The Role of Hilbert Space in Formalizing Elementary Quantum Mechanics

Alexander Berenbeim

Abstract

Quantum mechanics rapidly developed as a physical theory after several decades of serious empirical challenges to the model of classical mechanics at the subatomic level. This paper starts with a description of separable Hilbert spaces and basic definitions regarding operator theory. The second section describes the mathematical and experimental developments that led to the separable Hilbert space formulation of quantum mechanics developed by Von Neumann, who reconciled Heisenberg's matrix mechanical with Schrödinger's wave mechanical description of quantum mechanics. The third section examines the probability interpretation of Quantum Mechanics in light of the separable Hilbert space framework, as well as the limitations of that approach. The fourth, and final, section describes two different methods that developed in the 1930's and 1940's around addressing some of these limitations: the von Neumann algebras and Schwarz's field of distributions.

Contents

For Sam

Foreward

This paper grew out of notes for a project for Elcim Elgun's course on Measure Theory and Fourier Analysis, and was mainly written during July and August 2013 while I was away from the University of Waterloo, and attending to my family's affairs after my brother suddenly passed away in June.

The target audience of this paper is deliberately limited: these are private notes meant for my friends and for my professor Elcim Elgun. Some of the examples provided are from my course notes from Brian Greene's Quantum Mechanics sequence and do not reflect my own work. These examples appear starting in section 2.4. Some of the examples prior to this section were left as exercises in [8], but were found to be suitable for the exposition of the ideas in this paper.

Again, many thanks need to go out to both Brian Greene for introducing me to most of this material, and my professor Elcim Elgun, who was both incredibly kind and patient with me when I needed to withdraw from my studies for family reasons.

"*Nobody, except topologists, is interested in problems about Hilbert space; the people who work in Hilbert space are interested in problems about operators.*"- Paul Halmos, "Ten Problems in Hilbert Space"

1 Mathematical Preliminaries: Hilbert Spaces, Isomorphisms, and Operators

1.1 There's Always Room In Hilbert Space

Although David Hilbert was the first to investigate these extraordinary mathematical spaces that generalize Euclidean space and extend vector algebra to infinite dimensional cases, the term *Hilbert space* was coined by John von Neumann. It was von Neumann, whose pioneering research in quantum mechanics made extraordinary use of the properties which define Hilbert spaces, who saw their wide ranging applications to signal processing, operator theory, partial differential equations, engineering, economics, and many of the other areas of mathematics where von Neumann saw fit to write a paper. Formally speaking, we define them as follows:

Definition. A *Hilbert space* $\mathcal{H}$ is a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product $\langle , \rangle$.
Furthermore, the inner product satisfies the following properties

1. The inner product of an element with itself is positive definite

$$\langle x, x \rangle \geq 0$$

2. The inner product of a pair of elements is equal to the complex conjugate of the swapped elements

$$\langle x, y \rangle = \overline{\langle y, x \rangle}$$

3. The inner product is linear in its first argument

$$\langle \alpha x_1 + \beta x_2, y \rangle = \alpha \langle x_1, y \rangle + \beta \langle x_2, y \rangle$$

It follows that a complex inner product is antilinear in its second argument

Proof.

$$\langle x, \alpha y_1 + \beta y_2 \rangle = \overline{\langle \alpha y_1 + \beta y_2, \rangle} = \bar{\alpha}\overline{\langle y_1, x \rangle} + \bar{\beta}\overline{\langle y_2, x \rangle} = \bar{\alpha}\langle x, y_1 \rangle + \bar{\beta}\langle x, y_2 \rangle$$

and it follows a real inner product is bilinear.
Finally, the distance function is given by the norm $\|x\| = \sqrt{\langle x, x \rangle}$ by

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle}$$

□

One remarkable property of Hilbert spaces is that they enable geometric arguments. In fact, their geometric character is what intrigued von Neumann to apply them to Quantum Mechanics and inspired

his search for a continuous geometric formulation of quantum theory. With geometry in mind, the Cauchy-Schwarz inequality is of great importance:

Lemma 1. ***(Cauchy-Schwarz)****For all elements x and y of an inner product space*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

where equality occurs if and only if x and y are linearly dependent.

Proof. Consider arbitrary x, y in our inner product space over a field k, where k is either the field $\mathbb{R}$ or $\mathbb{C}$. Without loss of generality, if $y = 0$, it is clear that we have equality. Now assuming $x, y \neq 0$, we define

$$z = x - \frac{\langle x, y\rangle}{\langle y, y\rangle} y$$

And by the linearity of the inner product in the first argument, we find

$$\langle z, y\rangle = \left\langle x - \frac{\langle x, y\rangle}{\langle y, y\rangle} y, y \right\rangle = \langle x, y\rangle - \frac{\langle x, y\rangle}{\langle y, y\rangle}\langle y, y\rangle = 0$$

so z is a vector orthogonal to y. So now we have

$$x = z + \frac{\langle x, y\rangle}{\langle y, y\rangle} y$$

and thus

$$\|x\|^2 = \left|\frac{\langle x, y\rangle}{\langle y, y\rangle}\right|^2 \|y\|^2 + \|z\|^2 = \frac{|\langle x, y\rangle|^2}{\|y\|^2} + \|z\|^2 \geq \frac{|\langle x, y\rangle|^2}{\|y\|^2}$$

Then, multiplying both sides of the inequality by $\|y\|^2$

$$|\langle x, y\rangle|^2 \leq \|x\|^2 \|y\|^2$$

Moreover, equality is held when $\|z\|^2 = 0$, which, given the definition of z, occurs only when y is linearly dependent on x. □

Corollary 2. *The inner product between two vectors can be geometrically characterized by*

$$\langle x, y\rangle = \|x\|\|y\| \cos\theta$$

When applied to the inner product space of square-integrable complex functions, we see the inequality as

$$\left|\int_\Omega f\bar{g}\right|^2 \leq \int_\Omega |f|^2 \int_\Omega |g|^2$$

which is a generalization of Hölder's inequality.

Example. One important example of a Hilbert space is the sequence space

$$\ell_2(\mathbb{Z}) = \{\mathbf{x} = \{x_n\} \in \mathbb{C}^{\mathbb{Z}} : \sum_{n\in\mathbb{Z}} |x_n|^2 < \infty\}$$

and the inner product is defined for two sequences $\mathbf{x}, \mathbf{y}$ as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{-\infty}^{\infty} x_n \overline{y_n}.$$

Although we leave it to the reader to verify that this satisfies the definition of a Hilbert space, we take care to note that if $\|c\|_2 < \infty$, it can be "normalized" to 1. As we shall see, this normalization can apply to all such normed values.

1.2 Why There Is Always Room In The Hilbert Hotel: Countable Orthonormal Bases

One of the critical aspects of linear algebra which is generalized in Hilbert spaces is the notion of an orthonormal basis. This concept has wide ranging physical applications, of which this paper will try to make clear. Formally

Definition. An orthonormal basis is a family $\{e_i\}_{i \in I} \subset \mathcal{H}$ for some index I which satisfy the following conditions

1. Orthogonality: For $i \neq j \in I$, $\langle e_i, e_j \rangle = 0$
2. Normalization: Every element $i \in I$ has norm 1: $\|e_i\| = 1$
3. Completeness: $\overline{\text{Span}(\{e_i\})} = \mathcal{H}$

In a sense, these first two requirements can be given by the following condition:

$$\forall i, j \in I \qquad \langle e_i, e_j \rangle = \delta_{ij}$$

while the third condition, in the finite case can be conceived of as certainly saying

$$\forall x \in \mathcal{H} \exists \{c_i\} \subset k \qquad x = \sum^{n} c_i e_i$$

As an inner product space complete under a metric, Hilbert spaces are clearly topological spaces. However, there are a real diversity of Hilbert spaces concerning countability properties. Recall:

Definition. A topological space is called *separable* if it contains a countable dense subset.

