[ H\psi(x, y, z) = E\psi(x, y, z). \] (51)

The way to read this in words is: ‘the Hamiltonian operator acts on the wavefunction and gives back a number \(E\) times the wavefunction’. That number \(E\) that it gives back is the energy eigenvalue. This is just to give you some introduction to the language we’ll use as we get into the quantum mechanics.

The time dependent Schrödinger equation is more general and is written as:

\[ i\hbar \frac{\partial \psi}{\partial t} = H\psi. \] (52)

This is the most complicated partial differential equation we’ll see since it has derivatives wrt time and space. This equation can in principle describe all of chemistry.

### 3.3 Eigenvalue Equations

The definition of an eigenvalue equation is as follows. Let some operator act on a function. Then, if that operation gives you back some constant times the function itself, it is an eigenvalue equation. The operators can be the acts of differentiation, multiplication by a function, or even multiplication by a constant.

Here are a couple simple examples:

\[ \frac{de^{ax}}{dx} = ae^{ax}. \] (53)

The operator is \(\frac{d}{dx}\), the eigenfunction is \(e^{ax}\) and the eigenvalue is \(a\). A case where it doesn’t work is:

\[ \frac{d \log x}{dx} = \frac{1}{x} \neq \text{constant} \cdot \log x. \] (54)

We don’t get the same function back when we take the derivative. You should practise by taking some of the basic functions of mathematics (cos, sin, exponential, etc.) and then taking first and second derivatives to see if they are eigenfunctions of those operators.

### 4 Probability Concepts

Notions of probability are important in all areas of physics and chemistry (even in biology and economics, etc.). This is partly due to the fact that we most often
deal with so many particles or other objects that a complete description of the motions of all of them is hopeless. Therefore, we need to average over many events and compare that result to experiment. More fundamentally, the theory of quantum mechanics tells us that we can only make probabilistic predictions about the future evolution of physical systems if there are alternate possible events that can occur. (This is even true in classical mechanics if the system dynamics is at all chaotic, which is usually the case in nature). So we need some grounding in the most basic aspects of probability theory. We will see this brings us to study distribution functions, averages and moments, joint probabilities, the binomial distribution, the Gaussian and Poisson distributions, and finally the central limit theorem. Most of this discussion is based on Chapter 5 of the text “A Modern Course in Statistical Physics,” by L. E. Reichl. Two other good references are: “Probability, Random Variables, and Stochastic Processes,” by A. Papoulis, and “An Introduction to Probability Theory and Its Applications,” by W. Feller. These are a bit advanced.

4.1 Bayes’s Theorem

A very basic formula of probability is this:

\[ P(B|A) = \frac{P(A \cap B)}{P(B)}. \]  

(55)

where \( P(B|A) \) is the probability of observing the event \( B \) given the fact that \( A \) has occurred, \( P(A \cap B) \) is the probability that both \( A \) and \( B \) occur, and \( P(B) \) is the probability of \( B \) in the first place. This is actually the foundation for a great deal of modern statistics.

4.2 Distribution functions

There can be both discrete or continuous types of events. For example, the probability of heads or tails is for discrete events, while the probability distribution for the distance of the thrown horseshoe from the pole is continuous. The probability distribution function is essentially the probability of a given event, then normalized so that the sum (or integral) over all events is 1. That is, something must happen with probability of 1.

Generally for discrete events we have:

\[ f(x_i) \geq 0 \]  

(56)

and
\[ \sum_i f(x_i) = 1. \quad (57) \]

The average of some quantity \( g(x_i) \) is then:

\[ \langle g \rangle = \sum_i g(x_i) f(x_i) \quad (58) \]

An important special case of this is the set of “moments” of the distribution:

\[ \langle x^n \rangle = \sum_i x_i^n f(x_i) \quad (59) \]

For the case of \( n = 2 \) (the second moment), a new quantity is often defined, which is the standard deviation (the square of the standard deviation is called the variance):

\[ \sigma = (\langle x^2 \rangle - \langle x \rangle^2)^{1/2}. \quad (60) \]

This quantity gives a measure of the spread of the distribution about the mean or average \( \langle x \rangle \). For a Gaussian distribution, it is essentially the width (half width at half height) of the distribution.

