CHAPTER 1

BACKGROUND MATHEMATICS

1 Derivatives: Partial and Otherwise

1.1 A Necessary Formality: Limits

Before I can define and examine any element of calculus. I must first define limits. These are defined formally as:

Definition 1.1: Limits

For an function of x, $f : \mathbb{R} \to \mathbb{R}$, the limit of f as x approaches $c \in \mathbb{R}$ is $L \in \mathbb{R}$, which is denoted as:

$$\lim_{x \to c} \Big(f(x) \Big) = L, \tag{1.1}$$

if and only if:

$$\forall \epsilon \in \mathbb{R}, \exists \delta \in \mathbb{R} \text{ such that } \forall x \in (c - \delta, c + \delta), \big| f(x) - L \big| \leq \epsilon. \tag{1.2}$$

1.2 Initial Definition of Derivatives

We start from a formal definition of derivatives.

Definition 1.2: Derivative

Define the derivative of a function, f, at x is defined by:

$$\lim_{h \to 0} \left(\frac{f(x+h) - f(x)}{h} \right). \tag{1.3}$$

The derivative is not guaranteed to exist for all functions.

1.3 Some Example Derivatives

We can then consider some examples of using the definition of derivatives.
The Exponential Function
If $f(x) = e^x$,

$$(e^x)' = \lim_{h \to 0} \left(\frac{f(x+h) - f(x)}{h} \right) = \lim_{h \to 0} \left(\frac{e^{x+h} - e^x}{h} \right) = e^x \lim_{h \to 0} \left(\frac{e^h - 1}{h} \right) = e^x \lim_{h \to 0} \left(\frac{(e^h)'}{1} \right). \tag{1.4}$$

We can see this is satisfied if $(e^x)' = e^x$.
The Logarithm

If $f(x) = \ln(x)$, then:

$$(\ln(x))' = \lim_{h\to 0}\left(\frac{\ln(x+h)-\ln(x)}{h}\right) = \lim_{h\to 0}\left(\frac{\ln\left(\frac{x+h}{x}\right)}{h}\right) = \lim_{h\to 0}\left(\frac{1}{h}\ln\left(1+\frac{h}{x}\right)\right) \quad (1.5)$$

$$= \lim_{h\to 0}\left(\ln\left(\left[1+\frac{h}{x}\right]^{1/h}\right)\right). \quad (1.6)$$

Let $H = h/x$. Then,

$$(\ln(x))' = \lim_{H\to 0}\left(\ln\left(\left[1+H\right]^{(1/H)\cdot(1/x)}\right)\right) = \frac{1}{x}\lim_{H\to 0}\left(\ln\left(\left[1+H\right]^{1/H}\right)\right) = \frac{1}{x}. \quad (1.7)$$

1.4 Basic Theorems of Differential Calculus

Theorem 1.3: Product Rule

If u and v are once-differentiable functions of x, then

$$\big(u(x)v(x)\big)' = u'(x)v(x) + u(x)v'(x) \quad (1.8)$$

Proof for Theorem.

Using the definition of differentiation:

$$\big(u(x)v(x)\big)' = \lim_{h\to 0}\left(\frac{u(x+h)v(x+h)-u(x)v(x)}{h}\right) \quad (1.9)$$

$$= \lim_{h\to 0}\left(\frac{u(x+h)v(x+h)-u(x)v(x)+u(x+h)v(x)-u(x+h)v(x)}{h}\right) \quad (1.10)$$

$$= \lim_{h\to 0}\left(\frac{u(x+h)\big[v(x+h)-v(x)\big]}{h}\right) + \lim_{h\to 0}\left(\frac{\big[u(x+h)-u(x)\big]v(x)}{h}\right) \quad (1.11)$$

$$= \lim_{h\to 0}\Big(u(x+h)\Big)\lim_{h\to 0}\left(\frac{\big[v(x+h)-v(x)\big]}{h}\right) + \lim_{h\to 0}\left(\frac{\big[u(x+h)-u(x)\big]}{h}\right)v(x) \quad (1.12)$$

$$= u'(x)v(x) + u(x)v'(x) \quad (1.13)$$

■

Theorem 1.4: Generalized Product Rule

Presupposing the product rule for any finite number of functions of x, $f_1, f_2, ..., f_k$,

$$\Big(f_1(x)f_2(x)...f_k(x)\Big)' = \sum_{j=1}^{k}\left(f_j'(x)\prod_{\substack{q=1\\q\neq j}}^{k}\Big[f_q(x)\Big]\right). \tag{1.14}$$

Proof for Theorem.

I proceed by induction: the base case is given by the product rule. Then let $\Gamma(n)$ denote the statement that Eqn. 1.14 holds true when $k = n$. Assume $\psi(n)$ is true and consider $\Big(f_1(x)f_2(x)...f_n(x)f_{n+1}(x)\Big)'$. By the product rule:

$$\Big(f_1(x)f_2(x)...f_n(x)f_{n+1}(x)\Big)' = \Big(f_1(x)f_2(x)...f_n(x)\Big)'f_{n+1}(x) + f_1(x)f_2(x)...f_n(x)f_{n+1}'(x). \tag{1.15}$$

Then as $\Gamma(n)$ was assumed true:

$$\begin{aligned}
\Big(f_1(x)f_2(x)...f_n(x)f_{n+1}(x)\Big)' &= \Big(\big[f_1(x)f_2(x)...f_n(x)\big]f_{n+1}(x)\Big)' && (1.16)\\
&= \big[f_1(x)f_2(x)...f_n(x)\big]'f_{n+1}(x) + \big[f_1(x)f_2(x)...f_n(x)\big]\big(f_{n+1}(x)\big)' && (1.17)\\
&= \sum_{j=1}^{n}\left(f_j'(x)\prod_{\substack{q=1\\q\neq j}}^{n}\Big[f_q(x)\Big]\right)f_{n+1}(x) + f_1(x)f_2(x)...f_n(x)f_{n+1}'(x) && (1.18)\\
&= \sum_{j=1}^{n+1}\left(f_j'(x)\prod_{\substack{q=1\\q\neq j}}^{n+1}\Big[f_q(x)\Big]\right). && (1.19)
\end{aligned}$$

Therefore, $\Gamma(n+1)$ is true if $\Gamma(n)$ is, i.e. $\Gamma(n) \Rightarrow \Gamma(n+1)$, and the product rule shows that $\Gamma(2)$ is true. This implies $\Gamma(n)$ holds for all values of $n \in \mathbb{Z}$ greater than one. ∎

This then gives a much nicer way of deriving the derivative of x^n than the usual way.

Theorem 1.5: Derivative of x^n

$\big(x^n\big)' = nx^{n-1}$

Proof for Theorem.

By re-expressing x^n, and using the generalized product rule, we can derive:

$$\left(x^n\right)' = \left(\prod_{j=1}^{n}(x)\right)' = \sum_{k=1}^{n}\left[\prod_{j=1}^{k-1}(x)\cdot 1\cdot \prod_{j=k+1}^{n}(x)\right] = \sum_{k=1}^{n}\left[x^{n-1}\right] = nx^{n-1}. \tag{1.20}$$

■

Theorem 1.6: Quotient Rule

$$\left(\frac{u(x)}{v(x)}\right)' = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2} \tag{1.21}$$

Proof for Theorem.

$$\left(\ln\left(\frac{u(x)}{v(x)}\right)\right)' = \left(\ln\left(u(x)\right)\right)' - \left(\ln\left(v(x)\right)\right)' = \frac{u'(x)}{u(x)} - \frac{v'(x)}{v(x)} \tag{1.22}$$

But similarly,

$$\left(\ln\left(\frac{u(x)}{v(x)}\right)\right)' = \left(\frac{u(x)}{v(x)}\right)'\frac{v(x)}{u(x)}. \tag{1.23}$$

Equating these last two derivations:

$$\left(\frac{u(x)}{v(x)}\right)'\frac{v(x)}{u(x)} = \frac{u'(x)}{u(x)} - \frac{v'(x)}{v(x)} \tag{1.24}$$

$$\Rightarrow \left(\frac{u(x)}{v(x)}\right)' = \frac{u(x)}{v(x)}\left(\frac{u'(x)}{u(x)} - \frac{v'(x)}{v(x)}\right) = \frac{u(x)}{v(x)}\frac{u'(x)v(x) - u(x)v'(x)}{u(x)v(x)} \tag{1.25}$$

$$= \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2} \tag{1.26}$$

■

2 Differential and Partial Differential Equations

Within all of physics, the ability to solve differential equations is key. Most - in fact almost all - of the main equations describing physical systems are differential equations. Examples include:

- Newton's second law of motion: a central tool of classical mechanics
- Schrodinger Equation: the central equation of quantum mechanics
- Einstein Equation: a central equation of general relativity

- Navier-Stokes Equation: the equivalent of Newton's second law of motion in fluid dynamics
- Maxwell's Equations: a complete description of electromagnetism

They are also of great use further afield, such as the BlackScholes equation, in finance; Lanchester's laws, in military strategy; and the SIR model, in epidemiology.

2.1 First Order Ordinary Differential Equations

The most simple form of differential equations are those that only feature first order derivatives. In Sec. 2.1, I will proceed through a series of methods of solving such equations.

Separation of Variables

Definition 2.1: Separable Differential Equations

A first order differential equation is separable if and only if it can be re-written as:

$$\frac{dy}{dx} = f(x)g(y), \tag{1.27}$$

up to a different choice of variables.

Theorem 2.2: Solution of Separable Differential Equations

Any separable differential equation, as in Eqn. 1.27, has the solution:

$$\int \left(\frac{1}{g(y)}\right) dy = \int \Big(f(x)\Big) dx, \tag{1.28}$$

this can likely be simplified further.

Proof for Theorem.

Starting from Eqn. 1.27:

$$\frac{dy}{dx} = f(x)g(y) \iff \frac{1}{g(y)}\frac{dy}{dx} = f(x) \iff \int \left(\frac{1}{g(y)}\frac{dy}{dx}\right) dx = \int \Big(f(x)\Big) dx \tag{1.29}$$

$$\iff \int \left(\frac{1}{g(y)}\right) dy = \int \Big(f(x)\Big) dx \tag{1.30}$$

■

Homogeneous and Inhomogeneous Differential Equations

Definition 2.3: Homogeneous and Inhomogeneous Differential Equations

A first order homogeneous linear differential equation is a first order differential equation which may be expressed as:

$$\frac{dy}{dx} + f(x)y = 0, \tag{1.31}$$

where f is some function purely of x. A first order inhomogeneous linear differential equation is a first order differential equation which may be expressed as:

$$\frac{dy}{dx} + f(x)y = g(y), \tag{1.32}$$

where f is some function purely of x and g is some function purely of y. If given an inhomogeneous differential equation as in Eqn. 1.32, call the instance of Eqn. 1.31 - with the same $f(x)$ - the corresponding homogeneous differential equation.

I am very lucky that we can find a general solution for homogeneous linear differential equations, which we now do.

Theorem 2.4: General Solution of Homogeneous Linear Differential Equations

The solution to any differential equation of the form:

$$\frac{dy}{dx} + f(x)y = 0, \tag{1.33}$$

is:

$$y = e^{-\int \big(f(x)\big)dx} \tag{1.34}$$

Proof for Theorem.

To confirm the claimed solution, I simply substitute it into the left hand side of Eqn. 1.33 and using the chain rule:

$$\frac{d}{dx}\left(e^{-\int \big(f(x)\big)dx}\right) + f(x)e^{-\int \big(f(x)\big)dx} = \frac{d}{dx}\left(-\int \big(f(x)\big)dx\right)e^{-\int \big(f(x)\big)dx} + f(x)e^{-\int \big(f(x)\big)dx}. \tag{1.35}$$

By the fundamental theorem of calculus, this equates to:

$$\Big[-f(x) + f(x)\Big]e^{-\int \big(f(x)\big)dx} = 0. \tag{1.36}$$

This is exactly as required to prove that the claimed solution is a correct solution. ■

Note that, due to the linearity of the differential equation, multiplying the solution proven above by any complex number also provides a solution.
Before continuing on to look at inhomogeneous linear first order differential equations, I show that the solutions for homogeneous linear differential remain useful when examining first order inhomogeneous linear differential equations.

Theorem 2.5: General Solution of First Order Inhomogeneous Linear Differential Equations

For any first order inhomogeneous linear differential equation, with solution $y_I(x)$; if $y_H(x)$ is a solution of the corresponding homogeneous differential equation, then, for any $\lambda \in \mathbb{C}$, $y_I(x) + \lambda y_H(x)$ is also a solution for the aforementioned inhomogeneous differential equation.

Proof for Theorem.

By assumption,

$$\frac{dy_H}{dx} + f(x)y_H = 0 \text{ and } \frac{dy_I}{dx} + f(x)y_I = g(y_I). \tag{1.37}$$

Then substituting $y_I(x) + \lambda y_H(x)$ into the left hand side of Eqn. 1.32 gives:

$$\frac{d}{dx}\Big(y_I(x) + \lambda y_H(x)\Big) + f(x)\Big[y_I(x) + \lambda y_H(x)\Big] \tag{1.38}$$

$$= \frac{d}{dx}\Big(y_I(x)\Big) + \lambda\frac{d}{dx}\Big(y_H(x)\Big) + f(x)\Big[y_I(x)\Big] + \lambda f(x)\Big[y_H(x)\Big] \tag{1.39}$$

$$= \left\{\frac{dy_I}{dx} + f(x)y_I(x)\right\} + \lambda\left\{\frac{dy_H}{dx} + f(x)y_H(x)\right\} \tag{1.40}$$

$$= \Big\{g(y_I)\Big\} + \lambda\Big\{0\Big\} = g(y_I), \tag{1.41}$$

which is exactly as required to be a solution of the inhomogeneous differential equation we were examining. ■

Integrating Factor

Definition 2.6: Integrating Factors

If given a differential equation of the form:

$$\frac{dy}{dx} + f(x)y = g(x), \tag{1.42}$$

then the corresponding integrating factor, $I(x)$, is a function of x, defined as:

$$I(x) = e^{\int \Big(f(x) \Big) dx} \tag{1.43}$$

Theorem 2.7

The solution to any differential equation of the form:

$$\frac{dy}{dx} + f(x)y = g(x), \tag{1.44}$$

is:

$$y = \frac{1}{I(x)} \int \Big(I(x)g(x) \Big) dx. \tag{1.45}$$

Proof for Theorem.