We leave it to the reader to verify that every separable Hilbert space has an orthonormal basis[1].

Lemma 3. ***Linear Approximation Lemma*** *Suppose that $\{e_1, \ldots, e_n\}$ is an orthonormal set in an inner product space $(\chi, \langle, \rangle)$. Let $E = Span\{e_1, \ldots, e_n\}$ and define for $f \in \chi$*

$$d(f, E) = \inf \{\|f - g\| : g \in E\}$$

Then

$$d(f, E)^2 = \left\| f - \sum_{i=1}^{n} \langle f, e_i \rangle e_i \right\|^2 = \|f\|^2 - \sum_{i=1}^{n} |\langle f, e_i \rangle|^2$$

Moreover $\sum_{i=1}^{n} \langle f, e_i \rangle e_i$ is the unique vector $g \in E$ such that $\|f - g\| = d(f, E)$

[1] *Hint: This can be verified by taking a linearly independent subsequence of any countably dense sequence of elements from $\mathcal{H}$, and then applying Gram-Schmidt to the sequence.*

Proof. Let $g = \sum_{i=1}^{n} \alpha_i e_i \in E$. Then

$$\begin{aligned}
\|f-g\|^2 &= \langle f-g, f-g\rangle \\
&= \langle f,f\rangle - \langle f,g\rangle - \langle g,f\rangle + \langle g,g\rangle \\
&= \|f\|^2 - 2\mathrm{Re}\langle f,g\rangle + \|g\|^2 \\
&= \|f\|^2 - 2\mathrm{Re}\left(\sum_{i=1}^{n} \overline{\alpha_i}\langle f,e_i\rangle\right) + \sum_{i=1}^{n} |\alpha_i|^2 \\
&\geq \|f\|^2 - 2\sum_{i=1}^{n} |\alpha_i||\langle f,e_i\rangle|^2 + \sum_{i=1}^{n} |\alpha_i|^2 \\
&= \|f\|^2 - \sum_{i=1}^{n} |\langle f,e_i\rangle|^2 + \sum_{i=1}^{n} (|\langle f,e_i\rangle| - |\alpha_i|)^2 \\
&\geq \|f\|^2 - \sum_{i=1}^{n} |\langle f,e_i\rangle|^2
\end{aligned}$$

These inequalities become equalities whenever $\alpha_i = \langle f, e_i\rangle$ for all $1 \leq i \leq n$. Thus if

$$g = \sum_{i=1}^{n} \langle f,e_i\rangle e_i$$

then this g corresponds exactly to

$$\inf\{\|f-h\| : h \in E\} = \|f-g\|.$$

□

Corollary 4. ***(Bessel's Inequality)*** *For $\mathcal{H}$ with orthonormal basis $\{e_i\}$, for $f \in \mathcal{H}$,*

$$\sum_{i=1}^{\infty} |\langle f,e_i\rangle|^2 \leq \|f\|^2$$

In general, we characterize inner product spaces with orthonormal basis by the Orthonormal Basis Theorem, which can be stated as follows:

Theorem 5. ***(Orthonormal Basis Theorem)*** *Let χ be an inner product space and $\{e_i\}_{i=1}^{\infty}$ be an orthonormal sequence. Then the following are equivalent:*

1. *Span$\{e_i\}_{i=1}^{\infty} = \left\{\sum_{i=1}^{n} \alpha_i e_i : n \in \mathbb{N}, \alpha_i \in \mathbb{C}\right\}$ is dense in χ.*
2. *For every $f \in \chi$, we have* Bessel's Equality

$$\|f\|^2 = \sum_{i=1}^{\infty} |\langle f,e_i\rangle|^2$$

3. *For every $f \in \chi$, we have under the $\|\cdot\|$ limit*

$$\lim_{n\to\infty} \sum_{i=1}^{n} \langle f,e_i\rangle e_i = f$$

which can be written as

$$f = \sum_{i=1}^{\infty} \langle f,e_i\rangle e_i$$

4. For every $f, g \in \chi$, we have Parseval's identity

$$\langle f, g\rangle = \sum_{i=1}^{\infty} \langle f, e_i\rangle\langle e_i, g\rangle$$

This leads to an important theorem:

Theorem 6. ***(Abstract Plancheral Theorem)*** *Let χ be an inner product space and let $\{e_n\}_{n\in\mathbb{Z}} \subset \chi$ be an orthonormal basis. Then the operator*

$$U : \chi \to \ell_2$$

given by

$$Uf = \{\langle f, e_n\rangle\}$$

is an isometry. In other words,

$$\|Uf\| = \|f\|$$

and

$$\langle Uf, Ug\rangle = \langle f, g\rangle$$

Proof. For any $f \in \chi$

$$\|Uf\|_2^2 = \sum |\langle f, e_n\rangle|^2 \leq \|f\|^2$$

by Bessel's inequality. It is clear that U is a linear map into ℓ_2.
Next, we see by Parseval's identity that

$$\begin{aligned} \langle Uf, Ug\rangle &= \langle \{\langle f, e_n\rangle\}_{n\in\mathbb{Z}}, \langle g, e_n\rangle\}_{n\in\mathbb{Z}}\rangle \\ &= \sum_n \langle f, e_n\rangle\overline{\langle g, e_n\rangle} \\ &= \langle f, g\rangle \end{aligned}$$

Finally, setting $f = g$, we find our desired result □

Of great consequence are the following results:

Theorem 7. ***(Riesz-Fischer theorem)*** *Let $\{e_n\}$ be an orthonormal basis for an infinite dimensional Hilbert space $\mathcal{H}$. If $\{c_n\}$ is a sequence of numbers such that $\sum |c_n|^2$ converges, then there is an $x \in \mathcal{H}$ such that $x = \sum_{n-1}^{\infty} c_n e_n$ where $c_n = \langle x, e_n\rangle$.*

Corollary 8. *There is an isometry isomorphism between ℓ_2 and any infinite dimensional separable Hilbert space $\mathcal{H}$, and furthermore, any two such Hilbert spaces are isomorphic.*

Proof. (Sketch) Given an orthonormal basis $\{e_n\}_{n\in\mathbb{Z}} \subset \mathcal{H}$, define an operator $T : \ell_2 \to \mathcal{H}$ by

$$T(\{c_n\}) = \sum_n c_n e_n$$

□

1.3 Operator: What's your number?

Operators are ubiquitous in mathematics. To the layman, mathematics is merely arithmetic operators acting on the positive reals, however, the practical importance of more abstract operators cannot be overstated. In classical mechanics, the derivative is the de facto operator, and we are to show that linear operators are of crucial importance in the study of quantum mechanics.