### 4.3 Continuous Distributions

For continuous distributions, a coarse way of thinking about the change is that the sums above go over to integrals. For example, the probability of a single value of the continuous distribution \( f(x) \) is strictly speaking zero, but the probability to observe events in a range of \( x \) is:

\[ P(a \leq x \leq b) = \int_a^b f(x) dx. \quad (61) \]

where

\[ f(x) \geq 0 \quad (62) \]

and

\[ \int f(x) dx = 1 \quad (63) \]
The average of a quantity \( g(x) \) is:

\[
\langle g \rangle = \int g(x)f(x)dx
\]

(64)

and the moments are:

\[
\langle x^n \rangle = \int x^n f(x)dx
\]

(65)

### 4.4 Joint probabilities

Consider a case where we have two independent variables \( x \) and \( y \). Then we need to examine a probability distribution as a function of those two variables: \( f(x,y) \). This is the probability density to observe both \( x \) and \( y \). If we integrate over one of the variables, then we are left with a function of the other only:

\[
f(x) = \int f(x,y)dy
\]

(66)

If we integrate over both, of course we get one:

\[
1 = \int \int f(x,y)dxdy.
\]

(67)

The covariance is defined as follows:

\[
cov(x,y) = \int \int dxdy(x - \langle x \rangle)(y - \langle y \rangle)f(x,y)
\]

(68)

\[
= \int \int dxdyxyf(x,y) - \langle x \rangle\langle y \rangle
\]

(69)

\[
= \langle xy \rangle - \langle x \rangle\langle y \rangle.
\]

(70)

If two variables are independent, then the covariance is zero.

### 4.5 The Binomial Distribution

It is interesting that there is a close connection between “number theory” and probability theory. This can be seen by looking at the rather simple example
of the binomial expansion. For large exponent, the coefficient distribution (discrete) goes over to either the Gaussian or Poisson distribution, depending on the probability of a given outcome. These are two of the most useful distributions in all of probability theory.

Say we perform $N$ independent trials for flipping a coin, which could be a weighted coin. The probability for a head in one toss is $p$ and the probability for tails is $q$. Obviously, $p + q = 1$. Say the total is $N = N_H + N_T$. Then the probability for any measurement of $N_H$ and $N_T$ is:

$$P_N(N_H) = \frac{N!}{N_H!N_T!} p^{N_H} q^{N_T} \quad (71)$$

The sum over all these $N$ probabilities is clearly one since

$$\sum_{N_H=0}^N P_N(N_H) = \sum_{N_H=0}^N \frac{N!}{N_H!(N-N_H)!} p^{N_H} q^{N-N_H} \quad (72)$$

$$= (p + q)^N = 1 \quad (73)$$

After some algebra, you can show that the first moment $\langle N_H \rangle$ of the binomial distribution is $pN$, and the second moment is $(Np)^2 + Npq$.

Then the variance is $Npq$ and the standard deviation is thus $\sqrt{Npq}$. The fractional deviation, i.e. the standard deviation divided by the average (or first moment) is:

$$\frac{\sigma}{\langle N_H \rangle} = \sqrt{\frac{q}{p}} \frac{1}{\sqrt{N}} \quad (74)$$

This is the first indication of the very important role of $\sqrt{N}$ in statistical arguments. The larger the sample, the lower the statistical deviation in relation to the mean or average value. However, notice that the fractional deviation only decreases slowly with $N$, that is as the square root. So sometimes we have to work very hard to get good statistics experimentally or in computations.

### 4.6 The Gaussian Distribution

Now how do we get the Gaussian or normal distribution from the binomial distribution? If you plot the binomial distribution for $p = q$ and for large $N$, it starts to take the shape of a Gaussian centered about the mean, which for this case is $N/2$. Can we make that statement a bit more quantitative? The
way to do this is to use an approximation (Stirling) developed for the factorial function:

\[ \ln N! \approx N \ln N - N \]  

If we take the logarithm of the probability for one term in the binomial expansion, then make the Stirling's approximation, then re-exponentiate the result, we finally end up with the result:

\[ P(N_H) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \frac{(N_H - \langle N_H \rangle)^2}{\sigma^2} \right] \]  

The underlying assumptions here are that \( p \) and \( q \) are close to the same and \( N \) is large. Actually mathematically the assumption is that \( Npq \gg 1 \). Two parameters are required to specify a Gaussian (and only two): the mean and the standard deviation, \( i.e. \) the first two moments.

### 4.7 The Poisson Distribution

Another important limit of the binomial distribution occurs when \( N \to \infty \) and \( p \to 0 \) such that \( Np = a \ll N \) where \( a \) is a constant. This situation might arise if we very heavily weight the coin so that the tails dominate the tosses. In this limit, we get for the probability of heads, the Poisson distribution:

\[ P(N_H) = \frac{a^{N_H} e^{-a}}{N_H!} \]  

Notice that here the distribution is determined by only one parameter \( a \) which is the first moment.