Starting from Eqn. 1.44 and multiplying through by the relevant integrating factor:

$$I(x)\frac{dy}{dx} + I(x)f(x)y = I(x)g(x). \tag{1.46}$$

Then integrating both sides with respect to x:

$$\int \Big(I(x)\frac{dy}{dx} + I(x)f(x)y \Big) dx = \int \Big(I(x)g(x) \Big) dx. \tag{1.47}$$

Focusing - for just one second - exclusively on the left hand side of Eqn. 1.47:

$$\int \big(I(x)\frac{dy}{dx} + I(x)f(x)y \big) dx = \int \left(\frac{dy}{dx} e^{\int \Big(f(x) \Big) dx} + f(x)ye^{\int \big(f(x) \big) dx} \right) dx \tag{1.48}$$

$$= \int \left(\frac{d}{dx} \left[ye^{\int \big(f(x) \big) dx} \right] \right) dx = ye^{\int \big(f(x) \big) dx} = yI(x). \tag{1.49}$$

Substituting this back into Eqn. 1.47 gives:

$$yI(x) = \int \Big(I(x)g(x) \Big) dx \Rightarrow y = \frac{1}{I(x)} \int \Big(I(x)g(x) \Big) dx. \tag{1.50}$$

■

Examples

2.2 Second Order Ordinary Differential Equations

Homogeneous Second Order Ordinary Differential Equations with Constant Coefficients

Definition 2.8: Homogeneous Second Order Ordinary Differential Equations With Constant Coefficients

A homogeneous second order ordinary differential equation, with constant coefficients is a second order differential equation of the form:

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0. \tag{1.51}$$

Theorem 2.9: General Solution of Homogeneous Second Order Ordinary Differential Equations With Constant Coefficients

Any differential equation of the form:

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0. \tag{1.52}$$

has solutions of the form:

$$Ae^{\lambda_1 x} + Be^{\lambda_2 x}, \tag{1.53}$$

where A and B can be any complex numbers, and λ_1, λ_2 are the distinct[a] solutions of the quadratic equations:

$$a\lambda^2 + b\lambda + c = 0. \tag{1.54}$$

[a] If they are not distinct, then this solution does not apply

Proof for Theorem.

I use the ansatz $e^{\lambda x}$ and substitute it into Eqn. 1.52:

$$a\frac{d^2}{dx^2}\left(e^{\lambda x}\right) + b\frac{d}{dx}\left(e^{\lambda x}\right) + ce^{\lambda x} = 0 \iff a\lambda^2 e^{\lambda x} + b\lambda e^{\lambda x} + ce^{\lambda x} = 0 \tag{1.55}$$

$$\iff \left(a\lambda^2 + b\lambda + c\right)e^{\lambda x} = 0 \tag{1.56}$$

As the solution must obey this equation for any value of x, this can only be satisfied if:

$$a\lambda^2 + b\lambda + c = 0. \tag{1.57}$$

The solutions of this equations were already defined as λ_1 and λ_2. Hence both,

$$e^{\lambda_1 x} \text{ and } e^{\lambda_2 x} \tag{1.58}$$

are solutions to the differential equation in Eqn. 1.52. As Eqn. 1.52 is clearly linear, this implies that the general solution is:

$$Ae^{\lambda_1 x} + Be^{\lambda_2 x}, \tag{1.59}$$

where A and B can be any complex numbers. ■

I then just have to handle the case that was excluded from the above theorem: if $a\lambda^2 + b\lambda + c = 0$ has repeated roots.

Theorem 2.10: Homogeneous Second Order Linear Differential Equations (With Repeated Roots)

If $a\lambda^2 + b\lambda + c = 0$ has only a single solution, λ_1, then the differential equation:

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0 \tag{1.60}$$

has the solution:

$$\Big[Ax + B\Big]e^{\lambda_1 x}, \tag{1.61}$$

, where A and B can be any complex numbers.

Proof for Theorem.

Again, this can be demonstrated by substituting the candidate solution into the right hand side of Eqn. 1.60:

$$a\frac{d^2}{dx^2}\left(\Big[Ax + B\Big]e^{\lambda_1 x}\right) + b\frac{d}{dx}\left(\Big[Ax + B\Big]e^{\lambda_1 x}\right) + c\Big[Ax + B\Big]e^{\lambda_1 x} \tag{1.62}$$

$$= a\frac{d}{dx}\left(Ae^{\lambda_1 x} + \lambda_1\Big[Ax + B\Big]e^{\lambda_1 x}\right) + b\left(Ae^{\lambda_1 x} + \lambda_1\Big[Ax + B\Big]e^{\lambda_1 x}\right) + c\Big[Ax + B\Big]e^{\lambda_1 x} \tag{1.63}$$

$$= aA\lambda_1 e^{\lambda_1 x} + a\lambda_1 Ae^{\lambda_1 x} + a\lambda_1^2\Big[Ax + B\Big]e^{\lambda_1 x} + bAe^{\lambda_1 x} + b\lambda_1\Big[Ax + B\Big]e^{\lambda_1 x} + c\Big[Ax + B\Big]e^{\lambda_1 x} \tag{1.64}$$

$$= \Big(2aA\lambda_1 + a\lambda_1^2 B + bA + b\lambda_1 B + cB\Big)e^{\lambda_1 x} + \Big(a\lambda_1^2 A + b\lambda_1 A + cA\Big)xe^{\lambda_1 x} \tag{1.65}$$

As λ_1 is defined as the solution of $a\lambda^2 + b\lambda + c = 0$, $a\lambda_1^2 A + b\lambda_1 A + cA = 0$. Therefore the

above is equivalent to:

$$\Big(2aA\lambda_1 + a\lambda_1^2 B + bA + b\lambda_1 B + cB\Big)e^{\lambda_1 x} = \Big(2aA\lambda_1 + bA + \big[a\lambda_1^2 + b\lambda_1 + c\big]B\Big)e^{\lambda_1 x} \tag{1.66}$$

$$= \Big(2a\lambda_1 + b\Big)Ae^{\lambda_1 x}. \tag{1.67}$$

Considering the quadratic formula and that λ_1 is the only solution to $a\lambda^2 + b\lambda + c = 0$, $\lambda_1 = \frac{-b}{2a}$,

$$a\frac{d^2}{dx^2}\Big(\Big[Ax + B\Big]e^{\lambda_1 x}\Big) + b\frac{d}{dx}\Big(\Big[Ax + B\Big]e^{\lambda_1 x}\Big) + c\Big[Ax + B\Big]e^{\lambda_1 x} \tag{1.68}$$

$$= \Big(2a\Big[\frac{-b}{2a}\Big] + b\Big)Ae^{\lambda_1 x} \tag{1.69}$$

$$= 0. \tag{1.70}$$

2.3 Initial Value Problems and Other Boundary Conditions

The discerning reader may have noticed that in all the above differential equations no single function is derived as a solution. There were always As or Bs that may be arbitrary real numbers within the solution. This was not wrong, but nature is not like this: a ball thrown through the air does not have a range of solutions describing its motion. It has but one.
So how does the multitude of solutions we obtain from the differential equations describing the physics of the situation reduce to exactly one?
The answer lies in the initial conditions (or other boundary conditions).

Definition 2.11: Initial Conditions

An initial condition is a restriction on a solution to a differential equation that demands that the solution to the problem - a function - takes a particular value when the argument is set to a particular value (typically, that - if the solution is a function of time - the solution takes a particular value, perhaps being in a particular position, when $t = 0$).

These initial conditions are satisfied by taking our general solution to the differential equation, without the initial conditions, and using the required conditions derive the values of the unknown/arbitrary values - in the above cases these were As and Bs - and solve for the particular values of them that make the required conditions true.
This point may be best illustrated with an example.
Example Initial Value Problem

Solve the differential equation:

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} + y = 0, \tag{1.71}$$

subject to the initial condition:

$$y(0) = 0 \text{ and } y'(1) = 4e. \tag{1.72}$$

Solution
We already know that this differential equation, without the initial condition has solution:

$$y(x) = \Big[Ax + B\Big]e^x. \tag{1.73}$$

So to impose the conditions, I must solve - for A and B - the simultaneous equations:

$$y(0) = \Big[A0 + B\Big]e^0 = 0, \tag{1.74}$$

$$y'(1) = Ae^1 + \Big[A + B\Big]e^1 = 4e. \tag{1.75}$$

These simultaneous equations can be seen to have solution:

$$A = 2 \tag{1.76}$$

$$B = 0. \tag{1.77}$$

Hence, the solution to the initial value problem is:

$$y(x) = \Big[2x + 0\Big]e^x = 2xe^x. \tag{1.78}$$

3 Matrices

3.1 Basic Definitions

For the purposes of quantum mechanics, we only need to consider square matrices.

Definition 3.1: Square Matrices

A $n \times n$ square matrix over a set $\mathbb{F}$, that is typically $\mathbb{R}$ or $\mathbb{C}$, is a set of elements of $\mathbb{F}$ arranged into a two-dimensional grid. The element of $\mathbb{F}$ in the ith row and jth column of the matrix A is labelled A_{ij}. The set of all $n \times n$ square matrices over $\mathbb{F}$ is often denoted as $\mathbb{F}^{n\times n}$.

The important features of matrices are how they interact with each other under specific operations. The two main operations are addition and multiplication, defined as:

Definition 3.2: Matrix Addition

We define matrix addition of two matrices, A and B, in terms of the elements of the resulting matrix and the input matrices:

$$\big(A+B\big)_{ij} = A_{ij} + B_{ij}. \tag{1.79}$$

Definition 3.3: Matrix Multiplication

We define matrix multiplication of two matrices, A and B, in terms of the elements of the resulting matrix and the input matrices:

$$\big(AB\big)_{ij} = \sum_{k=1}^{n} \Big(A_{ik} B_{kj} \Big). \tag{1.80}$$

3.2 Commutators

For the purposes of quantum mechanics, perhaps the most important aspect of matrices is that they do not commute under matrix multiplication.

Definition 3.4: Commutation

Two objects, A and B, are said to commute under binary operation, $*$, if and only if:

$$A * B = B * A. \tag{1.81}$$

Therefore, given two matrices, we say that they do not commute under matrix multiplication. A key example of this is:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{1.82}$$

$$\text{Therefore, } \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.83}$$

The main way we capture the commutativity / non-commutativity of two matrices is commutators. We define these as:

Definition 3.5: Commutators

Given two matrices, A and B, define the commutator, $\big[A, B\big]$, of the two matrices as:

$$\big[A, B\big] = AB - BA. \tag{1.84}$$

The main way we use this concept is to rearrange the order of matrix multiplication.

Lemma 1. *For any four matrices, A, B, C, D:*

$$ABCD = ACBD + A\big[B, C\big]D \tag{1.85}$$

Proof. Start from the definition of a commutator between two matrices, B and C:

$$\big[B, C\big] = BC - CB \Rightarrow BC = \big[B, C\big] + CB. \tag{1.86}$$

Therefore,

$$ABCD = A\big(BC\big)D = A\big(\big[B, C\big] + CB\big)D = ACBD + A\big[B, C\big]D. \tag{1.87}$$

□

Note that we have achieved the required rearrrangement at the cost of the additional term. The single most important value a commutator can take is zero. This is the case if and only if the two argument matrices commute.

Theorem 3.6: Commutator of Commuting Matrices

The commutator of any two matrices is zero if and only if they commute.

Proof for Theorem.

I prove this theorem in two parts:
The commutator of any two matrices is zero if they commute
Let A and B be any two matrices that commute. Then - by definition of commutation - $AB = BA$. Therefore,

$$\big[A, B\big] = AB - BA = AB - AB = 0. \tag{1.88}$$

The commutator of any two matrices is zero only if they commute
For the other half of the theorem note that this is equivalent to if two matrices have a commutator equal to zero, then they must commute. Assume

$$\big[A, B\big] = 0 \Rightarrow AB - BA = 0 \Rightarrow AB = BA. \tag{1.89}$$

Therefore, A and B commute. ■

3.3 Matrices as Transforms on a Vector Space

The first thing to define is a vector space.

Definition 3.7: Vector Spaces (informally)

We - for our purposes herein - define a vector space over some set of numbers (such at $\mathbb{R}$ or $\mathbb{C}$)[a] as the set of all vectors of a particular size where the entries are all in the specified set. The set of vectors of size $n \in \mathbb{N}$ with all elements being in $\mathbb{R}$ or $\mathbb{C}$ is denoted as $\mathbb{R}^n$ or $\mathbb{C}^n$, respectively.

[a] Formally this can be any field, but do not worry about that for now

I now take a slightly different perspective on matrices but one motivated by something you should be familiar with.
You are likely aware that vectors can be represented as columns of numbers e.g.

$$\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix} \text{ or } \begin{pmatrix}1\\0\\0\\0\\0\end{pmatrix} \text{ or } \begin{pmatrix}1+i\\0\\i\\0\\7\end{pmatrix} \text{ or } \begin{pmatrix}8+7i\\\frac{2}{3}\\2i\\0\\\pi\end{pmatrix} \text{ or } 5\begin{pmatrix}1\\1\\2i\\0\\6\end{pmatrix} \tag{1.90}$$

But I then note that I can consider the column vectors - as above - as a matrix and apply the previously established rules of matrix multiplication to allow a matrix (of the right size) to act on the vector:

Definition 3.8: Multiplying a Vector by a Matrix

For any $n \times n$ matrix, A, and column vector of size n, $\vec{v}$, define the multiplication of $\vec{v}$ by A (also known as the application of A to $\vec{v}$) as:

$$\left(A\vec{v}\right)_i = \sum_{k=1}^{n} \left(A_{ik}\vec{v}_k\right). \tag{1.91}$$

The key thing to observe is that - generally - $A\vec{v} \neq \vec{v}$. Although there are - very! - important cases where $A\vec{v} = \vec{v}$, as we shall see later. Hence we can see that multiplication by a matrix, in this case A, has changed/transformed $\vec{v}$. As A can be applied to *any* vector of the correct size (i.e. of the same size as $\vec{v}$), A can be seen as transforming the set of all vectors of that size. Therefore, A, or any $n \times n$ matrix, can - and often is, like we will do later - be alternatively viewed as a transformation on the vector space of all vectors of size n. I now give an example that hopefully will prove illustrative.

Consider a classical[1] particle in three-dimensional space. Its position can be described entirely by a vector from $\mathbb{R}^3$. Any 3×3 matrix over the real numbers can be seen as transforming the vector representing the particles position: we can therefore interpret it as representing moving the particle. Consider the particular example:

$$\vec{v} = \begin{pmatrix}1\\0\\0\end{pmatrix}, \tag{1.92}$$

which described the particle as being in a particular position. If I then pick a 3×3 matrix over the real numbers (and label it A):

$$A = \begin{pmatrix}1,2,3\\4,5,6\\7,8,9\end{pmatrix}, \tag{1.93}$$

[1] If this does not mean anything to you yet, do not worry: just pretend that the word classical is not here.

we can view this matrix, A, as moving the particle from the location represented by $\vec{v}$ to the location represented by $A\vec{v}$, i.e. to:

$$A\vec{v} = \begin{pmatrix} 1,2,3 \\ 4,5,6 \\ 7,8,9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix}. \tag{1.94}$$

3.4 Determinants

I define the determinant of a matrix as:

Definition 3.9: Determinant of a Matrix

For a $n \times n$ square matrix, M, the determinant, denoted det, is defined recursively: $\forall i \in \mathbb{N}^{\leq n}$,

$$\det\big(M\big) = \sum_{j=1}^{n} \Big[\big(-1\big)^{i+j} M_{i,j} \det\big(m_{i,j}\big) \Big], \tag{1.95}$$

where $m_{i,j}^{M}$ is a matrix obtained from M by removing the ith row and the jth column.