Definition. Let U and V be two normed vector spaces. A linear operator $\hat{A} : U \to V$ is a *bounded operator* if there exists a $M > 0$ such that

$$\|\hat{A}\mathbf{x}\|_V \leq M\|\mathbf{x}\|_U$$

for all $\mathbf{x} \in U$.
We see that bounded operators form their own vector space, with a norm compatible with the norms of U and V, where the norm is defined

$$\|\hat{A}\| := \inf\{M : \|\hat{A}\mathbf{x}\|_V \leq M\|\mathbf{x}\|_U\}$$

In the case where we're dealing with endomorphic operators from U to U, this reduces to the familiar looking

$$\|\hat{A}\hat{B}\| \leq \|\hat{A}\| \cdot \|\hat{B}\|$$

Here, we notice that we can identify linear operators in our Hilbert space as linear transformations. Now, recall the following definitions which should be covered in an introductory abstract algebra course:

Definition. In an arbitrary vector space, a scalar λ is an *eigenvalue* and a nonzero *vector* x is an *eigenvector* of a linear transformation A if

$$Ax = \lambda x.$$

Definition. For a given operator, the set $\{\lambda_i\}$ of eignevalues is called the *spectrum* of A. If a given eigenvalue is a multiple root of the characteristic polynomial (given as $\det(A - \lambda I) = 0$), then the spectrum is said to be *degenerate*.

One particularly important class of operators are self-adjoint operators.

Definition. Let A be a linear transformation on a vector space V. For every A, the operator $A^\dagger$, which satisfies

$$\langle Ax, y\rangle = \langle x, A^\dagger y\rangle \text{ for all } x, y \in V$$

is called the *adjoint* of A.

Definition. If $A = A^\dagger$, then A is its own adjoint, and so A is called a *self-adjoint* operator. In real inner-product spaces, a self-adjoint operator is called a *symmetric* operator, and in complex inner-product spaces, it is called a *Hermitian* operator.

Theorem 9. *If $\hat{A}$ is a self-adjoint linear operator, then every eigenvalue of $\hat{A}$ is real valued.*

Proof. The eigenvalues of $\hat{A}$ are the λ such that $\hat{A}x = \lambda x$ where $x \neq 0$. Thus

$$\lambda^* \langle x, x \rangle = \langle x, \lambda^* x \rangle = \langle x, \hat{A}^\dagger x \rangle = \langle \hat{A}x, x \rangle = \langle \lambda x, x \rangle = \lambda \langle x, x \rangle$$

Since, x is an eigenvector, $x \neq 0$, and hence $\langle x, x \rangle \neq 0$, it follows that $\lambda^* = \lambda$ and so λ is real. □

Corollary 10. *The eigenvalues of a Hermitian operator $\hat{A}$ are real and* $\lambda = \dfrac{\langle x, \hat{A}x \rangle}{\|x\|^2}$

Of great theoretical importance is the result that a unitary diagonalizing matrix can be constructed from any Hermitian matrix, regardless of repeated eigenvalues.

Theorem 11. *Any Hermitian matrix A may be generalized by a unitary similarity transformation; i.e., there exists a unitary matrix U such that $U^{-1}AU$ is a diagonal matrix (the diagonal elements of which are the eigenvalues of A).*

The proof of this theorem can be found in [4]. Theorem 9 can be generalized to an infinite dimensional Hilbert space, but first we will need to consider *projection operators*.

Definition. A *projection operator* on a vector space V is an idempotent linear map $P_n : V \to V$, ie.

$$P_n^2 = P_n$$

P_n projects any vector into the one-dimensional subspace of V spanned by e_n, and any projection is associated with a direct sum decomposition. We can define a projection operator P_n as follows

$$x' = P_n x := \langle e_n, x \rangle e_n$$

where e_n is a unit vector.

Now consider an N-dimensional Hermitian operator $\hat{A}$ with eigenvalues $\{\lambda_n\}$ and eigenvectors $\{e_n\}$, and for convenience, assume the spectrum is non-degenerate.

Notation. Recall that since eigenvectors for a symmetric operator form an orthonormal basis, and $\hat{A}$ is a Hermitian operator, and hence symmetric, we can apply the orthonormal basis theorem to define identity operator $\hat{I}$ as follows:

$$\hat{I}g \equiv \sum_{n=1}^{N} \langle e_n, g \rangle e_n$$

for any $g \in V_N$, where V_N is the finite dimensional vector space. Now if we define our projection operator by

$$\hat{P}_n g = \langle e_n, g \rangle e_n$$

then $\hat{I}$ can be written as

$$\hat{I} = \sum_{n=1}^{N} \hat{P}_n$$

Notation. Moreover, with this formulation, it is clear that

$$\hat{A} = \sum_{n=1}^{N} \lambda_n \hat{P}_n.$$

Definition. *Orthogonal projections* are such that $P_n P_m = \delta_{nm} e_n \langle e_m, \cdot \rangle$. Specifically, in a Hilbert space, an *orthogonal projection* is an idempotent linear map $P_n : \mathcal{H} \to \mathcal{H}$ such that for all $x, y \in \mathcal{H}$

$$\langle P_n x, y \rangle = \langle x, P_n y \rangle$$

We notice that projections are orthogonal when the eigenvectors are orthogonal (which is not always the case!). Again, it is a well-known result of linear algebra that eigenvectors are always orthogonal for symmetric operators (which Hermitian operators are).

Notation. Finally, for spaces X, Y, we introduce the following notation

$$B(X, Y)$$

to denote all bounded, everywhere defined operators from X, Y.

So far we have been looking at bounded operators, but unbounded operators are of immense importance to quantum mechanics. Indeed, the theory of unbounded operators was developed in large part by von Neumann to put quantum mechanics on a rigourous mathematical foundation. However, unbounded operators are fairly difficult objects of study. They are often not part of any standard mathematics curriculum, and although they are ubiquitous in physics, their study is often left to graduate students looking for a challenging paper topic. Simply put,

Definition. An operator $\hat{A}$ is *unbounded* if there are elements of $\mathcal{H}$ on which $\hat{A}$ is <u>not defined</u>.

Finally, we will define the *spectrum* of unbounded operators in order to introduce Stone-von Neumann spectral theory.

First, let $\hat{A}$ be on operator on $\hat{A} : \mathcal{X} \to \mathcal{Y}$, two Banach space.

Definition. An operator $\hat{A}$ is invertible if and only if $\hat{A}^{-1} \in B(\mathcal{Y}, \mathcal{X})$.

Definition. A resolvent set of $\hat{A}$ is given as

$$\text{Res}\hat{A} := \{z \in \mathbb{C} : z\hat{I} - A \text{ is invertible}\}$$

Finally,

Definition. The *spectrum* of an unbounded operator $\hat{A}$ is given as

$$\text{Spec}\hat{A} := \mathbb{C} \backslash \text{Res}\hat{A}$$

It should be clear by this definition of the spectrum, that the set of all eigenvalues of $\hat{A}$, which we call the *point spectrum* and denote with $\text{Spec}_p\hat{A}$ is a subset of the spectrum of $\hat{A}$.

Theorem 12. *Let $T : \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator on a separable Hilbert space with domain Dom(T), which we will denote by D. Then there exists a finite measure space (M, μ), a unitary operator $U : \mathcal{H} \to L_2(M, \mu)$, and a real valued function $f : M \to \mathbb{R}$ such that*

- $\psi \in D \iff f \cdot (U\psi) \in L_2(M, \mu)$

- *If $\psi \in D$, then $U(T(\psi)) = f \cdot U(\psi)$, i.e.*

$$U(T(\psi))(x) = f(x)U(\psi)(x)$$

almost everywhere in M.

When considering orthogonal projections, we can generalize Thm 9, which applies to a finite dimensional vector space, to an infinite-dimensional Hilbert space when considering infinite-dimensional operators of the completely continuous type.