### 4.8 Brownian motion

There is a close connection between the Gaussian distribution and the random walk, or the ‘drunken sailor’ process, or Brownian motion. This phenomenon was first observed by a botanist named Brown who watched small particles moving in water (it turns out what he saw may have been due to vibrations of the apparatus and not due to real Brownian motion, but the name has stuck and the phenomenon is real). The theory was formally developed by Einstein and a great Polish scientist named Smoluchowski at about the same time. It led to the first truly convincing demonstration of the atomic theory of matter and a close prediction of Avogadro’s number.
Imagine a drunken sailor who is at some point on a line (for now we make him walk on a line which may be asking a lot). At any given point he is equally likely to move to the left or right, we can’t predict which. So if we ‘map’ this problem onto the previous sections where we derived the Gaussian distribution from the binomial expansion, we can say the sailor takes $N_r$ steps to the right and $N_l$ steps to the left, where $N_r + N_l = N$. Assume that the probabilities for left and right are the same, so $p = q = 1/2$. Introduce a new variable $m$ where $m = N_r - N_l$. Then $m = 2N_r - N$, and $N_r = (m + N)/2$. Substitute this in our previous expression for the Gaussian distribution, and we get:

\[ P_N(m) = \left[ \frac{2}{\pi N} \right]^{1/2} \exp \left[ -\frac{m^2}{2N} \right]. \quad (78) \]

This says the distribution of end points of the random walk is a Gaussian or normal distribution and the standard deviation of the distribution goes as $N^{1/2}$. The distribution is always centered about $m = 0$ so long as there is no ‘external force’ acting on the sailor. For example, if there was a steep incline he would tend to drift downhill.

Let the step size be $l$. Then we can show that in the limit of a very large number of steps, and letting the step size get small ($Nl^2$ goes to a limit which we will call $2Dt$), the distribution goes over to a continuous one as follows:

\[ P(x,t)dx = \left[ \frac{1}{4\pi Dt} \right]^{1/2} \exp \left[ -\frac{x^2}{4Dt} \right] dx. \quad (79) \]

where $t$ is the time from the beginning when the sailor was at the origin, $n$ is the number of steps per unit time, $N = nt$, and $D = \frac{1}{2}nl^2$. The quantity $D$ is called the diffusion constant, and is typically expressed in units of cm$^2$/sec. A common value for a molecule in a liquid is roughly $10^{-5}$ cm$^2$/sec. An interesting point about random walks is that, if $D = \frac{1}{2}nl^2$ goes to a limit as $l \to 0$, then $nl$ goes to infinity. This suggests that the ‘velocity’ of the particle is infinity, which is a peculiarity of the model based on the assumption that the random collisions instantaneously alter the velocities of the larger Brownian particle. This assumption is an extreme limit – it yields the correct results for long times but is of course strictly speaking incorrect.

When we pass to the continuum limit to get the formula for the probability above, it turns out this function is a solution to a very important (partial) differential equation, the diffusion equation:

\[ \frac{\partial P(x,t)}{\partial t} = D \frac{\partial^2 P(x,t)}{\partial x^2} \quad (80) \]
This equation describes the random motions of molecules in liquids, and if you look closely, you see it actually is close in appearance to the time dependent Schrödinger equation. You should try to show that the derived probability is indeed a solution to this PDE.

4.9 The Central Limit Theorem

A very deep theorem in probability theory is called the central limit theorem. We won’t prove it here (a good proof is given in Reichl), but will just state the result so that you get the idea of what it is and why it is so important. This theorem underlies the pervasive importance of the Gaussian distribution in probability, physics, chemistry, biology, and economics.

Let’s say we make $N$ measurements of some quantity $x$ which has a probability density $f(x)$. Then we ask: what is the distribution of the quantity $y$, which is simply the average value of all the $N$ measurements of $x$? We need to make three assumptions here 1) the moments of the distribution $f(x)$ are defined 2) the measurements are ‘statistically independent’ (often hard to really prove) 3) there are a large number of measurements (this is often say 30 or more). Then it can be proven mathematically that the distribution of $y$ is Gaussian, and the standard deviation of $y$ is the standard deviation of $x$ (i.e. the ‘natural’ fluctuations in $x$), divided by $\sqrt{N}$. This is how experimentalists get the error bars on their data — you take a large number of statistically independent samples, calculate the standard deviation in those numbers, then divide by $\sqrt{N}$. This gives a good estimate of the possible errors in the experiment. This theorem is also used when people perform Monte Carlo simulations of some process. The assumption of statistical independence is always lurking behind doing this so one must be very careful to be sure that condition is met to a high degree of accuracy to trust the outcome.