I then prove several important properties of the determinant:

Theorem 3.10

Exchanging any two neighbouring rows in a matrix multiplies its determinant by -1.

Proof for Theorem.

Consider two matrices: M and M', where M' is M but with neighbouring rows r_1 and r_2 exchanged (assume WLOG that $r_2 = r_1 + 1$). Then consider:

$$\det\big(M'\big) = \sum_{j=1}^{n} \Big[\big(-1\big)^{r_1+j} M'_{r_1,j} \det\big(m_{r_1,j}^{M'}\big) \Big] \tag{1.96}$$

$$\det\big(M\big) = \sum_{j=1}^{n} \Big[\big(-1\big)^{r_2+j} M_{r_2,j} \det\big(m_{r_2,j}^{M}\big) \Big] = \sum_{j=1}^{n} \Big[\big(-1\big)^{r_1+1+j} M_{r_2,j} \det\big(m_{r_2,j}^{M}\big) \Big]. \tag{1.97}$$

Due to the exchanging of the rows: $m_{r_2,j}^{M} = m_{r_1,j}^{M'}$ and $M_{r_2,j} = M'_{r_1,j}$. Therefore,

$$\det\big(M\big) = -\sum_{j=1}^{n} \Big[\big(-1\big)^{r_1+j} M_{r_1,j} \det\big(m_{r_1,j}^{M}\big) \Big] = -\det\big(M'\big). \tag{1.98}$$

■

Theorem 3.11: Determinant of a Matrix with Repeated Rows

Any matrix where two rows are identical has a determinant of zero.

Proof for Theorem.

Let M be a matrix where rows r_1 and r_2 are identical. Then let M' be M but with a series of exchanges performed such that the two identical rows are neighbouring. Then $\det\big(M'\big)$ is $\det\big(M\big)$ up to a possible factor of -1. Exchanging the two neighbouring and identical rows in M' must multiply its determinant by a factor of -1 but as those rows are identical it must also leave the determinant unchanged. Therefore the determinant must be zero, so that both statements can be satisfied. ∎

3.5 Inverse of a Matrix

An important class of matrices to examine are the identities:

Definition 3.12: The identity of size $n \in \mathbb{N}$

For each $n \in \mathbb{N}$, there is a unique $n \times n$ matrix such that for any $n \times n$ matrix, A:

$$AI = IA = A. \tag{1.99}$$

I then define the inverse of a matrix.

Definition 3.13: Inverse of a Matrix

The inverse, M^{-1}, of a matrix, M, is any matrix such that:

$$MM^{-1} = M^{-1}M = I. \tag{1.100}$$

I then turn to how to calculate the inverse.

Theorem 3.14: Calculating a Matrix Inverse

For any matrix, M, for which an inverse exists[a], that inverse is:

$$M^{-1} = \frac{C^T}{\det\big(M\big)}, \tag{1.101}$$

where $C_{i,j} = (-1)^{i+j}\det\big(m^M_{i,j}\big)$.

[a] This is the case whenever it does not have 0 for an eigenvalue.

Proof for Theorem.

Consider the definition of matrix multiplication, then for any $n \times n$ matrix,

$$\left(MM^{-1}\right)_{ij} = \sum_{k=1}^{n} \left(A_{ik}(-1)^{j+k} \det\left(m^{A}_{j,k}\right) \right) \frac{1}{\det\left(A\right)} \tag{1.102}$$

$$= \begin{cases} \det\left(A\right)\dfrac{1}{\det\left(A\right)} = 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}, \tag{1.103}$$

with the second case following from its summation's equivalence to the determinant of a matrix with repeated rows. Therefore,

$$MM^{-1} = I. \tag{1.104}$$

It is also required that $M^{-1}M = I$ but this is similar and I leave it as an exercise. ■

3.6 Eigenvalues and Eigenvectors

When an individual matrix is being considered, perhaps its most important properties are its eigenvalues and eigenvectors. These are defined by:

Definition 3.15: Eigenvalues and Eigenvectors

For any matrix, $M \in \mathbb{C}^{n\times n}$, an eigenvector, $|\lambda\rangle \in \mathbb{C}^n$, and its corresponding eigenvalue, $\lambda \in \mathbb{C}$, of M satisfy the equation:

$$M|\lambda\rangle = \lambda|\lambda\rangle. \tag{1.105}$$

Eigenvalues and eigenvalues are important, but perhaps the most important thing to know about them is how to find them. To consider this, examine the equation defining eigenvector and eigenvalues:

$$M|\lambda\rangle = \lambda|\lambda\rangle \iff \left(M - \lambda\hat{I}\right)|\lambda\rangle = 0\rangle. \tag{1.106}$$

Assuming $|\lambda\rangle$ is not zero, this only has a solution if:

$$\left|M - \lambda\hat{I}\right| = 0 \tag{1.107}$$

Eqn. 1.107 is called the characteristic equation of M; if M is an $n \times n$ matrix, it is an nth order polynomial in λ and hence has n roots which are the eigenvalues of M. Hence the eigenvalues can be obtained by solving the characteristic polynomial.

Using each of the now known eigenvalues, denoted λ_0, in:

$$M|\lambda_0\rangle = \lambda_0|\lambda_0\rangle, \tag{1.108}$$

provides a series of n (assuming again that M is a $n \times n$ matrix) simultaneous equations, each with n variables. Hence this system of equations can be solved to find the corresponding eigenvector $|\lambda_0\rangle$.

3.7 Diagonalizations of Matrices

As previously established, a $n \times n$ matrix has n eigenvectors. For ease, index them so that for each positive integer, $j, \leq n$ there is a corresponding distinct eigenvector, $|\lambda_j\rangle$. Similarly, let $\mathbf{1}_j$ denote the n-element vector where all elements but the jth are zero and the jth is one. E.g. if $n = 3$,

$$\mathbf{1}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \tag{1.109}$$

It is reasonably easy to see that there exists a matrix V such that, $\forall j \in \mathbb{N}^{\leq n}$:

$$|\lambda_j\rangle = V\mathbf{1}_j. \tag{1.110}$$

If I then define D_M to be the diagonal matrix where $\left(D_M\right)_{j,j}$ is the eigenvalue of M corresponding the the eigenvector $|\lambda_j\rangle$.

Theorem 3.16: Diagonalization of a Matrix

For any $n \times n$ matrix, M, with n linearly-independent eigenvectors, may be expressed as:

$$M = VD_MV^{-1}, \tag{1.111}$$

where V is defined by:

$$|\lambda_j\rangle = V\mathbf{1}_j. \tag{1.112}$$

Proof for Theorem.

As M is a $n \times n$ matrix with n linearly-independent eigenvectors, any n-element vector may be expresses as a linear combination of the eigenvectors of M. Therefore, if any other matrix has the exact same eigenvectors and eigenvalues as M it must affect every vector the same as M and so must be equal to M. In light of this, consider:

$$VD_MV^{-1}|\lambda_j\rangle = VD_MV^{-1}V\mathbf{1}_j = VD_M\mathbf{1}_j = V\left(D_M\right)_{j,j}\mathbf{1}_j = \left(D_M\right)_{j,j}V\mathbf{1}_j = \left(D_M\right)_{j,j}|\lambda_j\rangle. \tag{1.113}$$

As $\left(D_M\right)_{j,j}$ was defined as the eigenvalue of M corresponding to $|\lambda_j\rangle$, this shows:

$$M = VD_MV^{-1}. \tag{1.114}$$

■

3.8 Conjugate Transpose

One operation on matrices that does not immediately seem to be important but ends up being immensely important to quantum mechanics is the conjugate transpose.

Definition 3.17: Conjugate Transpose of a Matrix

For any matrix, M, the conjugate transpose of of M, denoted $M^\dagger$, is defined:

$$\left(M^\dagger\right)_{i,j} = \left[\left(M\right)_{j,i}\right]^*, \tag{1.115}$$

where $\left[\,\cdot\,\right]^*$ denotes the complex conjugate of its argument.

3.9 Commonly Used Subsets of Matrices

Throughout physics, we repeatedly find that certain subsets of matrices are especially important. Here I present two that are particularly important to quantum mechanics.

Definition 3.18: Hermitian Matrices

A matrix, M, is Hermitian if and only if it is its own conjugate transpose.

Definition 3.19: Unitary Matrices

A matrix, M, is unitary if and only if it has an inverse and that inverse is its conjugate transpose.

I then prove the most important features of these two families of matrices.

Theorem 3.20: Product of a unitary matrix and its Conjugate Transpose

For any unitary matrix, U,

$$UU^\dagger = U^\dagger U = I. \tag{1.116}$$

Proof for Theorem.

By definition of a unitary matrix, $U^\dagger = U^{-1}$, therefore:

$$UU^\dagger = UU^{-1} = I \tag{1.117}$$
$$U^\dagger U = U^{-1}U = I. \tag{1.118}$$

■

Theorem 3.21: Eigenvalue of Hermitian Matrices

For any Hermitian matrix, all of its eigenvalues are real.

Proof for Theorem.

Let v_λ be an eigenvector of matrix, M, with eigenvalue, $\lambda \in \mathbb{C}$. Then,

$$Mv_\lambda = \lambda v_\lambda \tag{1.119}$$

Taking the complex transpose of each side:

$$v_\lambda^\dagger M^\dagger = \lambda^* v_\lambda^\dagger. \tag{1.120}$$

Multiplying, from the right, both sides by v_λ and noting that - as it is Hermitian - $M^\dagger = M$ gives:

$$v_\lambda^\dagger M v_\lambda = \lambda^* v_\lambda^\dagger v_\lambda \Rightarrow \lambda v_\lambda^\dagger v_\lambda = \lambda^* v_\lambda^\dagger v_\lambda \tag{1.121}$$

$$\Rightarrow \lambda = lambda^*. \tag{1.122}$$

Hence, v_λ must be real. ∎

I leave as an exercise, the proof that the V in the diagonalization of any matrix is unitary. Finally, I state an important result without proof.

Theorem 3.22: Relation Between Hermitian and Unitary Matrices

Any unitary matrix may be expressed as e^{itH}, where H is some Hermitian matrix.

Proof for Theorem.

Proof Omitted. ∎

4 Traces and Norms

Definition 4.1: Traces

For any matrix, M, its trace, denoted as:

$$\mathrm{Tr}\Big(M\Big), \tag{1.123}$$

is the sum of the eigenvalues of M.

Definition 4.2: Schatten Norms

$\forall p \in \mathbb{N}$, define the Schatten p-norm as a function from matrices to the reals:

$$||M||_p = \left[\mathrm{Tr}\left(\left(M^\dagger M\right)^{p/2}\right)\right]^{1/p}. \tag{1.124}$$

As Schatten p-norms are defined on matrices, via representation theory, they are also defined on linear operators.

Definition 4.3: Trace Norm

The trace norm (also known as the nuclear norm) is the Schatten p-norm with $p = 1$.

Definition 4.4: Trace Distance

For any two matrices, ρ, σ, define the trace distance, $\mathcal{D}\big(\rho, \sigma\big)$, as:

$$\mathcal{D}\big(\rho, \sigma\big) = \frac{1}{2}\big|\big|\rho - \sigma\big|\big|_1. \tag{1.125}$$

Definition 4.5: Variation Distance

Given two probability distributions over the set of possible outcomes, Ω, (i.e. mappings that take a possible outcome, $s \in \Omega$, as input and return a number from $[0, 1]$), P and Q, define the variational distance between them, $VD\big(P, Q\big)$, by:

$$VD\big(P, Q\big) = \frac{1}{2}\sum_{s\in\Omega}\Big|P\big(s\big) - Q\big(s\big)\Big|. \tag{1.126}$$

5 Taylor Series

Theorem 5.1: Integration by Parts

For any differentiable functions u and v,

$$\int \Big[uv'\Big]dx = uv - \int \Big[u'v\Big]dx. \tag{1.127}$$

Proof for Theorem.

Let u, v be differentiable functions of x and start from the product rule (as in Def. **??**)

$$\big(uv\big)' = u'v + uv'. \tag{1.128}$$

I then take the integral of both sides of Eqn. 1.128:

$$\int \Big[(uv)'\Big]dx = \int \Big[u'v + uv'\Big]dx = \int \Big[u'v\Big]dx + \int \Big[uv'\Big]dx. \tag{1.129}$$

Hence, applying the fundamental theorem of calculus (axiom 14) to the left-most part of Eqn. 1.129 gives:

$$uv = \int \Big[u'v\Big]dx + \int \Big[uv'\Big]dx. \tag{1.130}$$

Eqn. 1.130 can then be re-arranged to give the formula for integration by parts (as in Eqn. **??**):

$$\int \Big[uv'\Big]dx = uv - \int \Big[u'v\Big]dx. \tag{1.131}$$

■

Theorem 5.2: Taylor's Theorem

For any function, f, such that for any interval on the real line, I, there exists some constant, $C \in \mathbb{R}$, such that:

$$\forall x \in I,\ \Big|f^{(k)}(x)\Big| < C^{k+1}k!, \tag{1.132}$$

where $f^{(k)}(x)$ denotes the kth derivative of f with respect to x[a]. $\forall k \in \mathbb{N}$, $\forall p \in \mathbb{R}$:

$$f(x) = \sum_{j=0}^{k}\left(\frac{f^{(j)}(p)}{j!}(x-p)^j\right) + \int_p^x \left(\frac{f^{(k+1)}(y)}{k!}(x-y)^k\right)dy. \tag{1.133}$$

[a] This condition is equivalent to assuming the function f is analytic.

Proof for Theorem.