Theorem 13. *To every self-adjoint operator $\hat{A}$ on a Hilbert space $\mathcal{H}$, there corresponds an operator-valued function $E(\lambda)$ such that*

1. $E(\lambda_1)E(\lambda_2) = E(\bar{\lambda})$; $\bar{\lambda} = \min\{\lambda_1, \lambda_2\}$;
2. $\lim_{\lambda \to -\infty} E(\lambda) = 0$, $\lim_{\lambda \to \infty} E(\lambda) = I$;
3. $I = \int_{\mathbb{R}} dE(\lambda)$;
4. $A = \int_{\mathbb{R}} \lambda dE(\lambda)$.

$E(\lambda)$ *is called a* resolution of the identity *belonging to $\hat{A}$. The set of points at which $E(\lambda)$ is non-constant is the* spectrum *of $\hat{A}$. For all λ, $E(\lambda)$ commutes with $\hat{A}$ and with any transformation which commutes with $\hat{A}$.*

A far more in-depth discussion of the spectral theory that plays a central role in classical and quantum mechanics can be found in [4].

To get a rough idea of what this alteration involves, we will consider the *Stieltjes* integral, and then an operator-valued function $E(\lambda)$.

Definition. The *Stieltjes* integral is a generalization of the Riemann integral, which we define for a function f on $[a, b]$ as

$$\int_a^b f(x)dg(x) = \lim_{N \to \infty} \sum_{i=1}^{N} f(x_i^*)[g(x_{i+1}) - g(x_i)]$$

with $x_i \in \mathcal{P}$, where $\mathcal{P}$ is a partition of the interval $[a, b]$. Note that g is not required to be continuous!

Now let us define an operator valued function

$$E(\lambda) = \begin{cases} 0 & \lambda \leq \lambda_1 \\ \sum_{n=1}^{\nu} P_n & \lambda_\nu \leq \lambda < \lambda_{\nu+1} \quad \nu = 1, \ldots, N-1 \\ \sum_{i=1}^{N} P_n & \lambda \geq \lambda_N \end{cases}$$

We see that all four properties of the operator valued function described in the Spectral theorem are satisfied. The following proof is courtesy of [4]

Proof. Let $\{\lambda_n\}$ be the corresponding eigenvalues to the eigenfunctions $\{f_n\}$.

It follows immediately if $\lambda_a < \lambda_1$ that

$$E(\lambda_a)E(\lambda_b) = 0$$

Without loss of generality, let $\lambda_a < \lambda_b < \lambda_N[Fuller]$. So now, consider without loss of generality that $\lambda_1 \leq \lambda_a < \lambda_\nu \leq \lambda_b$. Then,

$$E(\lambda_a)E(\lambda_b) = \sum_{n=1}^{\nu_1} P_n \sum_{m=1}^{\nu_2} P_m = \sum_{n=1}^{\nu_1}\sum_{m=1}^{\nu_2} P_n P_m = \sum_{n=1}^{\nu_1}\sum_{m=1}^{\nu_2} \delta_{nm} e_n \langle e_m, \cdot \rangle = \sum_{n=1}^{\nu_1} P_n = E(\bar{\lambda})$$

2. First, it is clear that

$$\lim_{\lambda \to -\infty} E(\lambda) = 0$$

by definition. Now we also see that for N dimensional vector space V,

$$\lim_{\lambda \to \infty} E(\lambda) = \sum_{n=1}^{N} P_n = I$$

3. Next, we see that the definition of the Stieltjes integral also us to write

$$\int_{-\infty}^{\lambda} dE(\lambda) = \sum_{n=1}^{\nu} P_n$$

if $\lambda \in [\lambda_\nu, \lambda_{\nu+1})$. Since the contributions to the integral come from the points at which $E(\lambda)$ changes discontinuously, by construction at each such point λ_n must be an eigenvalue of A. Thus, in the limit, we pick up each projection operator onto the one-dimensional subspace of V spanned by e_n. In this way

$$I = \int_{-\infty}^{\infty} dE(\lambda)$$

4. This follows immediately from (3) and the formulation that $A = \sum_{i=1}^{N} \lambda P_n$.

□

The virtue of this formulation is that it allows us to imagine the limit $E(\lambda)$ tending to a continuous operator-valued function, and thus to a "continuous spectrum". In practice, since continuum cannot be observed, for those eigenspaces with theoretically continuous eigenvalues, physicists make us of the *Stieltjes* integral.

2 A Formal History of the Development of Quantum Mechanics

"The most incomprehensible thing about the universe is that it is comprehensible."- Albert Einstein, "From Physics to Reality" (1936)

2.1 The Buildup of the Classical World: A Tale of Three Formalisms

Although several schools of philosophy of the ancient Greeks conjectured that the universe fundamentally obeyed mathematical laws, in particular, those of geometry, their surviving descriptions are never formal. Rather, these schools merely speculated about the fundamental character of the universe, likening the physical world to metaphors of flow and fire, cubes and spheres. Plato had the universe nested inside concentric spheres, with the outer most corresponding to a creative intelligence. Even the physics of Aristotle is not quantitative, but qualitative, reducing the world to four causes. For most of human history, it appeared to philosophers that the intelligibility of the world was that which we could describe with natural language, made rigourous only by geometric analogies.

The decisive leap from an Aristotlean description of the physical world, the one which predominated the Western world through the continued intellectual patronage of the Scholastic monasteries and the endorsement of the Catholic church, was brought about by the development of the calculus by Isaac Newton and Gottfried Leibniz. It was Isaac Newton's linking calculus to the study of the physical world and his assertion that the world could be made comprehensible by mathematics, an assertion that arguably has been met with great success, that finally began to lift mathematics above theology and physics above natural philosophy. Although we take for granted now that we can describe the physical world by mathematical functions, it was a profound paradigm shift counterintuitive to the received wisdom of how the world worked at the time.

Moreover, as Einstein himself noted, the most incomprehensible thing about the universe is that the rigourous mathematical description afforded by Newton's approach to physics actually works. Whereas before Newton we had descriptions of objects falling to the earth due to their "essential" nature, we now understand that there is a mathematical relationship between bodies, and this relationship could be understood without appealing to the unknowable essences and substances of a particular body. Before we only had philosophical speculation; now we had quantifiable observation, and with it a framework for predicting the future. The world became comprehensible. The universe became likened to an elaborate clock.

2.1.1 The Dawn of Mathematical Physics: Newtonian Formalism

Although astronomers from antiquity up until the Renaissance relied on geometry to formally describe their observations, the emerging study of analytic geometry added new levels of rigour to their works. For a generation of European researchers, among them Kepler, Hooke, and Isaac Barrow, this new mathematical formalism describing classical geometric concepts developed in tandem with their own physical research. For instance, the relationship between the study of large astronomical bodies, and the instruments which enabled refined study of them fueled the development of optical theory. Specifically, Isaac Barrow's application of

analytic geometry to study the reflection and refraction of light would influence his most renowned pupil, Isaac Newton[2].

Within Newton's *Philosophiae Naturalis Principia Mathematica*, he proposed three laws of motion that completely describe the behaviour of any particle body. These laws were argued for along the lines of an implicit Euclidean geometry, and for over two centuries, Newton's second law:

$$\vec{F} = m\frac{\partial^2}{\partial t^2}\vec{x} = \frac{d}{dt}\vec{p} \tag{2.1}$$

was *the* picture of how the world evolved. That is, Newton proposed that the world was such that the net force acting on an object was the rate of change of its linear momentum[3] in an inertial reference frame.