5 Matrix Operations

Linear algebra plays a major part in modern physics and chemistry. It is often true that a differential equation problem (such as the Schrödinger equation) can be cast as an equivalent matrix problem. This is what is done for example in quantum chemistry where the differential equations are rewritten as matrix eigenvalue problems by introducing basis functions. Maybe the simplest way to see the connection is to simply take the Schrödinger differential equation and write it in discrete form using finite differences for the second derivative operator. We won’t get into this approach here, but mention that the topic is discussed even in the old quantum chemistry text by Pauling and Wilson. The correspondence between eigenvalue differential equations and matrix eigenvalue
problems is discussed in any modern text on quantum mechanics. We’ll just give some of the most basic linear algebra operations for matrices and vectors, and use the simplest 2×2 matrices.

5.1 Vectors and dot products

A vector is just a collection of numbers in one direction. You can have row vectors:

\[ \mathbf{a} = (a_1 a_2) \] (81)

or column vectors:

\[ \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \] (82)

The dot product of two vectors \( \mathbf{a} \cdot \mathbf{b} \) is simply

\[ \mathbf{a} \cdot \mathbf{b} = (a_1 a_2) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = a_1 b_1 + a_2 b_2. \] (83)

5.2 Matrices

A matrix is a two dimensional collection of numbers. The rows go across and the columns down:

\[ \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \] (84)

The product of two matrices \( \mathbf{AB} = \mathbf{C} \) goes as follows:

\[ C_{ik} = \sum_j A_{ij} B_{jk} \] (85)

5.3 Matrix eigenvalue equation

An eigenvalue equation in matrix form is:

\[ \mathbf{A} \mathbf{v} = \lambda \mathbf{v} \] (86)
where \( \mathbf{A} \) is a matrix, \( \mathbf{v} \) is a vector, and \( \lambda \) is a number. How do we solve this equation? First, we need the eigenvalues. For an \( n \times n \) matrix we will have \( n \) eigenvalues. Rearrange the equation like this:

\[
\mathbf{A}\mathbf{v} - \lambda \mathbf{v} = 0. \tag{87}
\]

This is a linear algebraic equation, and it has a solution only if the determinant (see above) is zero:

\[
|\mathbf{A} - \lambda \mathbf{I}| = 0. \tag{88}
\]

Here \( \mathbf{I} \) is the unit matrix which has all ones along the diagonal and zeros everywhere else. This equation can be rewritten as:

\[
\det \begin{bmatrix}
  a_{11} - \lambda & a_{12} \\
  a_{21} & a_{22} - \lambda
\end{bmatrix} = 0 \tag{89}
\]

This leads to an algebraic equation (of order \( n \)):

\[
(a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0 \tag{90}
\]

which for the 2x2 case is just a quadratic equation. You solve this for the two \( \lambda \)'s and you get the eigenvalues. Then you go back to the eigenvalue equation, plug in the eigenvalues one at a time, and solve the equations for the two eigenvectors. We’ll do an example to show this. Take the matrix:

\[
\mathbf{A} = \begin{bmatrix}
  0 & 1 \\
  1 & 0
\end{bmatrix} \tag{91}
\]

Then the determinental equation is:

\[
\det \begin{bmatrix}
  -\lambda & 1 \\
  1 & -\lambda
\end{bmatrix} = 0 \tag{92}
\]

or just \( \lambda^2 - 1 = 0 \). So the two eigenvalues are +1 and -1. If we plug in the eigenvalues to the original eigenvalue equation, we get, for the +1 case: \( v_2 = v_1 \) and \( v_1 = v_2 \). This gives us a lot of lee-way to choose the vector, so let’s say it is:

\[
\mathbf{v} = \begin{bmatrix}
  1 \\
  1
\end{bmatrix}. \tag{93}
\]
For the -1 eigenvalue, we get a vector such as:

\[ \mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \]  

(94)

Notice that the two vectors we got give a value of zero when we take the dot product of the two. This is an example of what is called orthogonality of eigenvectors. (It turns out this is also true for the eigenfunctions which are the solutions to the eigenvalue differential equations we talked about earlier). This is a very important mathematical property of the eigenvectors.