I start with a definition of an integral, $J[x,n]$. This is not new maths or a new axiom, it is just notation.

$$J[x,n] = \int_0^x \left(\frac{f^{(n+1)}(y)}{n!}(x-y)^n\right)dy. \tag{1.134}$$

Then, integrating Eqn. 1.134 by parts:

$$\left[\frac{f^{(n+1)}(y)}{n!}\int\Big((x-y)^n\Big)dy\right]_0^x - \int_0^x\left(\frac{f^{(n+2)}(y)}{n!}\int\Big((x-y)^n\Big)dy\right)dy. \tag{1.135}$$

The integral $\int \Big((x-y)^n \Big) dy$ can be evaluated, hence:

$$J[x,n] = \Big[-\frac{f^{(n+1)}(y)}{n!} \frac{\big[x-y\big]^{n+1}}{n+1} \Big]_0^x + \int_0^x \Big(\frac{f^{(n+2)}(y)}{n!} \frac{\big[x-y\big]^{n+1}}{n+1} \Big) dy \tag{1.136}$$

$$= \Big[\frac{f^{(n+1)}(y)}{(n+1)!} \big[x-y\big]^{n+1} \Big]_x^0 + \int_0^x \Big(\frac{f^{(n+2)}(y)}{(n+1)!} \big(x-y\big)^{n+1} \Big) dy \tag{1.137}$$

$$= \frac{f^{(n+1)}(0)}{(n+1)!} x^{n+1} + J[x,n+1]. \tag{1.138}$$

To prove the existence of a Taylor series for any value of n, I then start from a base case using the fundamental theorem of calculus:

$$f(x) - f(0) = \int_0^x \Big(f'(y) \Big) dy \Rightarrow f(x) = f(0) + \int_0^x \Big(f'(y) \Big) dy, \tag{1.139}$$

which is the the required Taylor series when $n = 0$. Then, assuming the required Taylor series exists for a given value of $n \in \mathbb{Z}$:

$$f(x) = \sum_{j=0}^{n} \Big(\frac{f^{(j)}(0)}{j!} (x)^j \Big) + \int_0^x \Big(\frac{f^{(k+1)}(y)}{k!} (x-y)^n \Big) dy = \sum_{j=0}^{n} \Big(\frac{f^{(j)}(0)}{j!} (x)^j \Big) + J[x,n]. \tag{1.140}$$

Then, using Eqn. 1.138 to re-write J[x, n]:

$$f(x) = \sum_{j=0}^{n} \Big(\frac{f^{(j)}(0)}{j!} (x)^j \Big) + \frac{f^{(n+1)}(0)}{(n+1)!} x^{n+1} + J[x,n+1] = \sum_{j=0}^{n+1} \Big(\frac{f^{(j)}(0)}{j!} (x)^j \Big) + J[x,n+1]. \tag{1.141}$$

Therefore, the required Taylor series of $n+1$ exists. Therefore, axiom 18 allows me to - by induction - assert that the required Taylor series exists for any positive integer. As all the Taylor series produced above were always about zero, they are technically Maclaurin series. I now show that a Taylor series can always be obtained from a suitable Maclaurin series.

Define a new function in terms of the previously established general function: $g(x) = f(x-z)$, where $z \in \mathbb{R}$ and can be chosen as needed. For any chosen g a corresponding f exists, and for any chosen, f a corresponding g exists. Then,

$$g(z) = f(0), \tag{1.142}$$

$$g'(z) = f'(0), \tag{1.143}$$

$$g''(z) = f''(0), \tag{1.144}$$

$$g'''(z) = f'''(0), \tag{1.145}$$

$$\textit{and so on.} \tag{1.146}$$

This provides for a Taylor series to be taken about any point. To calculate this, an appropriate f can always be found. ■

Theorem 5.3: Infinite Taylor's Theorem

For any function, f, meeting all conditions of Taylor's theorem, there exists some neighbourhood[a] of any point, $p \in \mathbb{R}$, such that:

$$f(x) = \sum_{j=0}^{\infty} \left(\frac{f^{(j)}(p)}{j!}(x-p)^j \right). \tag{1.147}$$

[a] A neighbourhood of a specified point just means some area around that point of unspecified size.

Proof for Theorem.

By the inherited conditions for Taylor's theorem, for any interval centred on $p \in \mathbb{R}$, there exists some constant, $C \in \mathbb{R}$, such that:

$$\left| f^{(k)}(x) \right| < C^{k+1}k!. \tag{1.148}$$

Hence,

$$\frac{\left|f^{(k)}(x)\right|}{k!}|x-y|^k < \frac{C^{k+1}k!}{k!}|x-y|^k = C^{k+1}|x-y|^k. \tag{1.149}$$

This is then used to show that, $\forall k \in \mathbb{N}$,

$$\int_p^x \left(\frac{f^{(k+1)}(x)}{k!}(x-y)^k \right) dy \leq \left| \int_p^x \left(\frac{f^{(k+1)}(y)}{k!}(x-y)^k \right) dy \right| \tag{1.150}$$

$$\leq \int_p^x \left(\left| \frac{f^{(k+1)}(y)}{k!}(x-y)^k \right| \right) dy \leq \int_p^x \left(\frac{\left|f^{(k+1)}(y)\right|}{k!}\left|x-y\right|^k \right) dy \leq C^{k+1} \int_p^x \left(|x-y|^k \right) dy. \tag{1.151}$$

If I restrict x to be within ξ^{-1} - where $\xi = 1.1C$ - of p, then if y is restricted to within the limits of the integral in Eqn. 1.151, $|x-y| \leq \xi^{-1}$ and $x - p \leq \xi^{-1}$. Therefore, using this restriction:

$$\int_p^x \left(\frac{f^{(k+1)}(x)}{k!}(x-y)^k \right) dy \leq C^{k+1} \int_p^x \left(|x-y|^k \right) dy \leq C^{k+1}\xi^{-k} \int_p^x \left(1 \right) dy \tag{1.152}$$

$$\leq C^{k+1}\xi^{-k}(x-p) \leq C^{k+1}\xi^{-(k+1)} = \frac{C^{k+1}}{C^{k+1}(1.1)^{k+1}} \leq (1.1)^{-k}. \tag{1.153}$$

It can be shown, that $\forall k \in \mathbb{N}$, $(1.1)^{-k} > (1.1)^{-(k+1)}$, so I can extend the series to be infinite with zero error, i.e.

$$f(x) = \sum_{j=0}^{\infty} \left(\frac{f^{(j)}(p)}{j!}(x-p)^j \right). \tag{1.154}$$

■

6 Functions and Taylor Series of Matrices

For many functions that we can apply to the real or complex numbers, we can also apply them to matrices. We define these functions via their Taylor series. The most important one of these is exponentials. The Taylor expansion of the exponential function is: $\forall x \in \mathbb{C}$,

$$e^x = \sum_{j=1}^{\infty} \left(\frac{x^j}{j!} \right). \tag{1.155}$$

Therefore, for any matrix M, define e^M by:

$$e^M = \sum_{j=1}^{\infty} \left(\frac{M^j}{j!} \right). \tag{1.156}$$

I note that, in practice, this is not how anybody finds the exponential of a matrix. Instead, start from the diagonalization of the matrix to be exponentiated:

$$M = V D_M V^{-1}. \tag{1.157}$$

Then take the above Taylor series of this new form of M:

$$e^M = \sum_{j=1}^{\infty} \left(\frac{\left[V D_M V^{-1} \right]^j}{j!} \right) = \sum_{j=1}^{\infty} \left(\frac{V \left[D_M V^{-1} V \right]^j V^{-1}}{j!} \right) = V \sum_{j=1}^{\infty} \left(\frac{\left[D_M \right]^j}{j!} \right) V^{-1} = V e^{D_M} V^{-1}. \tag{1.158}$$

This final expression of e^M is much easier to compute: the exponential of a diagonal matrix - which D_M is - is easily done by taking the exponential of each diagonal element.

7 Special Functions

7.1 Hermite Polynomials

The Hermite polynomials are an infinite set of polynomials indexed by a parameter $n \in \mathbb{N}$. They are defined recursively as in Def. 7.1.

Definition 7.1: Hermite Polynomials

Define the nth Hermite polynomial, $H_n(x)$, by:

$$H_n(x) = \left(2x - \frac{d}{dx} \right)^n 1. \tag{1.159}$$

Therefore, the first few Hermite polynomials can be calculated as:

- $H_0 = 1$,

- $H_1 = 2x$,
- $H_2 = 4x^2 - 2$,
- $H_3 = 8x^3 - 12x$,
- $H_4 = 16x^4 - 48x^2 + 12$,
- $H_5 = 32x^5 - 160x^3 + 120x$.

I then derive our first simple result about Hermite polynomials.

Lemma 2. $H_{n+1}(x) = 2xH_n(x) - H_n'(x)$

Proof. Using the definition of the nth Hermite polynomial - in Def. 7.1:

$$H_{n+1}(x) = \left(2x - \frac{d}{dx}\right)^{n+1} 1 = \left(2x - \frac{d}{dx}\right)\left(2x - \frac{d}{dx}\right)^{n} 1 = \left(2x - \frac{d}{dx}\right) H_n(x) \tag{1.160}$$

$$= 2xH_n(x) - H_n'(x) \tag{1.161}$$

□

Lemma 2 is used immediately in Theorem 7.1.

Theorem 7.2: Derivative of Hermite Polynomials

For any $n \in \mathbb{N}$,

$$H_{n+1}'(x) = 2\big(n+1\big)H_n(x) \tag{1.162}$$

Proof for Theorem.

I proceed by induction. For the base case, it is easy to see that this formula holds when $n = 0$. To begin the inductive step, assume that the formula holds for all $k \leq n$. Then consider,

$$H_{k+1}'(x) = \frac{d}{dx}\left[\left(2x - \frac{d}{dx}\right)\left(2x - \frac{d}{dx}\right)^{n} 1\right] = \frac{d}{dx}\left[\left(2x - \frac{d}{dx}\right) H_k(x)\right] \tag{1.163}$$

$$= \frac{d}{dx}\Big[2xH_k(x)\Big] - \frac{d^2}{dx^2}\Big[H_k(x)\Big] = 2H_k(x) + 2xH_k'(x) - H_k''(x). \tag{1.164}$$

Using the assumption that the formula holds for H_k:

$$H_{k+1}'(x) = 2H_k(x) + 4xkH_k'(x) - 4k(k-1)H_{k-1}(x) \tag{1.165}$$

$$= 2H_k(x) + 2k\big(2xH_k'(x) - 2(k-1)H_{k-1}(x)\big). \tag{1.166}$$

The application of Lemma 2 then allows this to be rewritten as:

$$H_{k+1}'(x) = 2\big(k+1\big)H_k(x). \tag{1.167}$$

Therefore, by induction, the formula holds for all $n \in \mathbb{N}$. ■

I then conclude with a final theorem that will be useful later.

Theorem 7.3

$\forall n \in \mathbb{N}^{\geq 2}$,

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H'_{n-2}(x) \tag{1.168}$$

Proof for Theorem.

Using Lemma 2, $H_{n+1}(x) = 2xH_n(x) - H'_n(x)$. Then using Theorem 7.1:

$$H_{n+1}(x) = 2xH_n(x) - H'_n(x) = 2xH_n(x) - 2nH'_{n-1}(x) \tag{1.169}$$

$$\iff H_n(x) = 2xH_{n-1}(x) - H'_{n-1}(x) = 2xH_{n-1}(x) - 2(n-1)H'_{n-2}(x) \tag{1.170}$$

CHAPTER 2

A FIRST LOOK AT QUANTUM MECHANICS AND WHY WE CAN TREAT WAVE FUNCTIONS AS VECTORS

Our initial look at quantum mechanics takes as very limited postulates:

- For any system at any time, there exists a function - known as a wavefunction - completely describing the system.
- The state of the system evolves according to Schrodinger's Equation (given later).
- The probability of measuring a particle as being within a given region is the integral - over that region - of the square of the absolute value of the wavefunction.

1 Schrodinger's Equation

Definition 1.1: A Hamiltonian

A Hamiltonian is an operator corresponding to the total energy of the system. When applied to any state with a definite energy, it maps that state to itself but multiplied by the corresponding energy.

Definition 1.2: The Hamiltonian of a Particle

The Hamiltonian considering to a particle in a potential, $V(x)$, is:

$$\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \hat{V}(x) \tag{2.1}$$

Definition 1.3: Schrödinger equation

The Schrödinger equation is:

$$i\hbar\frac{d\psi}{dt} = \hat{H}\psi, \tag{2.2}$$

where $\hbar = \frac{6.62607015\times10^{-34}}{2\pi} kg\ m^2 s^{-1}$ (the exactness follows from kilograms being defined in terms of this constant) and is called the reduced Planck constant.

What I have given as the Schrödinger equation is often known as the time dependent Schrödinger equation. Another useful equation - known as the time independent Schrödinger equation - may be derived from it.

Theorem 1.4: the Time Independent Schrödinger Equation

For any quantum system with Hamiltonian, $\hat{H}$, for any stationary state of the system, $\psi_n(x)$, with energy, $E_n \in \mathbb{R}$, satisfies:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{n,x}(x)}{dx^2} + \hat{V}(x)\psi_{n,x}(x) = E_n\psi_{n,x}(x) \tag{2.3}$$

Proof for Theorem.

Start from the time dependent Schrödinger Equation and look for solutions of the form

$\psi_n(x,t) = \psi_{n,x}(x) \cdot \psi_{n,t}(t)$:

$$i\hbar\frac{d\psi_n}{dt} = \hat{H}\psi_n \tag{2.4}$$

$$\iff i\hbar\frac{d}{dt}\left(\psi_{n,t}\psi_{n,x}(x)\right) = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\left(\psi_{n,x}(x)\cdot\psi_{n,t}(t)\right) + \hat{V}(x)\psi_{n,x}(x)\cdot\psi_{n,t}(t) \tag{2.5}$$

$$\iff i\hbar\psi_{n,x}(x)\frac{d\psi_{n,t}}{dt} = -\frac{\hbar^2}{2m}\psi_{n,t}(t)\frac{d^2\psi_{n,x}(x)}{dx^2} + \hat{V}(x)\psi_{n,x}(x)\psi_{n,t}(t) \tag{2.6}$$

$$\iff \frac{i\hbar}{\psi_{n,t}(t)}\frac{d\psi_{n,t}}{dt} = -\frac{\hbar^2}{2m}\frac{1}{\psi_{n,x}(x)}\frac{d^2\psi_{n,x}(x)}{dx^2} + \hat{V}(x). \tag{2.7}$$

I see that I have achieved a separation of variables. To have this equation hold in all places at all times, I requires that both sides of the equation equate to a constant. I - suspiciously - call this constant E_n. Therefore, I now have two differential equations to solve:

$$\frac{i\hbar}{\psi_{n,t}(t)}\frac{d\psi_{n,t}}{dt} = E_n \tag{2.8}$$

$$-\frac{\hbar^2}{2m}\frac{1}{\psi_{n,x}(x)}\frac{d^2\psi_{n,x}(x)}{dx^2} + \hat{V}(x) = E_n. \tag{2.9}$$

Fortunately, the first of these is in a form I can easily solve. A fact more easily seen if the two differential equations are expressed as:

$$\frac{d\psi_{n,t}}{dt} = \frac{-iE_n}{\hbar}\psi_{n,t}(t) \tag{2.10}$$

$$-\frac{\hbar^2}{2m}\frac{d^2\psi_{n,x}(x)}{dx^2} + \hat{V}(x)\psi_{n,x}(x) = E_n\psi_{n,x}(x). \tag{2.11}$$

The second differential equation is the Time-Independent Schrodinger Equation (TISE) I've been aiming for. It provides a way to derive the stationary states (the states which the system will stay in under the time evolution governed by $\hat{H}$) of system being considered. With the stationary state labelled $\psi_{n,x}(x)$ having energy E_n. ■

2 Examples Showcasing Quantum Effects

Our first particular examples to examine quantum mechanics consist of a single free particle in various potentials, in a single dimension. Each of the following examples are determined by and differ only in the potentials a particle is placed in.