The second law implies that the evolution of the material world is *fully* determined by the force of motion and initial conditions. As it so happened, classical mechanics which developed from this law, developed alongside mathematical analysis, through the works of Euler, Lagrange, Laplace, and Hamilton, and much empirical observation. For two centuries, this picture corresponded both with observation and what seemed to be our intuition.[1]

In the Newtonian picture of the world, when the force $\vec{F}(\vec{x})$ acting on a body could be expressed in terms of a scalar function $V(\vec{x})$ as $\vec{F}(\vec{x}) = -\nabla V(\vec{x})$, the force is called a *conserved force* and the function V is called the *potential energy*, or simply, the *potential*. This formulation of the energy of a conserved force is given by

$$E = \frac{m}{2}\left(\frac{d}{dt}\vec{x}\right)^2 + V(\vec{x}) = \frac{m}{2}\vec{v}^2 + V(\vec{x}) \tag{2.2}$$

where the function (2.2) describes energy as the sum of kinetic and potential energy, respectively.

When $\vec{F}$ is a conservative force, both mathematically, and for most observable scales, it was found that (2.2) is conserved, as can be seen

$$\frac{dE}{dt} = \sum_i \left(m\frac{dx_i}{dt}\frac{d^2x_i}{dt^2} + \frac{\partial V}{\partial x_i}\frac{dx_i}{dt}\right) = \sum_i \left(m\frac{d^2x_i}{dt^2} + \frac{\partial V}{\partial x_i}\right)\frac{dx_i}{dt} = 0$$

One useful example of a conservative force is the simple harmonic oscillator, where the potential is given as $V(x) = \frac{1}{2}kx^2$ which yields the force $F(x) = -kx$, where k is a constant coefficient.

*A*nother example of the Newtonian formalism is that of a *free* particle. This is a particle where no external force is acting on the particle, and so it's evolution is entirely determined

[2]Although Isaac Barrow was the one who proved the fundamental theorem of calculus, Newton and Leibniz are generally both credited with developing calculus.

[3]In this case, linear momentum is a vector value denoted $\vec{p} = m\vec{v} = m\frac{d}{dt}\vec{x}(t)$, where v is the velocity of a particle given by a position mapping $\vec{x}(t)$, and m is mass, as usual.

by the equation for motion. Explicitly $V(\vec{x}) \equiv 0$ and

$$\vec{F} = m\frac{\partial^2}{\partial t^2}\vec{x} = 0 \Rightarrow p = m\frac{\partial \vec{x}}{\partial t} = C$$

for some constant C. That is, an object moves in a straight line unless otherwise interrupted.

2.1.2 Lagrangian Formalism and The Euler-Lagrange equation

Newton's formalism of vector equations presupposed an orthogonal coordinate system which we now associate with Euclidean space. However, several difficulties with Newton's original formulation, such as analyzing the global properties of a system or handling constraints given the formulation of motion as a second order equation, led to two reformulations of classical mechanics, which in turn sowed the seeds for quantum mechanics. The first of these formalisms belongs to Lagrange.

Lagrange wished to consider a system whose state is described by N parameters drawn from some configuration space M (which later became identified in the abstract as a manifold). For a generalized coordinate $q(t) \in M$, $t \in [t_i, t_f]$, and each coordinate has an associated generalized velocity defined as

$$\dot{q} = \frac{dq}{dt}.$$

Definition. *Action* is a functional S, which takes given trajectories $q(t)$ of a particle satisfying conditions

$$q(t_i) = q_i$$

$$q(t_f) = q_f$$

with $S[q, \dot{q}] \in \mathbb{R}$, defined by

$$S[q, \dot{q}] = \int_{t_i}^{t_f} L(q, \dot{q})\, dt \tag{2.3}$$

The Lagrangian function $L(q, \dot{q})$ in (2.3) relates these parameters and satisfies the *principle of least action*, which claims that the physically realized trajectory corresponds to the extremum of the functional.

In practice, Lagrangian functions are chosen satisfying the principle of least action. This requires that L is a twice-differentiable function for all arguments. What is remarkable about the Lagrangian formalism is that allows us to derive Newton's second 'law'. This is remarkable because while Newton asserted his law's based off of observation, Lagrange's (and Hamilton's) formalism derived the laws a priori from a mathematical formalism of optimality.

When deriving Newton's law, it is easier to write this principle in a local form as a differential equation (which in fact, allows us to derive Newton's second law). But first, it will be useful to remind the reader of the following definitions:

Definition. *Extrema* are either maxima or minima of a given functional, and a functional $S[f]$ is said to have an extremum with respect to elements f of a given function space defined

over a given domain if

$$\Delta S = S[g] - S[f]$$

have the same sign for all g in an arbitrarily small neighborhood of f.

Definition. The *variational derivative* relates the change of a functional J to the change in a function f by an arbitrarily small function denoted δf. Given configuration space M (a manifold representing functions ρ) and a functional J defined by

$$J : M \to k$$

where k is either the real or complex field, then we formally define the variational derivative of $J[\rho]$ as

$$\frac{\delta J}{\delta \rho} = \int \frac{\delta J}{\delta \rho(t)} \phi(t)\, dt = \lim_{\varepsilon \to 0} \frac{J[\rho + \varepsilon\phi] - J[\rho]}{\varepsilon} = \left[\frac{d}{d\varepsilon} J[\rho + \varepsilon\phi] \right]_{\varepsilon = 0} \tag{2.4}$$

where $\varepsilon\phi$ is the variation of ρ and ϕ is an arbitrary differentiable function.

Incidentally, in Banach spaces (and thus Hilbert spaces), the variational derivative in fact becomes the Fréchet derivative.

Definition. The *variation* or *differential* of a functional J is defined as

$$\delta J = \int \frac{\delta J}{\delta \rho(t)} \delta\rho(t)\, dt \tag{2.5}$$

where $\delta\rho = \varepsilon\phi$ is the variation of ρ.

Given these definition, suppose that q is a path which is an extremum of S. Let $\delta q(t)$ be a variation of q such that $\delta q(t_i) = \delta q(t_f) = 0$. Furthermore, for $\delta q = \varepsilon\rho$, let $\delta q = \varepsilon\rho$ and all its derivatives through the econd order be continuous functions on t, ε.

Now recall that a necessary, but not sufficient condition for a minimum is that a function vanishes in the first derivative, and so we have

$$\left. \frac{dS}{d\varepsilon} \right|_{\varepsilon = 0} = 0$$

and so we find

$$\begin{aligned}
\delta S &= \int_{t_i}^{t_f} \frac{dL(q + \delta q, \dot{q} + \delta\dot{q})}{d\varepsilon} dt \\
&= \int_{t_i}^{t_f} \left(\frac{\partial L}{\partial (q + \delta q)} \frac{d(q + \delta q)}{d\varepsilon} + \frac{\partial L}{\partial (\dot{q} + \delta\dot{q})} \frac{d(\dot{q} + \delta\dot{q})}{d\varepsilon} \right) dt \\
&= \int_{t_i}^{t_f} \frac{\partial L}{\partial (q + \delta q)} \frac{d(q + \delta q)}{d\varepsilon} + \frac{\partial L}{\partial (\dot{q} + \delta\dot{q})} \frac{d}{dt} \left(\frac{d(q + \delta q)}{d\varepsilon} \right) dt
\end{aligned}$$