One last topic we should touch on is that the eigenvector should be normalized. If we take the dot product of an eigenvector with itself, the result should be one. This is true because the square of each element of the vector is somehow proportional to the probability for some event to occur. Thus, when we sum over all probabilities, we should get one for the total probability for something to happen. The dot product of each eigenvector we got above with itself is 2. Therefore, we should multiply each individual vector above by \(1/\sqrt{2}\) to obtain a normalized eigenvector.

6  Miscellaneous

6.1  Gaussian integral

A very important integral is the Gaussian integral, a definite integral. I’ll just give you the formula for it; you could look it up in the CRC Handbook.

\[ \int_{-\infty}^{\infty} e^{-ax^2+bx} \, dx = \sqrt{\frac{\pi}{a}} e^{b^2/4a}. \]  

(95)

6.2  Fourier Transform

We won’t use this mathematical technique much but I would like you to at least see it. It is very important in all forms of spectroscopy and now in medicine: CAT scans and MRI imaging would not be possible without it (there is a fancy method called the fast Fourier transform FFT which allows it to work quickly). What the Fourier transform does is to go from a variable in one space and re-display the same information in another space. An example is in pulsed NMR spectroscopy where a time dependent signal is measured and then Fourier transformed to give the frequency (or energy) dependent signal. Finally, it gives a
good pictorial representation of the uncertainty principle in position and momentum space. We’ll use that as our example.

Consider the two variables $k$ and $x$, where $k$ is like a momentum and $x$ is the coordinate. The Fourier transform from $x$ to $k$ space is defined as:

$$f(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$  \hspace{1cm} (96)

This formula can be thought of as adding up a lot of waves (from the $e^{ikx}$ term) multiplied by the function value. This tells us how important the component of the momentum $f(k)$ is in the signal $f(x)$. The reverse process is written as:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(k) e^{ikx} dk$$  \hspace{1cm} (97)

Both $f(x)$ and $f(k)$ have the same information content but sometimes redisplaying the information can make a world of difference (for example the NMR example mentioned above – the time dependent data is a mess, but the frequency spectrum shows sharp peaks like we are used to seeing in NMR of organic compounds).

### 6.3 The Uncertainty Principle

The uncertainty principle in quantum mechanics tells us that if we make many measurements of one variable and nail down its value to very high accuracy, then if we make the series of corresponding measurements of a complementary variable (like position and momentum) the spread in the measurements of the second variable will be large. The famous equation of Heisenberg is:

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$  \hspace{1cm} (98)

In words, if we repeatedly try to tie down a series of electrons in one spot, we will have no predictability about where they’re going. This concept led to a lot of debate between physicists and philosophers. The principle is now widely accepted as part of our understanding of nature.

We can use the formulas given above for the Gaussian integral and for the Fourier transform to analyze pictorially what the uncertainty principle is about. In quantum mechanics we often represent a particle (like an electron) as a Gaussian wavepacket:
This is the equation for the famous normal distribution that all of us professors use to assign grades. It is a localized function, and its half width is $\sigma$ at half the height at the origin.

Now as one of the exercises I will have you plug this function into the formula for the Fourier transform and then use the Gaussian integral formula to get $\psi(k)$. You will find that its width is proportional to $1/\sigma$. So the product of the widths in $x$ space and $k$ space is roughly one, or some constant not too different from one. Well, the variable $k$ is actually the momentum divided by $\hbar$, so we find crudely that the spread in $x$ times the spread in $p$ is close to $\hbar$, the uncertainty principle. You should plot these functions to get some idea what they look like. If $\psi(x)$ is wide, $\psi(k)$ is narrow, and vice versa.

7 Units

The units we will use in this course are SI units and Gaussian units, mainly SI units at first and Gaussian when we get into atomic structure. These topics are covered in depth in Levine’s text so I will not discuss them here. The point I want to emphasize (and that’s not stressed enough these days) is that when you derive formulas, you should always make sure the units work out right! There are two reasons for this: 1) the first is obvious, that you want to get the ‘right’ answer, but also 2) if you see that your units are not correct at the end, it can give you a hint at what you did wrong in the derivation. Every time you derive a formula, look to make sure the units make sense! If you’re trying to derive something that in the end gives length, your formula should reflect that. This will save you a lot of trouble. You should know the units for length, time, velocity, momentum, charge, and energy.