2.1 Superposition: Infinite Square Well

Our first - slightly unphysical - example is defined by a potential that is zero within a specified one-dimensional region and infinite everywhere else. This results in the particle remaining trapped within this specified region. We formally define the problem to be solved

as examining the behaviour of a particle within the potential:

$$V(x) = \begin{cases} 0, & \text{if } 0 \leq x \leq a, \\ \infty, & \text{otherwise} \end{cases}. \tag{2.12}$$

Examining the TISE within the square well (where $V(x) = 0$):

$$\hat{H}\psi = -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} \Rightarrow -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = E\psi \Rightarrow \frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2}\psi. \tag{2.13}$$

For convenience, let $k^2 = \dfrac{2mE}{\hbar^2}$. So the above differential equation has the solution:

$$\psi(x) = A\cos(kx) + B\sin(kx), \tag{2.14}$$

where A and B are arbitrary constants. They are decided by the boundary conditions that are imposed. In this case, the boundary conditions come from the wavefunction disappearing (equalling zero) in the regions where $v(x) = \infty$. This is as there is no chance of measuring the particle being there. This translates to the conditions:

- $\psi(0) = 0 \iff A\cos(0) + B\sin(0) = A = 0.$
- $\psi(a) = 0 \iff A\cos(ka) + B\sin(ka) = B\sin(ka) = 0$

Hence, k must take some value such that ka is a multiple of π. However, there are many such values - of k - that satisfy this constraint: *all* are valid. In equations, we need to find the values of k_n such that, for some integer value of n:

$$k_n a = n\pi \iff k_n = \frac{n\pi}{a}. \tag{2.15}$$

Therefore,

$$\psi_n(x) = B_n \sin\left(\frac{n\pi}{a}x\right). \tag{2.16}$$

Note that there are many different solutions, indexed by the integer value n. The values of B are decided by the *normalization condition*. The normalization condition is derived from the fact that the probability of measuring the particle anywhere must be one i.e. $\forall n \in \mathbb{N}$

$$\int_{-\infty}^{\infty} \left(|\psi_n(x)|^2\right) dx = 1. \tag{2.17}$$

If we use the above derived formula for $\psi_n(x)$ in this condition:

$$\int_0^a \left(\left|B_n \sin\left(\frac{n\pi}{a}x\right)\right|^2\right) dx = \left|B_n\right|^2 \int_0^a \left(\left|\sin\left(\frac{n\pi}{a}x\right)\right|^2\right) dx = 1 \tag{2.18}$$

$$\iff B_n = + \sqrt{\frac{1}{\int_0^a \left(\left|\sin\left(\frac{n\pi}{a}x\right)\right|^2\right) dx}}. \tag{2.19}$$

We don't actually need to bother evaluating this right now. What matters is that B_n takes a value that can be evaluated.

Definition 2.1: Superposition of States

An we noted in the chapter on the pre-requisite mathematics, when we considered linear differential equations if two functions are a solution to a homogeneous linear differential equation, then any linear combination (basically any sum of the two of them, each multiplied by some coefficient) is also a solution to that same differential equation. Such solutions are known as superpositions and correspond to a genuine quantum phenomena: that a system can exist in a combination of different classical states, when not being measured.

I note that, within the well, the Schrodinger equation of the particle is a homogeneous linear differential equation, so any combination of found solutions to the Schrodinger equation would also satisfy the Schrodinger equation. This new solution would have to be normalized - so the probabilities still sum to one - but this does not prevent these superposition solutions from existing.

2.2 Entering Classically Forbidden Regions: Finite Square Well

Making our first example a little more physical, we now define a new potential that is zero within the zero-potential region and takes a specific real - generally unspecified, as its exact value is immaterial - value, V_0, everywhere else.

$$V(x) = \begin{cases} V_0 \in \mathbb{R}, & \text{if } -a \leq x \leq a, \\ \infty, & \text{otherwise} \end{cases}. \tag{2.20}$$

Is the particle still trapped within the zero-potential region? Examining the TISE within the square well (where $V(x) = 0$), we derive:

$$\psi(x) = A\cos(kx) + B\sin(kx), \tag{2.21}$$

but by the symmetry of the situation, about $x = 0$, we reason that $B = 0$. We then examine the schrodinger equation within the regions of potential V_0, considering the cases where the energy, E, is less than V_0 (i.e. $E - V_0 < 0$):

$$\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = \big(V_0 - E\big)\psi. \tag{2.22}$$

Letting $k' = \dfrac{\sqrt{2m\big(V_0 - E\big)}}{\hbar}$, this has the solution:

$$\psi(x) = Ce^{k'x} + De^{-k'x}, \tag{2.23}$$

where C and D are - as yet - unknown values. If we impose the condition that infinitely far from the well the probability of finding the particle is zero (which is reasonable), then the overall solution of the wave function is:

$$\psi(x) = \begin{cases} B\cos(kx) & \text{if } -a \leq x \leq a \\ Ce^{k'x} & \text{if } -a > x \\ De^{-k'x} & \text{if } a < x \end{cases}. \tag{2.24}$$

Again we evaluate the values of the coefficients using the boundary conditions, which are now different. What we do is known as wavefunction matching; basically joining up all of the functions within the different regions at the known points the meet. The conditions for doing this are:

- The wavefunction must be continuous everywhere.
- The derivative of the wavefunction must be continuous everywhere.

Therefore, considering matching the wavefunctions at $x = a$, the conditions are:

$$A\cos(ka) = Ce^{-k'a}, \tag{2.25}$$

$$-Ak\sin(ka) = -Ck'e^{-k'a}. \tag{2.26}$$

The interesting thing about satisfying these two equations simultaneously is that it is not possible for all values of k and k'. This is the root of quantisation! Dividing the latter of the above by the former:

$$\frac{A\cos(ka)}{-Ak\sin(ka)} = \frac{Ce^{-k'a}}{-Ck'e^{-k'a}} \iff \frac{\cos(ka)}{k\sin(ka)} = \frac{1}{k'} \iff \tan(ka) = \frac{k'}{k}. \tag{2.27}$$

I do not actually much care about matching the wavefunctions. The interesting thing here is the quantisation. Recall that the values of k and k' are defined in terms of energies. Hence, we can rewrite the above as:

$$\tan\left(\sqrt{2mE}a/\hbar\right) = \frac{\sqrt{2m(V_0 - E)}/\hbar}{\sqrt{2mE}/\hbar} \iff \tan\left(\sqrt{2mE}a/\hbar\right) = \sqrt{\frac{(V_0 - E)}{E}}. \tag{2.28}$$

This is actually a pain to solve analytically so we are far better off attempting to solve this numerically (I'm not actually going to as I can use this as an excuse to use approximations). We can rearrange to make the equation slightly nicer, by letting $x = \sqrt{E}$ and choosing units such that $\hbar = 1$,

$$\tan\left(\sqrt{2m}ax\right) = \sqrt{\frac{V_0}{x^2} - 1}, \tag{2.29}$$

but to solve this numerically we have to choose a particular value of m, the particle's mass, a, the radius of the well, and V_0, the value of the potential outside the region where it is zero. We don't actually care what values they take; this example is merely illustrative, so just set

- $a = 1$,
- $m = \frac{1}{2}$,
- $V_0 = 1$.

This gives reduces the equation to solve to:

$$\tan\left(x\right) = \sqrt{\frac{1}{x^2} - 1}. \tag{2.30}$$

Therefore, $x \approx 0.739085133215161...$ is the first solution but there are infinitely many.

Definition 2.2: Tunnelling

Tunnelling is the phenomena in quantum mechanics where a partice appears where it theoretically can be but should not have been able to get to due to energy barriers. We say the particle has tunnelled through the barrier.

2.3 A Textbook Confinement: Quantum Harmonic Oscillator

The quantum harmonic oscillator is defined as a system of a single particle in a potential:

$$V(x) = \frac{1}{2}qx^2, \tag{2.31}$$

where q is some real, positive value characterizing the exact potential. Therefore, q can be expressed as $m\omega^2$, where m is the particle's mass and ω is an arbitrary value - for now - that makes the equivalence to q true. Then consider the Shrodinger equation of this system:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \frac{1}{2}m\omega^2x^2\psi = E\psi. \tag{2.32}$$

To make life easier, I define a new function[1], ϕ, by:

$$\phi(x) = \psi(x\sqrt{m\omega/\hbar}). \tag{2.33}$$

A further simplification comes from defining a change of variables:

$$y = x\sqrt{m\omega/\hbar}. \tag{2.34}$$

The key consequence of both these transforms is that:

$$\frac{d^2\psi}{dx^2} = \frac{m\omega}{\hbar}\frac{d^2\phi}{dx^2}. \tag{2.35}$$

[1] which is just a rescaling of the wavefunction

Using these results in the Schrodinger equation - and re-expressing E as $\frac{\hbar\omega}{2}\epsilon$ gives:

$$-\frac{\hbar^2}{2m}\frac{m\omega}{\hbar}\frac{d^2}{dy^2}\Big(\phi(y)\Big)+\frac{1}{2}m\omega^2\frac{y^2}{m\omega/\hbar}\phi(y)=\frac{\hbar\omega}{2}\epsilon\phi(y) \tag{2.36}$$

$$\Rightarrow -\frac{\hbar\omega}{2}\frac{d^2}{dy^2}\Big(\phi(y)\Big)+\frac{\hbar\omega}{2}y^2\phi(y)=\frac{\hbar\omega}{2}\epsilon\phi(y) \tag{2.37}$$

$$\Rightarrow -\frac{d^2}{dy^2}\Big(\phi(y)\Big)+y^2\phi(y)=\epsilon\phi(y) \tag{2.38}$$

I then seek to solve this differential equation with the ansatz: $f(y)e^{-y^2/2}$. Substituting this into the Schrodinger equation:

$$-\frac{d^2}{dy^2}\Big(f(y)e^{-y^2/2}\Big)+y^2f(y)e^{-y^2/2}=\epsilon f(y)e^{-y^2/2} \tag{2.39}$$

$$\iff -\Big(f''(y)e^{-y^2/2}-yf'(y)e^{-y^2/2}-f(y)e^{-y^2/2}-yf'(y)e^{-y^2/2}+y^2f(y)e^{-y^2/2}\Big) \tag{2.40}$$

$$+y^2f(y)e^{-y^2/2}=\epsilon f(y)e^{-y^2/2} \tag{2.41}$$

$$\iff \Big(f''(y)-2yf'(y)+\big(\epsilon-1\big)f(y)\Big)e^{-y^2/2}=0 \tag{2.42}$$

$$\iff f''(y)-2yf'(y)+\big(\epsilon-1\big)f(y)=0. \tag{2.43}$$

I then turn to a purely mathematical result, using earlier derived properties of the Hermite polynomials.

Theorem 2.3: Solution to a Differential Equation

The differential equation:

$$f''(y)-2yf'(y)+\big(\epsilon-1\big)f(y)=0 \tag{2.44}$$

is solved by $H_n(y)$ if $\epsilon=2n+1$.

Proof for Theorem.

Substituting the claimed solution into the differential equation gives:

$$H_n''(y)-2yH_n'(y)+\big(\epsilon-1\big)H_n(y)=0. \tag{2.45}$$

Using Theorem 7.1, this can be re-written as:

$$4n(n-1)H_{n-2}(y)-4ynH_{n-1}(y)+\big(\epsilon-1\big)H_n(y)=0 \tag{2.46}$$

$$\iff 2n\Big(2(n-1)H_{n-2}(y)-2yH_{n-1}(y)\Big)+\big(\epsilon-1\big)H_n(y)=0. \tag{2.47}$$

Using Theorem 7.1, this reduces to:

$$-2nH_n(y)+\big(\epsilon-1\big)H_n(y)=0 \tag{2.48}$$

$$\iff \big(\epsilon-(2n+1)\big)H_n(y)=0. \tag{2.49}$$

Then the assumption that $\epsilon = 2n + 1$ implies that this final equation is true. ■

I note that if $\epsilon \neq 2n + 1$ then the differential equation has no solution that tends to zero at $\pm\infty$ as we require[2]. So then the solution to the problem of finding the wavefunctions is:

$$\phi(y) = H_n(y)e^{-y^2/2}. \tag{2.50}$$

All that remains to do is undo the substitutions that were made to simplify the problem:

$$\psi(y\sqrt{m\omega/\hbar})_n = C'_n H_n(y)e^{-y^2/2} \Rightarrow \psi_n\big(x\big) = C_n H_n\left(\frac{x}{\sqrt{m\omega/\hbar}}\right)e^{-x^2 m\omega/2\hbar}, \tag{2.51}$$

where C_n and C'_n are normalization coefficients. I have not explicitly calculated a full wavefunction yet so I fell like I owe you one. I'll calculate the two lowest energy states.

Lowest Energy State

The lowest energy state of a system is often called its ground state. The ground state is:

$$\psi_0\big(x\big) = C_0 H_0\left(\frac{x}{\sqrt{m\omega/\hbar}}\right)e^{-x^2 m\omega/2\hbar} = C_0 e^{-x^2 m\omega/2\hbar}. \tag{2.52}$$

I find C_0 by imposing the normalization condition (using the same substitution as before):

$$1 = \int_{-\infty}^{\infty}\left(\big|\psi(x)\big|^2\right)dx = C_0^2\int_{-\infty}^{\infty}\left(e^{-x^2 m\omega/2\hbar}\right)dx = C_0^2\sqrt{\hbar/m\omega}\int_{-\infty}^{\infty}\left(e^{-y^2}\right)dy \tag{2.53}$$

Fortunately, the integral at the end of the above equation is a well known integral, known as the Gaussian integral. It is known to equate to $\sqrt{\pi}$. Therefore,

$$1 = C_0^2\sqrt{\hbar/m\omega}\sqrt{\pi} \tag{2.54}$$

$$\Rightarrow C_0 = \left(\frac{m\omega}{\hbar\pi}\right)^{1/4}. \tag{2.55}$$

Therefore, the wavefunction of the ground state of the quantum harmonic oscillator is:

$$\psi_0\big(x\big) = \left(\frac{m\omega}{\hbar\pi}\right)^{1/4} e^{-x^2 m\omega/2\hbar}. \tag{2.56}$$

The First Excited State

The first excited state has the wavefunction:

$$\psi_1\big(x\big) = C_1 H_1\left(\frac{x}{\sqrt{m\omega/\hbar}}\right)e^{-x^2 m\omega/2\hbar} = C_1\frac{2x}{\sqrt{m\omega/\hbar}}e^{-x^2 m\omega/2\hbar}. \tag{2.57}$$

[2]Which allows for the total probability to be finite.