And now we integrate the second term by parts, finding:

$$\begin{aligned}
\delta S &= \int_{t_i}^{t_f} \frac{\partial L}{\partial(q+\delta q)}\frac{d(q+\delta q)}{d\varepsilon}dt + \left(\frac{d(q+\delta q)}{d\varepsilon}\frac{\partial L}{\partial(\dot{q}+\delta\dot{q})}\right)\Bigg|_{t_i}^{t_f} - \int_{t_i}^{t_f} \frac{d(q+\delta q)}{d\varepsilon}\frac{d}{dt}\frac{\partial L}{\partial(\dot{q}+\delta\dot{q})}dt \\
&\quad \left(\text{as } \delta q(t_i) = \delta q(t_f) = 0 \Rightarrow \frac{dq_i}{d\varepsilon} = \frac{dq_f}{d\varepsilon} = 0 \text{ at } t_i, t_f\right) \\
&= \int_{t_i}^{t_f} \frac{\partial L}{\partial q}\frac{dq}{d\varepsilon}dt - \int_{t_i}^{t_f} \frac{dq}{d\varepsilon}\frac{d}{dt}\frac{\partial L}{\partial\dot{q}}dt \\
&= \int_{t_i}^{t_f} \frac{\partial L}{\partial(q+\delta q)}\frac{d(q+\delta q)}{d\varepsilon} - \frac{d(q+\delta q)}{d\varepsilon}\frac{d}{dt}\frac{\partial L}{\partial(\dot{q}+\delta\dot{q})}dt \\
&= \int_{t_i}^{t_f} \left(\frac{\partial L}{\partial(q+\delta q)} - \frac{d}{dt}\frac{dL}{d(\dot{q}+\delta\dot{q})}\right)\frac{d(q+\delta q)}{d\varepsilon}dt
\end{aligned}$$

And now, by requiring δS to be minimized at $\varepsilon = 0$, we find

$$\begin{aligned}
0 = \left.\frac{dS}{d\varepsilon}\right|_{\varepsilon=0} &= \int_{t_i}^{t_f} \left(\frac{\partial L}{\partial(q+\delta q)} - \frac{d}{dt}\frac{dL}{d(\dot{q}+\delta\dot{q})}\right)\Bigg|_{\varepsilon=0} \left.\frac{d(q+\delta q)}{d\varepsilon}\right|_{\varepsilon=0} dt \\
&= \int_{t_i}^{t_f} \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial}{\partial\dot{q}}\right)\phi\, dt
\end{aligned}$$

and since ϕ is an arbitrary twice, differentiable function, it follows

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial}{\partial\dot{q}} = 0 \tag{2.6}$$

Physically speaking, the result that $\delta S = 0$ follows from our hypothesis that q yields an extremum of action S. Since this holds true for any slight perturbation of q, we obtain the Euler-Lagrange equation (2.6) as the integrand must vanish.

Next, we introduce the concept of *generalized momentum* p_k conjugate to our coordinate q_k defined by

$$p_k = \frac{\partial L}{\partial\dot{q_k}} \tag{2.7}$$

and now we see that the Euler-Lagrange equation can be expressed as

$$\frac{dp_k}{dt} = \frac{\partial L}{\partial q_k}.$$

This change of expression allows one to find the Lagrangian in the classical mechanics of a particle.To see this, consider the arbitrary Lagrangian

$$L = \frac{1}{2}m\vec{\dot{q}}^2 - V(\vec{q}) \tag{2.8}$$

and substitute this into (2.4). We see then that the Lagrangian in (2.6) becomes

$$m\ddot{q}_k + \frac{\partial V}{\partial q_k} = 0 \tag{2.9}$$

Now consider the soon to be all too familiar example of the simple harmonic oscillator.

Example. Consider the following Lagrangian

$$L(q, \dot{q}) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}kq^2$$

The reader can derive the Newtonian formulation

$$m\ddot{x} = m\frac{d}{dt}\dot{x} = -kx$$

by considering that our generalized coordinate $q \leftrightarrow x$, where x is the real line in $\mathbb{R}$.[4]

Lagrange's formalism allows for analysis of symmetries, and also allowed physicists to take into account constraints acting on particle bodies. More importantly, it still finds application today by providing the formalism for the path integral formation of quantum theory.[5] In short, the practical successes of the Lagrangian formalism come from its transformation of its inputs into a second-order ordinary differential equation. However, even this formulation has some limitations.

2.1.3 Hamiltonian Formalism

Hamilton's formulation of mechanics can be thought of as restatement of Lagrange's formulation with an explicit symplectic structure. The Hamiltonian formalism predicts the same outcomes as the Lagrangian formulation. However, this would be to ignore that it gives the equations of motion as first order in the time derivative, and thus it is a formulation that allows physicists to study flows in phase space.

To find a Hamiltonian, H, we first start with a given Lagrangian L, and through a *Legendre transformation of variables* find

$$H(q, p) \equiv \sum_k p_k \dot{q}_k - L(q, \dot{q}).$$

For this transformation to be defined, the Jacobian must satisfy

$$\det\left(\frac{\partial p_i}{\partial \dot{q}_j}\right) = \det\left(\frac{\partial^2 L}{\partial \dot{q}_i \dot{q}_j}\right) \neq 0$$

[4]Hint: Consider $\frac{d}{dt}L$ and note that the principle of least action implies that this L corresponds with an extremum.

[5]The interested reader can check out Richard Feynman's work on this, as it is both mathematically rigourous, and eminently readable.

where the space of (q_k, p_k) is the *phase space* of the system.

Now consider an infinitesimal change in the Hamiltonian by $\delta q_k, \delta p_k$:

$$\delta H = \sum_k \left(\delta p_k \dot{q}_k + p_k \delta \dot{q}_k - \frac{\partial L}{\partial q_k} \delta q_k - \frac{\partial L}{\partial \dot{q}_k} \right) = \sum_k \left(\delta p_k \dot{q}_k - \frac{\partial L}{\partial q_k} \delta q_k \right).$$

It follows

$$\dot{q}_k = \frac{\partial H}{\partial p_k}$$

and

$$\frac{\partial H}{\partial q_k} = -\frac{\partial L}{\partial q_k}.$$

What is exceptionally about Hamilton's formulation is that it lends itself to description by commutation relations $[A, B]$ and brackets $\{A, B\}$

Definition. A *Lie Bracket* is a commutation relation $[A, B]$ which satisfies the following

1. (linearity)
$$[A, c_1 B_1 + c_2 B_2] = c_1 [A, B_1] + c_2 [A, B_2]$$
2. (skew-symmetry)
$$[A, B] = -[B, A]$$
3. (Jacobi Identity)
$$[[A, B], C] + [[C, A], B] + [[B, C], A] = 0$$

We see that for two functions $A(q, p)$ and $B(q, p)$ defined on the phase space of a Hamiltonian H, that we can define a corresponding Lie bracket known as the *Poisson bracket*, which is given by

$$\{A, B\} = \sum_k \left(\frac{\partial A}{\partial q_k} \frac{\partial B}{\partial p_k} - \frac{\partial A}{\partial p_k} \frac{\partial B}{\partial q_k} \right).$$

Of great importance are the fundamental Poisson brackets,

$$\{p_i, p_j\} = \{q_i, q_j\} = 0 \tag{2.10}$$

and

$$\{q_i, p_j\} = \delta_{ij} \tag{2.11}$$

which we leave to the reader to verify. The power of Hamilton's formalism and the Poisson bracket in particular is that we can describe the time development of a physical quantity $A(q, p)$ as

$$\begin{aligned} \frac{dA}{dt} &= \sum_k \left(\frac{dA}{dq_k} \frac{dq_k}{dt} + \frac{dA}{dp_k} \frac{dp_k}{dt} \right) \\ &= \sum_k \left(\frac{dA}{dq_k} \frac{\partial H}{\partial p_k} - \frac{dA}{dp_k} \frac{\partial H}{\partial q_k} \right) \\ &= \{A, H\} \end{aligned}$$

Thus, we see that Hamilton's equations of motion can be written as

$$\frac{dp_k}{dt} = \{p_k, H\}, \frac{dq_k}{dt} = \{q_k, H\} \tag{2.12}$$

Finally, we note that if $\{A, H\} = \dfrac{dA}{dt} = 0$, then the quantity A is conserved.
Now let us revisit the simple harmonic oscillator again, this time formulated as a Hamiltonian

Example. Recall that the Lagrangian before was given as

$$L(q, \dot{q}) = \frac{1}{2}(m\dot{q}^2 - kq^2)$$

Now, we should make the following substitution $k = m\omega^2$, where ω represents angular momentum. Then the Lagrangian becomes

$$L = \frac{1}{2}m(\dot{q}^2 - (\omega q)^2)$$

And thus, the Hamiltonian can be written as

$$H(q, p) = p\dot{q} - L(q, \dot{q}) = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2$$

Furthermore, the reader can verify that Hamilton's equations of motion for the simple harmonic oscillator are

$$\frac{d}{dt}p = -m\omega^2 q$$

$$\frac{d}{dt}q = \frac{p}{m}$$

2.2 The Breakdown of the Classical World: Physicists See the Light

Empirical observations, and the development of the Lagrangian and Hamiltonian formalism consistently verified the predictions made by this second law until the late 19^{th} century. Initially, with the emergence of Maxwell's formulation of electromagnetism, it was assumed that the Newtonian picture was merely incomplete. Maxwell's formulation led many to conclude that light had a maximum velocity, and that it travelled as a wave. However, as further experimentation and observation was performed under electromagnetic framework, whether on bodies of very small or very large mass, the model given of classical mechanics began to breakdown in many spectacular and bizarre ways.

As it so happened, the abstract mathematical tools that developed in analysis and differential geometry would help resolve the discrepancies between the classical picture of the world given by the second law, and what we began observing. Moreover, it is remarkable that the revolution in physics on both sides can in large part be attributed to the work of a single man, Albert Einstein, and his investigations into the properties of light. While he is more

well known for his theory of relativity, addressing the universe at a large scale, his early work for which he won his Nobel prize (and the work relevant to this paper) on the photo-electric effect argued that light behaves as *both* a particle and a wave.

Physical theories describing light have alternated throughout the history of mathematics between those which had that light was a particle and those that described light as a wave. Isaac Newton first proposed the *corpuscular* theory of light, which was held to be the model of the propagation of light until the 19th century. Then, starting with Young's double slit experiment in 1803, it was established that light exhibited a clear wave interference pattern. Throughout the 19th century, light was held to be a wave which would strike a particle, and that it was the intensity of light striking a surface which determined the energy of the electron emitted from the surface. The only thing missing experimental verification was the medium which light was propagating through as a wave.[6] The first chinks in this wave theory came with the negative experimental results of the Michelson-Morley experiment regarding the proposed medium that light was passing through. However, Max Planck's work attempting to explain black body radiation, the type of electromagnetic radiation surrounding a body in thermodynamic equilibrium, suggested that light waves could only gain or lose energy in finite amounts related to their frequency. These finite amounts were named *quanta*

2.2.1 Planck Breaks The Universe Down

Since the middle of the nineteenth century, physicists had been investigating an idealized physical body called a *black body*, which absorbs all incident electromagnetic radiation regardless of the frequency or angle of incidence. What they noticed is that a black body held at a constant temperature emits electromagnetic radiation, which they called *black body radiation.* The pioneering work of Kirchhoff, Stefan, Boltzmann, and Wien started with the observation that the 'hotter an object', the 'brighter the glow' of an object. Physically speaking, when a heated object is in equilibrium with light, the amount of light absorbed is equal to the amount of light that is emitted, and so a black body absorbs all the light that hits it while emitting a maximum amount of radiation.

Max Planck resolved this by introducing a formalism to describe black-body radiation that was consistent with theoretical electromagnetism and theoretical thermodynamics, as well as the emerging experimental data. Succinctly, Planck's Law, for wavelength λ, and absolute temperature T, can be given as

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \tag{2.13}$$

where k_B is the Boltzmann constant, c is the speed of light, B is the spectral radiance, and h was a new constant called the Planck constant. This law describes the unique spectral distri-

[6]The hypothesis was that light was passing through the luminiferous aether, which was first proposed by Huygens in 1678.

bution for electromagnetic radiation in thermodynamic equilibrium. Although this description could in part be accounted for by the formalism of classical mechanics, the introduction of h had only a formal precursor without a physical referent.

This new universal constant was derived by Planck's theoretical supposition that the distribution of electromagnetic energy could be described over different modes of several oscillators. Planck's hypothesis that the equations of motion for light could be described as a linear combination of oscillators with finitely many characteristic frequencies was in part based off of the emerging subject of statistical mechanics, as well as a heuristic approach of Boltzmann's to describe energy distribution by an arbitrarily small energy constant ϵ. Planck's big breakthrough was to realize that the constant in Boltzmann's work was not merely a mathematical formalism, but an actual quantity. In supposing that the total energy was distributed to each oscillator by an integer multiple of a definite physical unit of energy characteristic of the respective oscillation frequencies ν, Planck was able to reconcile observation with theory. By relating the minimum amount of energy in any given system to this constant h, the energy of a system could now be given by

$$E = h\nu \tag{2.14}$$

This was the birth of quantum mechanics, as energy could no longer be thought of as existing in a continuum where $\epsilon > 0$. Rather, there is a lower bound

$$\epsilon > h > 0$$

Building off Planck's work, Einstein overturned the view that it was the intensity of light which led to electron emission by mathematically proving that it is the *frequency of light*, denoted ν, which determines the energy emitted. Einstein further showed that the intensity of the light determines the number of electrons that are emitted. In doing so, he argued that photons possess both their well-known wave behaviour, while also exhibiting a clear particle-like behavior. Specifically, Einstein predicted that the energy of individual ejected electrons off of a surface increase linearly with the frequency of light. For this work, experimentally verified in 1914, Einstein won his Nobel Prize.

What Einstein showed was that Planck's relation (2.14)

$$E = h\nu.$$

implies that a photon must have a frequency at or beyond the threshold frequency to have the energy needed to eject a single electron. This is the so called photo-electric effect. Subsequent work, most of all by deBroglie, demonstrated was that the momentum of a photon can be described as

$$p = \frac{h}{2\pi}\frac{2\pi}{\lambda}$$

where λ is the wavelength of light. What Planck and Einstein argued, and what has been empirically verified since, is that the reduced Planck's constant $\hbar := \frac{h}{2\pi}$, is the universal constant describing energy quanta. It is the scale where quantum effects are manifested. Setting $k = \frac{2\pi}{\lambda}$ and $\omega = 2\pi\nu$, we can elegantly describe the energy and momentum of photons, as well as other quanta, by the deBroglie relations:

$$E = \hbar\omega \tag{2.15}$$

$$p = \hbar k. \tag{2.16}$$

For a sense what these variables describe, ω is called the angular frequency, or sometimes the angular velocity of a quanta, k is called the angular wave number, and it is inversely proportional to wavelength, λ.