Then applying the normalization condition and absorbing a factor of $\frac{2}{m\omega/\hbar}$ into C_1^2, for convenience:

$$1 = \int_{-\infty}^{\infty} \left(C_1^2 x^2 e^{-x^2 m\omega/\hbar} \right) dx = C_1^2 \int_{-\infty}^{\infty} \left(x^2 e^{-x^2 m\omega/\hbar} \right) dx. \tag{2.58}$$

Using another known solution of an integral:

$$1 = C_1^2 \int_{-\infty}^{\infty} \left(x^2 e^{-x^2 m\omega/\hbar} \right) dx = C_1^2 \frac{\sqrt{\pi}}{2\left(m\omega/\hbar\right)^{3/2}} \tag{2.59}$$

$$\Rightarrow C_1^2 = \frac{2\left(m\omega/2\hbar\right)^{3/2}}{\sqrt{\pi}} = \left(4\frac{m^3\omega^3}{\hbar^3\pi} \right)^{1/2} \Rightarrow C_1 = \left(4\frac{m^3\omega^3}{\hbar^3\pi} \right)^{1/4}. \tag{2.60}$$

Therefore, the wavefunction of the first excited state is:

$$\psi_1(x) = \left(4\frac{m^3\omega^3}{\hbar^3\pi} \right)^{1/4} x e^{-x^2 m\omega/2\hbar} \tag{2.61}$$

3 Practical Usage: Using Wells as Qubits

I now present the first instance of practically using quantum phenomena, by examining what is a quantum computer at its most basic level. The core component is a quantum version of a bit. We will go into more detail about what a bit is later, but for now just consider a qubit to be a quantum object with exactly two states. We can build a qubit - effectively - by considering only the two lowest energy states of a physical system. Then, consider two of the harmonic potential wells side by side. Assume that they are far enough apart and the potentials are steep enough that the probability of the particle in one oscillator tunnelling to the other is negligible. So how to encode two states - importantly with the same energy - in this system?

First, isolate the system so the overall energy it contains is completely fixed. Then define one state, that we denote $|0\rangle$ as one of the oscillators being in the first excited state and the other being in the ground state. Define the other state, which we denote as $|1\rangle$ as the other oscillator being the one in the excited state.

Due to the isolation, there are exactly two states that the system can be in.

Note the notation we have use to denote the two states. Each has full wavefunctions that we could easily write down - and in fact, already know due to the above - but this notation - known as Dirac notation - is more convenient. And all the properties of the state can be derived from the known wavefunction as required.

4 Setting Up Treating Wavefunctions as Vectors

In the immediately preceding section, we saw that the state of some systems, such as the qubit we constructed, can be represented by objects such as $|0\rangle$ and $|1\rangle$. In fact these two objects are vectors and so any state of a two-dimensional system can be represented as a vector of the appropriate size. We can then use matrices to apply operations, time evoltions, and measurements to them.

CHAPTER 3

INTRODUCTION TO QUANTUM MECHANICS AS LINEAR ALGEBRA

1 Classical Systems in a Convenient Formalism

First, consider a classical system that can be in any of n_l states. I index each possible state from 1 to n_l and label each of those states as $|b_j\rangle$, where j is the index of the states. I can then define a set, $\mathcal{H}_{class}$, containing all possible states of the classical system i.e. $\mathcal{H}_{class} = \{|b_j\rangle | 1 \leq j \leq n_l\}$.[1]
A good example is a set of n_s switches (e.g. light switches), then each of the possible states, i.e. $|b_j\rangle$, is defined by a particular combination of each of the n_s switches being on or off (e.g. if $n_s = 2$, $|b_1\rangle$ is the state where the first switch is off and the second switch is on). Therefore, the state of the system can be completely described by giving the index of the state, for example, $|b_k\rangle \in \mathcal{H}_{class}$, that the set of switches are in. Setting this example aside for a second, an important subset of classical systems is when there are only two possible states of the system, i.e. $|\mathcal{H}_{class}| = 2$, then I term such a system a bit.

Definition 1.1

A bit is a system with exactly two possible states.

For bits, it is conventional to label the two possible states $|0\rangle$ and $|1\rangle$.

In the earlier example about switches, each individual switch, in isolation, can be considered as a bit, as it can be either on or off. Therefore having two states.
This gives a convenient way to look at $\mathcal{H}_{class}^{switches}$ (the set of all states of n_s switches): each state in $\mathcal{H}_{class}^{switches}$ consists of a single choice from each of the possible states of each switch.
More mathematically and concretely, indexing the switches and letting $\mathcal{H}_j^{single}$ be the set of states for a single switch (with index j). Then, using the labelling convention, mentioned above, for the states of a bit (which each individual switch can be considered as):

$$\mathcal{H}_j^{single} = \{|0_j\rangle, |1_j\rangle\}, \tag{3.1}$$

where the j inside the bracket denotes which qubit the state refers to. This allows $\mathcal{H}_{class}^{switches}$ to be written as the tensor product of these single switch sets ($\mathcal{H}_j^{single}$):

$$\mathcal{H}_{class}^{switches} = \bigotimes_{j=1}^{n_s} \left(\mathcal{H}_j^{single} \right). \tag{3.2}$$

This then implies that the size of $\mathcal{H}_{class}^{switches}$ is $|\mathcal{H}_{class}^{switches}| = 2^{n_s}$. The final component of the classical systems to consider is that there exists a function of the system's state that returns the energy of the system when in that state, referred to as the Hamiltonian.

Definition 1.2

A classical Hamiltonian is a function from the set of possible states of the corresponding system to $\mathbb{R}$, that returns the energy of the input state of the system.

[1] This presentation of states is known as braket or Dirac notation: any state, in this notation, is denoted as $|label\rangle$, where *label* is the unique identifier / "name" of the state.

2 Quantum Systems in the Same Formalism

I can then expand the set of possible states of the system, to form the state space of a quantum system, by expanding the set of possible states into a vector space over the field of complex numbers. This vector space is determined by being the span of the set of classical states considered before, subject to a normalisation condition [2].

Definition 2.1

For a system that can be in a superposition of the basis states[a] $\{|b_j\rangle | 1 \leq j \leq n_l\}$, i.e. for any state of the system, $|\psi\rangle$, such that:

$$|\psi\rangle = \sum_{j=1}^{n_l} \Big(\alpha_j |b_j\rangle\Big), \tag{3.3}$$

where $\forall |b_j\rangle \in \{|b_j\rangle | 1 \leq j \leq n_l\}$, $\alpha_j \in \mathbb{C}$.
The normalisation condition, followed by all quantum systems, is:

$$\sum_{j=1}^{n_l} \Big(|\alpha_j|^2\Big) = 1. \tag{3.4}$$

[a] A basis state is the element of a set such that any state of the system can be expressed as a linear combination - with complex coefficients - of elements in the basis set. The set is known as the basis.

So the set of possible states of a single quantum bit (a qubit) is the set:

$$\big\{\alpha_0|0\rangle + \alpha_1|1\rangle \big| \alpha_0, \alpha_1 \in \mathbb{C} \textit{ such that } |\alpha_0|^2 + |\alpha_1|^2 = 1\big\}, \tag{3.5}$$

where a qubit is defined as in Definition 2.

Definition 2.2

A qubit is a quantum system with two basis states[a].

[a] A set of basis states of a quantum is a set such that every state the system could be in is expressible as a linear combination (with complex coefficients) of basis states i.e. the set of possible states of the system is the span of the basis states.

States that are not a basis state of this vector space but still in the vector space, are said to be in a superposition. Though, in fact, the term superposition is always defined relative to a basis and no basis has a particular claim to being *the* basis.
As this vector space also admits an inner product that defines a metric on the vector space and is complete, it can also be referred to as a Hilbert space and is often referred to as *the* Hilbert space of the system.

[2] This normalization condition is a consequence of the Born rule (which governs measurement in quantum systems and is defined in Definition 4) and the requirement for the probabilities each measurement outcome to always sum to one.

Definition 2.3

A Hilbert space is a real or complex inner product space that is a complete metric space with respect to the metric induced by the inner product. For the purposes of this thesis, the Hilbert space of a quantum system can just be considered as the set of all possible states (including superpositions) that the quantum system being considered can be in.

For a single qubit - as the set of possible states of a quantum system can be considered as the span of the possible states of the corresponding classical system - its set of possible states, i.e. its Hilbert space, follows from Eqn. (3.1) and can be expressed as:

$$\mathcal{H}_j^{\textit{single, quantum}} = \mathit{Span}\{|0\rangle_j, |1\rangle_j\}. \tag{3.6}$$

Returning to the example of n_s switches, each switch's two states in the classical system become two basis states - as shown in Eqn. (3.6). Each switch's Hilbert space is then defined as the Hilbert space spanned by the two basis states, subject to the normalisation condition (i.e. Definition 2).
For the whole system of n_s switches, considered as a quantum system, the Hilbert space, $\mathcal{H}_{quantum}^{switches}$, is the tensor product of the Hilbert space of each switch (as the tensor product of two Hilbert spaces can be considered as the Hilbert space of their respective systems combined into a single system), analogously to Eqn. (3.2) i.e.

$$\mathcal{H}_{quantum}^{switches} = \bigotimes_{j=1}^{n_s} \left(\mathcal{H}_j^{\textit{single, quantum}} \right). \tag{3.7}$$

For example, if there are two switches (i.e. $n_s = 2$), the Hilbert space of the two switches - considered as a single system - is:

$$\mathit{Span}\big(|0\rangle_1 \otimes |0\rangle_2, |0\rangle_1 \otimes |1\rangle_2, |1\rangle_1 \otimes |0\rangle_2, |1\rangle_1 \otimes |1\rangle_2\big), \tag{3.8}$$

i.e any state of the two qubits can be expressed as:

$$\alpha_{00}|0\rangle_1 \otimes |0\rangle_2 + \alpha_{01}|0\rangle_1 \otimes |1\rangle_2 + \alpha_{10}|1\rangle_1 \otimes |0\rangle_2 + \alpha_{11}|1\rangle_1 \otimes |1\rangle_2, \tag{3.9}$$

where each $\alpha_{j,k}$ is a complex number such that $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$. It is worth noting that expressing states as formally and fully as $|0\rangle_1 \otimes |0\rangle_2$ is quite a laborious notation and so, for example, it is quite common - both in this thesis and more widely - to express this state as $|0\rangle_1|0\rangle_2$ - omitting the tensor product symbol - or, going even further to be concise, as $|00\rangle$. I note that any choice of four orthogonal states in the Hilbert space of two qubits can be used to express any state as in Eqn. 3.9. This particular choice is known as the computational basis and may be extended to any number - $N \in \mathbb{N}$ - qubits as the N-fold tensor product of elements in $\{|0\rangle, |1\rangle\}$

In the quantum setting, the Hamiltonian, which was a function for classical systems, is upgraded to an operator (defined in next section) and for that reason I suspend a full description of quantum Hamiltonians until the next section.

3 Operators and Time Evolutions in Quantum Mechanics

Most simply, operators are things that do something to states of the system. Mathematically, I define them, factoring in the requirements of quantum mechanics, as linear maps from the system under consideration's Hilbert space onto itself.
An operator, $\hat{A}$, (that acts on a Hilbert space, $\mathcal{H}$) is said to be linear if, for any states, $|\psi\rangle, |\phi\rangle \in \mathcal{H}$; and $\alpha_\phi, \alpha_\psi \in \mathbb{C}$:

$$\hat{A}\Big(\alpha_\psi|\psi\rangle + \alpha_\phi|\phi\rangle\Big) = \alpha_\psi\hat{A}|\psi\rangle + \alpha_\phi\hat{A}|\phi\rangle. \tag{3.10}$$

For ease of "bookkeeping," I denote the set of all operators on a given Hilbert space, $\mathcal{H}$, by $\mathbb{B}\big(\mathcal{H}\big)$.

The most important operator for any quantum system is its Hamiltonian: in moving from a classical to a quantum system, the Hamiltonian is "upgraded" to a Hermitian operator so it can now act on states. I define the Hamiltonian to map an energy basis state (eigenstate of the Hamiltonian) to itself multiplied by it's energy, a real value (so then the energy is an eigenvalue of the Hamiltonian).
Defining the eigenstates (ensuring they form an orthogonal basis of the Hilbert space) and eigenvalues uniquely defines the Hamiltonian (up to physically irrelevant variations).

Definition 3.1

The Hamiltonian of a quantum system is a Hermitian operator such that the eigenstates of the Hamiltonian are the stationary states (i.e. the states such that a system in that state does not change - except for an irrelevant overall phase - with time) of the system, and the corresponding eigenvalues are the energy of the respective stationary states.

A large part of the centrality of a system's Hamiltonian is its role in the system's equation of motion. However, there is no short, simple derivation of the governing equation for the dynamics of quantum systems. The dynamics were observed experimentally and the justification for the Schrödinger equation, presented below, is its good agreement with experiments. For any quantum system with Hamiltonian, $\hat{H}$, if $|\psi\rangle$ represents the state of the system, the Schrödinger equation is:

$$i\hbar\frac{d|\psi\rangle}{dt} = \hat{H}|\psi\rangle, \tag{3.11}$$

where $\hbar = \frac{6.62607015\times10^{-34}}{2\pi} kg\ m^2 s^{-1}$ and is called the reduced Planck constant. But, because theorists are too cool to write out long numbers, I use "God-given" units for the rest of this thesis, that allows me to set $\hbar = 1$.
For my purposes, it is more convenient to derive an operator that applies the time evolution for a specified amount of time, rather than solve the Schrödinger equation every time. To be clear, I seek an operator, $\hat{U}_t$, such that, for any $t \in \mathbb{R}$, if $|\psi(0)\rangle$ is an arbitrary state at

an arbitrary time and $|\psi(t)\rangle$ is the result of evolving that state for time t under the relevant Hamiltonian:

$$|\psi(t)\rangle = \hat{U}_t|\psi(0)\rangle. \tag{3.12}$$

To find such an operator, I start from the Schrödinger equation (Eqn. (3.11)) and derive:

$$\left(i\frac{d}{dt} - \hat{H}\right)|\psi(t)\rangle = 0 \iff \left(i\frac{d}{dt} - \hat{H}\right)\hat{U}_t|\psi(0)\rangle = 0. \tag{3.13}$$

Then, as this must hold for any choice of initial state, $|\psi(0)\rangle$, I require that:

$$\left(i\frac{d}{dt} - \hat{H}\right)\hat{U}_t = 0. \tag{3.14}$$

This is a differential equation of a known, standard form and so the correct solution can easily be seen but, for completeness, I check that the solution $\hat{U}_t = e^{-i\hat{H}t}$ is correct:

$$\left(i\frac{d}{dt} - \hat{H}\right)e^{-i\hat{H}t} = \hat{H}e^{-i\hat{H}t} - \hat{H}e^{-i\hat{H}t} = 0. \tag{3.15}$$

Likewise, any multiple of $\hat{U}_t$ is a solution by the same argument as the above. As the Hamiltonian is defined to be Hermitian, this form of the time evolution operator, $\hat{U}_t$, means that time evolutions, for any time period, t, are applied by unitary operators. I.e. allowing the time evolution for any length of time, according to any Hermitian Hamiltonian, will apply a unitary operator. This will be important later.

A final important operator is the identity. The identity applied to a state or system represents doing nothing to it. It is denoted $\hat{I}$.
Sometimes when the aim is to represent doing nothing to a sub-part, S_{sub} of a system, the identity (on that sub-part only) is represented by $I_{S_{\text{sub}}}$. Though occasionally the subscript is also omitted.

4 Measurement

It is necessary to start with a definition of what can be observed, before the process of observing it can be considered.

Definition 4.1

In quantum mechanics, an observable is a quantity that can be measured. It is represented by a Hermitian matrix.

Like was the case for the Schrödinger equation, the rules for measurement cannot be derived even if given the Schrödinger equation. So, based on experimental observation, I assert the Born rule Born, "Zur Quantenmechanik der StoSSvorgänge" (as in Definition 4) to govern measurement outcomes in quantum mechanics:

Definition 4.2

For any observable with matrix representation, A, if that observable is measured, the outcome will be an eigenvalue of A. Additionally, for a system in the state $|\psi\rangle$, the probability of measuring the eigenvalue, λ, is $\langle\psi|\hat{\Pi}_\lambda|\psi\rangle$, where $\hat{\Pi}_\lambda$ is the projector into the eigenspace of λ. This is known as the Born rule.

5 From Dirac Notation to Density Matrices

While Dirac notation (the style of quantum mechanics constructed so far) functions perfectly well, it does have the unfortunate habit of treating quantum and classical probability separately. Meaning if I want to consider both simultaneously (and mixing them can be advantageous), I would have to manually handle the classical probabilities. This could quickly become onerous. So re-consider the previously established mechanics to handle quantum and classical probability on the same footing.
The fundamental hurdle to this is the Born rule. So I need some way to re-express quantum states (and how operators act on them) such that quantum (via the Born rule) and classical probabilities are managed similarly.
In Dirac notation, for any state, $|\psi\rangle = \sum_{j=1}^{n_l}\left(\alpha_j|b_j\rangle\right)$ (where $\{|b_j\rangle\}_{j=1}^{n_l}$ is an orthonormal basis), the probability of measuring a given basis state[3], $|b_k\rangle$, is:

$$\left|\langle b_k|\psi\rangle\right|^2 = \left|\sum_{j=1}^{n_l}\left(\alpha_j\langle b_k|b_j\rangle\right)\right|^2 = \left|\alpha_k\right|^2. \tag{3.16}$$

Inspired by this, let $|\psi\rangle$ be any state from the previous formalism. In the new formalism, the same state would be represented by:

$$|\psi\rangle\langle\psi|, \tag{3.17}$$

which, considering the state $|\psi\rangle$ as a vector[4], allows us to consider $|\psi\rangle\langle\psi|$ as a matrix. Call this a density matrix, which is generally denoted as ρ. This means that a state in an equal superposition of the measurement basis states $|b_1\rangle\langle b_1|$ and $|b_2\rangle\langle b_2|$, would be represented as:

$$\frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2| + \frac{1}{2}|\psi_1\rangle\langle\psi_2| + \frac{1}{2}|\psi_2\rangle\langle\psi_1|. \tag{3.18}$$

In this case, the probability of measuring each basis states is $\frac{1}{2}$.
Similarly, if there is (classical) uncertainty about which state a system is in I can just sum the probabilities multiplied by the relevant density matrix. e.g. if there is an equal, classical probability of the state $|\psi_1\rangle\langle\psi_1|$ or $|\psi_2\rangle\langle\psi_2|$ being prepared, the system can be considered to be in the state:

$$\frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2|. \tag{3.19}$$

[3] when performing a measurement that projects into one of the basis states.
[4] and $\langle\psi|$ is the Hermitian conjugate of $|\psi\rangle$.

This has the nice benefit that the probability of measuring each basis state is the coefficient of that state's density matrix (i.e. $|\psi_1\rangle\langle\psi_1|$ or $|\psi_2\rangle\langle\psi_2|$) in the state's representation in Eqn. (3.19). Therefore it is easy so see (in Eqn. (3.18) and Eqn. (3.19)) that classical uncertainties and superpositions are represented on an equal footing. Each is represented in the coefficients of the density matrices (e.g. $|\psi_1\rangle\langle\psi_1|$ or $|\psi_1\rangle\langle\psi_2|$) that sum to give the density matrix representing the state.
As a technicality, this new formalism does not represent the set of all states of a system as a Hilbert space. Instead note that density matrices can be viewed as operators (as they are matrices and hence must represent linear operators on a corresponding vector space, decided by the matrix size and the field the matrices are over) that act on the Hilbert space of the system (the same system but in the old formalism). This set of density matrices is exactly the set of valid linear operators, $\mathbb{B}(\mathcal{H})$, on the Hilbert space of the old formalism.

I now present the application of operators to density matrices. For this, assume the aim is to perform the equivalent of (in the old formalism) applying the operator $\hat{U}$ to the state $|\psi\rangle$.
It is easy to see how this is done by defining the state $|\phi\rangle = \hat{U}|\psi\rangle$ and equating the density matrix representations of the two states:

$$|\phi\rangle\langle\phi| = \hat{U}|\psi\rangle\langle\psi|\hat{U}^\dagger. \tag{3.20}$$

It is therefore simple to conclude that for any operator, $\hat{U}$, and any density matrix, ρ, I apply the operator as:

$$\hat{U}\rho\hat{U}^\dagger. \tag{3.21}$$

The one outstanding element of density matrices left to cover is measurement. In the density matrix formalism, a measurement is defined via a set of projection operators (with the condition they project into orthogonal subspaces). Denote the projector corresponding to each outcome measurement, α, as $\hat{\Pi}_\alpha$.
$\hat{\Pi}_\alpha$ projects any state into the subspace defined by any state in that subspace having the property such that the quantity measured by the measurement takes the value α i.e. any state in the subspace would give measurement outcome α if measured.
Then the probability of getting outcome α from the measurement (on ρ) is:

$$Prob\big(measuring\ \alpha\big) = \mathrm{Tr}\bigg(\hat{\Pi}_\alpha\rho\bigg). \tag{3.22}$$

The state of the system after the measurement is the state the projectors project it into, but normalised:

$$\rho' = \frac{\hat{\Pi}_\alpha\rho\hat{\Pi}_\alpha^\dagger}{\mathrm{Tr}\big(\hat{\Pi}_\alpha\rho\hat{\Pi}_\alpha^\dagger\big)} = \frac{\hat{\Pi}_\alpha\rho\hat{\Pi}_\alpha^\dagger}{\mathrm{Tr}\big(\hat{\Pi}_\alpha^\dagger\hat{\Pi}_\alpha\rho\big)}. \tag{3.23}$$

Definition 5.1

While density matrices are more general than Dirac notation, a fair fraction of the time they just mean all gates / time evolutions / have to be written out twice. I therefore reserve the right to use either Dirac notation or density matrices depending on which is convenient.

A more general notion of measurement that that presented above comes from Positive Operator Valued Measures (POVMs) defined in Definition 5. This notion of measurement will not be used in considerations of quantum systems or when designing protocols/algorithms but can be used for considering more general measurements in proofs.

Definition 5.1

A Positive Operator Valued Measure (POVM) is a set, $\{\hat{E}_i\}_i$, of positive semi-definite Hermitian operators such that:

$$\sum_i \left(\hat{E}_i\right) = \hat{I}. \tag{3.24}$$

Each element of the set, $\hat{E}_i$, corresponds to a measurement outcome, indexed by i, and the probability of that outcome is:

$$\mathrm{Tr}\left(\hat{E}_i \rho\right), \tag{3.25}$$

where ρ is a density matrix the measurement is being performed on.

CHAPTER 4

BASIC QUANTUM COMPUTING

As the first line of every quantum computing paper tells us, quantum computers promise to radically transform what is computationally possible. Quantum computing was originally conceived of via the quantum Turing machine, but this is unwieldy for practical or even theoretical usage and overly abstract for algorithm design.
An alternative, and yet still universal Deutsch, "Quantum Theory, the ChurchTuring principle and the Universal Quantum Computer", approach is the circuit model.

1 Components of a Quantum Computation in the Circuit Model

In brief, a quantum computation is executed (in the circuit model), by preparing a number of qubits in specific states, applying operations to those qubits, and then measuring the qubits. During this computation, the qubits may experience quantum effects e.g. being entangled with each other and being in superpositions; giving rise to the expected greater computational power of quantum computers. It is worth pausing to consider each component of the circuit model.

1.1 Wires

The basis of the circuit model is that qubits are represented by wires (i.e. horizontal lines), with each qubit represented by a single wire. The flow of time is then represented by moving rightward along each wire: the same distance along different wires represents the same point in time for different qubits. A trivial example of a wire is given in Fig. 4.1. Fig. 4.1 is not

---

Figure 4.1: Example of a single wire, representing a single qubit. This represents nothing happening to the qubit.

the most exciting figure, but I'll add each component of an interesting circuit to Fig. 4.1 as it is introduced.

1.2 State Preparation

This is the initial step in the circuit model and consists of taking the required number of qubits from unknown, arbitrary states (as, in general, you're unaware of what state the qubits were used for before you started using the computer) and preparing them in a known, specified (by the algorithm being executed) state. Typically, this is a tensor product of single-qubit gates. In the circuit model, the state each qubit is prepared in is shown to the left of the left-most end of the wire (representing the earliest point in time of the qubit represented by the wire), as in Fig. 4.2.

Figure 4.2: Example circuit of a single wire, representing a single qubit prepared in the state $|0\rangle$ and then left.

1.3 Gates / Operations

The representation of the operations applied to qubits in the course of the circuit is via gates. These are defined in Definition 1.3.

Definition 1.1

A gate depicts an operation on a set of qubits, with each qubit depicted by both an input wire (representing the qubit before the operation) and an output wire (representing each qubit after the operation). For an example of a gate see Fig. 4.3, which has input wires on the left and output wires on the right.

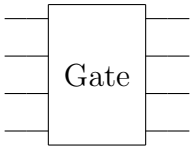


Figure 4.3: Example gate

These gates are physically applied via the time evolution of a specific Hamiltonian, for a specific duration, to produce the intended gate. For this reason, due to reasons discussed in Sec. 3, all gates appearing in a circuit must be unitary.

An important aspect of the application of operations to a qubit, and hence the application of gates to wires, is denoting which qubits an operator is applied on and in what order the operations are performed. The first point is conveyed by which wires a gate acts on; the second is conveyed by the left-most gates being applied first, followed by the next left-most. So the order gates are applied follows the representation of the flow of time in the wires.

In light of this, for a circuit to make sense, gates are required to be time-wise connected, as defined in Definition 1.3.

Definition 1.2

Two gates are said to be time-wise connected if no output wire of one connects to an output wire of the other or itself, and no input wire of one connects to an input wire of the other or itself. For example see Fig. 4.4, which features two copies of the gate in Fig. 4.3 connected time-wise.

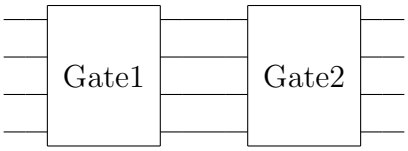


Figure 4.4: Two time-wise connected gates, representing the operations represented by the two gates applied sequentially, progressing from left to right.

Example gates can be added to the example circuit developed in Fig. 4.1 and Fig. 4.2, and this is shown in Fig. 4.5.

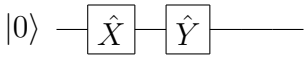


Figure 4.5: Example circuit of a single wire, representing a single qubit prepared in the state $|0\rangle$, followed by the application of a Pauli $\hat{X}$ gate then a Pauli $\hat{Y}$ gate.

Typically, a physical quantum device is restricted to applying a finite number of gates, by the physics of how it applies gates usually. The set of all possible gates on a given device is known as the gateset. Some gatesets are known to be 'universal,' meaning all efficiently implementable circuits can be expressed using only gates from the gateset.

1.4 Measurement

The final element of a circuit is the measurements. These provide the results of a circuit, and do exactly as the name suggests. She measurements of a circuit are the last thing to happen to their respective qubits and, as such, are represented by 'D's at the end of a wire, as in Fig. 4.6. A measurement is applied, typically, to a single qubit, though not all qubits are

required to be measured. Any wire in a circuit that just ends abruptly can be disregarded past that point. It is not measured or considered any further.

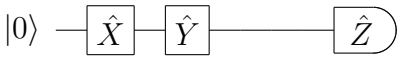


Figure 4.6: Example circuit of a single wire, representing a single qubit prepared in the state $|0\rangle$, followed by the application of a Pauli $\hat{X}$ gate then a Pauli $\hat{Y}$ gate, and then measurement in the Pauli $\hat{Z}$ basis.

2 Commonly Used Gates

There are some gates that must be presented for use in this thesis. I stick to a few, well-worn, and commonly used gates. The full characterization of these gates follows from their matrices, provided in the tables below: the single-qubit gates are shown in Table. 4.7 and the two qubit gates shown in Table. 4.8.

Any other gates used are denoted by their operator, as defined, being contained in a gate.

Some gates can be form groups, with specific properties, two such sets that will continually reappear are the Pauli gates and the Clifford gates.

Definition 2.1

Define $\mathbb{P}_1$ as the set $\{\hat{I}, \hat{X}, \hat{Y}, \hat{Z}\}$. Where $\hat{I}$ is the identity on a single qubit and $\hat{X}, \hat{Y}, \hat{Z}$ are the single-qubit Pauli operators.
Similarly, $\mathbb{P}_1^{\otimes N}$ denotes the set of operators defined by a single element of $\mathbb{P}_1$ acting on each of N qubits (the elements of $\mathbb{P}_1$ acting on different qubits can be different).

Clifford gates were originally defined in Ref. Gottesman, "Theory of Fault-Tolerant Quantum Computation" with a different definition but, for my purposes, the following suffices:

Definition 2.2

A two-qubit gate, $\hat{\mathcal{M}}$, is defined to be Clifford if $\forall \hat{x} \in \mathbb{P}_1^{\otimes 2}$,

$$\hat{\mathcal{M}}^\dagger \hat{x} \hat{\mathcal{M}} \in \mathbb{P}_1^{\otimes 2}. \tag{4.1}$$

I will sometimes refer to this as two-qubit Clifford gates normalising $\mathbb{P}_1^{\otimes 2}$.

Gate Name	Circuit Depiction	Matrix Representation
Identity	$\hat{I}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Pauli X	$\hat{X}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Pauli Y	$\hat{Y}$	$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$
Pauli Z	$\hat{Z}$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
Hadamard	H	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
Phase	$\hat{S}$	$\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$
Inverse Phase	$\hat{S}^\dagger$	$\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$
T	$\hat{T}$	$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$
$R_x(\theta)$	$\hat{R}_x(\theta)$ or $e^{-i\hat{X}\theta/2}$	$\begin{pmatrix} \cos(\theta/2) & -i\sin(\theta/2) \\ -i\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}$
$R_z(\theta)$	$\hat{R}_z(\theta)$ or $e^{-i\hat{Z}\theta/2}$	$\begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}$
Arbitrary Gate	$\hat{U}$	$\begin{pmatrix} U_{1,1} & U_{1,2} \\ U_{2,1} & U_{2,2} \end{pmatrix}$

Table 1.1: A table of the names, circuit depictions, and matrix representations (in the basis $\{|0\rangle, |1\rangle\}$) of commonly used single-qubit gates. Note that the Hadamard gate is denoted as H, *without a hat*, as opposed to Hamiltonians which will always be written with a hat. $\theta \in [0, 2\pi]$ is a parameter that specifies a particular gate in a family of gates.

Figure 4.7: Table of Common Single-Qubit Gates

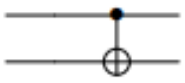
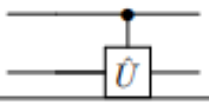
Gate Name	Circuit Depiction	Matrix Representation
controlled-X / CNOT / $c\hat{X}$		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$
controlled-Z / $c\hat{Z}$		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$
controlled-$\hat{U}$ / $c\hat{U}$		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & U_{1,1} & U_{1,2} \\ 0 & 0 & U_{2,1} & U_{2,2} \end{pmatrix}$

Table 1.2: A table of the names, circuit depictions, and matrix representations (in the basis $\{|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle\}$) of commonly used two-qubit gates. Note that $\hat{U}$ can be any single-qubit gate.

Figure 4.8: Table of Common Two-Qubit Gates

CHAPTER 5

NOISE IN QUANTUM SYSTEMS

Start with idea system is actually a lot larger than we would like. and is an open system and then look at how we represent this with

1 A Background on CPTP Maps

The "gold standard" way to represent error in quantum operations is that an erroneous implementation of an operator/measurement/state preparation is equivalent to the ideal/errorless operation followed by a completely positive trace-preserving map. A such, Completely Positive trace-preserving (CPTP) maps are a key notion in any consideration of error. These maps, as the name suggests, are any map from the system under consideration to itself that is completely positive and trace-preserving. I pause here to introduce these maps and their important properties that will be used later in this chapter.

Definition 1.1

Any linear operator, A, acting on any Hilbert space is positive if, for any element, $|x\rangle$, in the Hilbert space:

$$\langle x|A|x\rangle \geq 0. \tag{5.1}$$

Definition 1.2

For any Hilbert space, $\mathcal{H}$, any map, $\Phi : \mathbb{B}\big(\mathcal{H}\big) \to \mathbb{B}\big(\mathcal{H}\big)$, is positive if positive operators are mapped exclusively to positive operators. If, $\forall N \in \mathbb{N}$, $\Phi \otimes I_N$ (where I_N is the identity on N qubits) is positive, then Φ is completely positive.

Definition 1.3

For any Hilbert space, $\mathcal{H}$, a map, Φ, is trace-preserving if $\forall \rho \in \mathbb{B}(\mathcal{H})$,

$$\mathrm{Tr}\Big[\Phi\big(\rho\big)\Big] = \mathrm{Tr}\Big(\rho\Big). \tag{5.2}$$

For completeness, once completely-positive and trace-preserving maps have been defined, I formally define CPTP maps.

Definition 1.4

A completely positive trace-preserving map between matrices (or, via their representations, states) is a map that is :
- Completely Positive.
- trace-preserving.

A central theorem of use in presenting the following properties of CPTP maps is the Stinespring Dilation Theorem. The most quantum-information-friendly background and proof of the Stinespring Dilation Theorem (to my knowledge) can be found in Ref. Watrous, "Basic Notions of Quantum Information", following from the proof of Theorem 2.22.

Theorem 1.5: Stinespring Dilation Theorem (Special Case)

Let $\mathcal{H}_A$ be a Hilbert space, then for any completely-positive trace-preserving map, $\Phi : \mathbb{B}\big(\mathcal{H}_A\big) \to \mathbb{B}\big(\mathcal{H}_A\big)$, there exists: another Hilbert space, $\mathcal{H}_C$; an element of that space, $|\phi\rangle \in \mathcal{H}_C$; and a unitary, $\hat{\mathcal{U}}_{AC} : \mathcal{H}_A \otimes \mathcal{H}_C \to \mathcal{H}_A \otimes \mathcal{H}_C$, such that, $\forall \rho_A \in \mathbb{B}(\mathcal{H}_A)$:

$$\Phi\big(\rho_A\big) = \mathrm{Tr}_c\Big[\hat{\mathcal{U}}_{AC} \circ \Big(\rho_A \otimes |\phi\rangle\langle\phi|\Big) \circ \hat{\mathcal{U}}_{AC}^{\dagger}\Big], \tag{5.3}$$

Where $\circ$ denotes operator composition but is additionally overloaded to be matrix multiplication between matrices.

Proof for Theorem.

Proof omitted ∎

A first application of Theorem 1 is in Lemma 3, which details and proves the existence of a useful form for considering CPTP maps.

Lemma 3. *For any Hilbert space, $\mathcal{H}$, any CPTP map, Φ, acting on $\mathbb{B}\big(\mathcal{H}\big)$, has Kraus decomposition i.e. can be expressed as:*

$$\Phi\big(\rho\big) = \sum_{j=1} \Big(\hat{\mathcal{K}}_j \circ \rho \circ \hat{\mathcal{K}}_j^{\dagger}\Big), \tag{5.4}$$

Where each $\hat{\mathcal{K}}_j$ is a linear operator on $\mathcal{H}$ (called a Kraus operator), and,

$$\sum_{j=1}\left(\hat{\mathcal{K}}_j^\dagger \circ \hat{\mathcal{K}}_j\right) = \hat{I}. \tag{5.5}$$

Note that the number of Kraus operators is always finite.

Proof. Via the Stinespring Dilation Theorem, any CPTP map, $\Phi : \mathbb{B}\big(\mathcal{H}\big) \to \mathbb{B}\big(\mathcal{H}\big)$, can be expressed as: $\forall \rho \in \mathbb{B}\big(\mathcal{H}\big)$

$$\Phi\big(\rho\big) = \mathrm{Tr}_c\Big[\hat{\mathcal{U}}_{AC} \circ \Big(\rho \otimes |\phi\rangle\langle\phi|\Big) \circ \hat{\mathcal{U}}_{AC}^\dagger\Big]. \tag{5.6}$$

Let $\big\{|e_j\rangle\big\}_{j=1}^{N_E}$ be a basis of the extra Hilbert space added (denoted $\mathcal{H}_c$ in the Stinespring Dilation Theorem). Then $\Phi\big(\rho\big)$ can be re-expressed as:

$$\Phi\big(\rho\big) = \sum_{j=1}^{N_E}\Big[\big(\hat{I}_A \otimes \langle e_j|\big) \circ \hat{\mathcal{U}}_{AC} \circ \big(\rho \otimes |\phi\rangle\langle\phi|\big) \circ \hat{\mathcal{U}}_{AC}^\dagger \circ \big(\hat{I}_A \otimes |e_j\rangle\big)\Big]. \tag{5.7}$$

Defining $\hat{\mathcal{K}}_j$ as $\big(\hat{I}_A \otimes \langle e_j|\big) \circ \hat{\mathcal{U}}_{AC} \circ \big(I_A \otimes |\phi\rangle\big)$. Therefore,

$$\Phi\big(\rho\big) = \sum_{j=1}^{N_E}\left(\hat{\mathcal{K}}_j \circ \rho \circ \hat{\mathcal{K}}_j^\dagger\right). \tag{5.8}$$

All that then remains is to show that the condition in Eqn. (5.5) is satisfied:

$$\sum_{j=1}\left(\hat{\mathcal{K}}_j^\dagger \circ \hat{\mathcal{K}}_j\right) = \sum_{j=1}\Big[\big(I_A \otimes \langle\phi|\big) \circ \hat{\mathcal{U}}_{AC}^\dagger \circ \big(I_A \otimes |e_j\rangle\big) \circ \big(I_A \otimes \langle e_j|\big) \circ \hat{\mathcal{U}}_{AC} \circ \big(I_A \otimes |\phi\rangle\big)\Big] \tag{5.9}$$

$$= \big(\hat{I}_A \otimes \langle\phi|\big) \circ \hat{\mathcal{U}}_{AC}^\dagger \circ \sum_{j=1}\left(\hat{I}_A \otimes |e_j\rangle\langle e_j|\right) \circ \hat{\mathcal{U}}_{AC} \circ \big(\hat{I}_A \otimes |\phi\rangle\big) = \hat{I}. \tag{5.10}$$

□

The converse of Lemma 3 also holds, as shown in Lemma 4.

Lemma 4. *For any Hilbert space, $\mathcal{H}$, and any map, Φ acting on $\mathbb{B}\big(\mathcal{H}\big)$, of the form,*

$$\Phi(\rho) = \sum_{j=1}\left(\hat{\mathcal{K}}_j \circ \rho \circ \hat{\mathcal{K}}_j^\dagger\right), \tag{5.11}$$

where each $\hat{\mathcal{K}}_j$ is a linear operator on $\mathcal{H}$, and,

$$\sum_{j=1}\left(\hat{\mathcal{K}}_j^\dagger \circ \hat{\mathcal{K}}_j\right) = \hat{I}. \tag{5.12}$$

(i.e. ϕ has a Kraus decomposition) is a completely positive trace-preserving map.

Proof. I need to show that the map in Eqn. (5.11) is both trace-preserving and completely positive. Denote this map by Φ.
Trace-preserving

$$\mathrm{Tr}\Big[\Phi\big(\rho\big)\Big] = \mathrm{Tr}\Big[\sum_{j=1}\Big(\hat{\mathcal{K}}_j \circ \rho \circ \hat{\mathcal{K}}_j^\dagger\Big)\Big] = \sum_{j=1}\Big[\mathrm{Tr}\Big(\hat{\mathcal{K}}_j^\dagger \circ \hat{\mathcal{K}}_j \circ \rho\Big)\Big] = \mathrm{Tr}\Big[\sum_{j=1}\Big(\hat{\mathcal{K}}_j^\dagger \circ \hat{\mathcal{K}}_j\Big) \circ \rho\Big] \tag{5.13}$$

$$= \mathrm{Tr}\Big(\rho\Big). \tag{5.14}$$

Hence the trace remains unchanged after the application of Φ.
Completely-Positive
Let $\mathcal{H}'$ be an arbitrary Hilber space of dimension $n \in \mathbb{N}$ and $|x\rangle \in \mathcal{H} \otimes \mathcal{H}'$, then:

$$\langle x|\big[\Phi\big(\rho\big) \otimes \hat{I}_n\big]|x\rangle = \langle x|\sum_{j=1}\Big[\big(\hat{\mathcal{K}}_j \otimes \hat{I}_n\big) \circ \big(\rho \otimes \hat{I}_n\big) \circ \big(\hat{\mathcal{K}}_j^\dagger \otimes \hat{I}_n\big)\Big]|x\rangle \tag{5.15}$$

$$= \sum_{j=1}\Big\{\langle x|\big[\big(\hat{\mathcal{K}}_j \otimes \hat{I}_n\big) \circ \big(\rho \otimes \hat{I}_n\big) \circ \big(\hat{\mathcal{K}}_j^\dagger \otimes \hat{I}_n\big)\big]|x\rangle\Big\} = \sum_{j=1}\Big[\langle x_j|\big(\rho \otimes \hat{I}_n\big)|x_j\rangle\Big], \tag{5.16}$$

where $|x_j\rangle$ is defined as $\big(\hat{\mathcal{K}}_j^\dagger \otimes \hat{I}_n\big)|x\rangle$ and is still in $\mathcal{H} \otimes \mathcal{H}'$. I then note that ρ being positive implies that $\rho \otimes \hat{I}_n$ is positive, hence:

$$\langle x|\big[\Phi\big(\rho\big) \otimes \hat{I}_n\big]|x\rangle = \sum_{j=1}\Big[\langle x_j|\big(\rho \otimes \hat{I}_n\big)|x_j\rangle\Big] \geq 0. \tag{5.17}$$

So Φ is completely positive. Therefore Φ is a completely positive, trace-preserving map, a CPTP map. □

2 A Characterisation of Error

In this thesis, error in an operation will always be modelled as CPTP maps following or preceding the correct implementation of the intended (ideal) operation. There may be some other conditions on the error, but it will always be modelled as CPTP maps.

The characterisation of all error as being CPTP error is a strong claim. This can be argued for by considering all error that could affect the system as being the result of a unitary (as all changes in a known closed system are representable by a unitary operator) acting on the entire universe, including the system. However the state of the universe is not known and so, from our perspective, is stochastic. Thus, the actual error can be considered as consisting of different unitaries applied with varying probabilities and therefore can be characterised by the map:

$$\mathcal{E}\big(\rho\big) = \sum_{k=1}\Big(p_k \hat{\mathcal{U}}_k \rho \hat{\mathcal{U}}_k^\dagger\Big), \tag{5.18}$$

where,
• $\mathcal{E}(\,\cdot\,)$ is the CPTP map representing the error acting on a density matrix (given as the argument).
• Each $\hat{\mathcal{U}}_k$ is a unitary that acts on both the system and environment, there is a unknown number of them so the summation above has no upper limit.
• p_k is a probability such that $\sum_{k=1}(p_k) = 1$.
Therefore, by Lemma 4, the error applied can be modelled as a CPTP map acting on both the system and its environment. Alternatively, I can require certain things of post-error states and use this to imply the error map is a CPTP map. These requirements are:
Let ρ be any density matrix of the system the error acts on, and let $\mathcal{E}$ be the error map,

1. $\mathcal{E}(\rho)$ can be treated identically to a density matrix for measurement purposes.
2. The outcome probabilities for any measurement on $\mathcal{E}(\rho)$ are positive.
3. The outcome probabilities for any measurement on $\mathcal{E}(\rho)$ sum to one.

Theorem 2.1

The above requirements imply that $\mathcal{E}$ is a CPTP map.

Proof for Theorem.

For a measurement of the state of ρ (in some orthogonal basis) e.g. measuring a single qubit in the Z-basis; let $|\alpha\rangle$ be any output state, Π_α be a projector onto that state, $[\alpha]$ be the set of all possible measurement outcomes and $\mathbb{P}_\alpha$ be the probability of measuring that state.
$\mathcal{E}$ is completely positive
Let ρ' be the density matrix of any state of any system in which the system $\mathcal{E}$ acts on is a subsystem, similarly let α' be any possible outcome of any measurement on the entire system.
Requirement 2 means $\mathbb{P}_{\alpha'} \geq 0$, therefore, for any measurement outcome, α':

$$0 \leq \mathrm{Tr}\Big[\Pi_{\alpha'}(\hat{I}\otimes\mathcal{E})(\rho')\Big] = \sum_{b'\in[\alpha']}\Big[\langle b'|\alpha'\rangle\langle\alpha'|(\hat{I}\otimes\mathcal{E})(\rho')|b'\rangle\Big] = \langle\alpha'|(\hat{I}\otimes\mathcal{E})(\rho')|\alpha'\rangle. \quad (5.19)$$

As any density matrix is a positive operator, this implies $\mathcal{E}$ is a completely positive map.
$\mathcal{E}$ is trace-preserving
Requirement 3 means $\sum_\alpha(\mathbb{P}_\alpha) = 1$, therefore:

$$1 = \sum_\alpha\Big[\mathrm{Tr}\Big(\Pi_\alpha\mathcal{E}(\rho)\Big)\Big] = \mathrm{Tr}\Big[\sum_\alpha\Big(\Pi_\alpha\Big)\mathcal{E}(\rho)\Big] = \mathrm{Tr}\Big[\mathcal{E}(\rho)\Big], \quad (5.20)$$

where the last equality follows from the sum of projectors onto disjoint subspaces, the union of which is the entire space, is equivalent to the identity.
As any density matrix also has trace one, $\mathcal{E}$ is trace-preserving. ∎