2.2.2 Double Slit Experiment

Einstein's wave-particle duality account of light confounded physicists who had come to view light as exhibiting a wavelike behaviour. Notably, this behaviour was established by Thomas Young in 1803 with his double-slit experiment. In this experimental setting, a beam of light passes through two narrow parallel slits and is projected onto a screen, creating an interference pattern which corresponds to the interference pattern of waves.

However, it isn't just light that behaves like both a particle and a wave. More mysteriously, *electrons*, the subatomic particles that had been thought to orbit nucleus of atoms at the turn of the twentieth century were also observed to behave in a wavelike manner. Although this will be a simplified account of the double slit experiment, it should clarify the experiment, the observed phenomena, and the mathematics.

The experiment started as a lab accident. Scientists were firing an electron gun at a block of Nickel crystal when the evacuated chamber exploded. As part of cleaning up the lab for further experiments, the scientists cleaned the Nickel crystal by heating it up to remove the foreign contaminants. When they started firing the electron gun again, because the structure of the crystal had been altered, they scientists observed that the electrons exhibited a new pattern of behaviour, one which violated the classical assumptions of an electron path.[1]

Rather than following a clear, Newtonian trajectory, the scientists witnessed interference patterns suggestive of a wave. This raised new questions about what an electron was exactly. Given the observed interference pattern, the natural question to ask was: are individual electrons exhibiting a wavelike motion?

The result of these experiments was that the electron was no longer seen to have an orbit around a nucleus described by q, with momentum $p = mv$ as a function of time, as in the case of classical mechanics. Instead, it was seen that physicists would have to reconcile wavelike

behaviour of a particle path with wave-momentum described by

$$p = \frac{h}{\lambda} = \hbar k$$

2.3 Heisenberg's Matrix Mechanics

The first logically consistent attempt to incorporate Einstein's groundbreaking work with the vast body of new experimental observations in a new physical model comes from the work of Heisenberg, and a subsequent collaborations with Born and Jordan [7]. Heisenberg's breakthrough came as he was studying the problem of calculating spectral lines of the hydrogen atom by working within Bohr's model describing quantum jumps. His startling observation was that describing atomic systems in terms of *observables* saw that the classical commutativity of variables p, q where

$$pq = qp$$

no longer had empirical support. His big insight was that the physical quantities which were measured were not variables, but rather, corresponded to *matrices*. Working with the experimental data on hydrogen spectroscopy, and treating position and momentum as matrices $\hat{q}$ and $\hat{p}$ respectively, Heisenberg assumed the following commutation relation

$$[\hat{p}, \hat{q}] = -i\hbar I$$

where I is a unit matrix.[8] This formulation corresponded with the values of the frequencies and the strengths of the spectra of the hydrogen atom.[9]
Heisenberg's formulation built on the premise that all physical observables (which are measurable material phenomena such as energy, position or momentum), could be represented by Hermitian matrices, and indexed by energy levels. In turn, the eigenvalues of the matrix of an observable correspond to the possible values that an observable may take, while an eigenvector was the state of the system at measurement.

This formulation was quite spooky, as it implies that the physical world existed in multiple states simultaneously, but that observation, in this case, measuring a phenomenon, would collapse the state of a physical system from all the possible values it might take to the one that was observed! What makes this spooky is that in practice, most observables do not share eigenvectors, and so most observables cannot be measured simultaneously. The biggest leap yet from Newton's world was that position $\hat{p}$ and momentum $\hat{q}$ do not share common eigenvectors, and cannot both be known at the same time (we will explore why this is the case rigourously in section 3.)

[7] The seminal papers are "On A Quantum Interpretation of Kinematical and Mechanical Relations" by Heisenberg; "On Quantum Mechanics", from Born and Jordan; and "On Quantum Mechanics II" from Born, Jordan, and Heisenberg.

[8] Strangely enough, Schrödinger was able to show that $p \to -i\hbar\frac{d}{dx}$ and $x \to x$, which would yield this commutation relation.

[9] Spectroscopy gives observational data on the transition at the atomic level arising from the interaction of atoms with light quanta. The experimentalists following Bohr believed that only what was measurable by spectroscopy should appear in a theory of quantum mechanics. In this case, energy level and intensity of an emission could be measured, but not location!

The picture of the world given by matrix mechanics, known as the *Heisenberg picture*, can be succinctly described as noting that observables are time dependent while the state-vectors, are time independent. Heisenberg formulated that for an observable $\hat{A}$ where the Hamiltonian H describing the total energy of the quantum system being observed does not vary with time, the time-evolution operator for the observable could be written as

$$U(t) = e^{\frac{-iHt}{\hbar}} \tag{2.17}$$

and so the time evolution of an observable could be given by

$$A(t) = U^* \hat{A} U = e^{\frac{iHt}{\hbar}} \hat{A} e^{\frac{-iHt}{\hbar}} \tag{2.18}$$

And thus

$$\begin{aligned}
\frac{d}{dt}A(t) &= \frac{i}{\hbar} H e^{\frac{iHt}{\hbar}} \hat{A} e^{\frac{-iHt}{\hbar}} + e^{\frac{iHt}{\hbar}} \frac{\partial \hat{A}}{\partial t} e^{\frac{-iHt}{\hbar}} - \frac{i}{\hbar} e^{\frac{iHt}{\hbar}} A \cdot H e^{\frac{-iHt}{\hbar}} \\
&= \frac{i}{\hbar} e^{\frac{iHt}{\hbar}} (HA - AH) e^{\frac{-iHt}{\hbar}} + e^{\frac{iHt}{\hbar}} \left(\frac{\partial \hat{A}}{\partial t} \right) e^{\frac{-iHt}{\hbar}} \\
&= \frac{i}{\hbar}[H, A(t)] + e^{\frac{iHt}{\hbar}} \left(\frac{\partial \hat{A}}{\partial t} \right) e^{\frac{-iHt}{\hbar}} \\
&= \frac{i}{\hbar}[H, A(t)] + \frac{\partial A(t)}{\partial t} \\
&= = \frac{i}{\hbar}[A(t), H] + \frac{\partial A(t)}{\partial t}
\end{aligned}$$

And so, the Heisenberg equation of motion,

$$\frac{d}{dt}A(t) = -\frac{i}{\hbar}[A(t), H] + \frac{\partial A(t)}{\partial t} \tag{2.19}$$

can be realized as a commutation relation. And moreover, for operators $\hat{A}$ which are not time dependent

$$\frac{d}{dt}\hat{A} = -\frac{i}{\hbar}[\hat{A}, H] \tag{2.20}$$

Thus, we see is that Hamilton's equations of motion have become

$$\frac{d\hat{q}_i}{dt} = -\frac{i}{\hbar}[\hat{q}_i, H]$$

$$\frac{d\hat{p}_i}{dt} = -\frac{i}{\hbar}[\hat{p}_i, H]$$

Finally, Heisenberg's picture of motion preserves classical mechanics given the following correspondence between the commutator $[A, B]$ and the Poisson bracket $\{A, B\}$.

$$i\hbar\{A, H\} \leftrightarrow [A, H].$$

Now, this may seem a bit too abstract, so let us walk through the motivating example for this formulation: the spectra of the energy operator on a simple harmonic oscillator:

Example. Heisenberg's formulation describes the task of finding the allowed energies of a system, denoted E_n, is equivalent to the task of finding the corresponding eigenvalues and eigenvectors of an operator. In the case of the single dimensional harmonic oscillator, the eigenvalues can be found by considering the following two operators, dubbed the *annihilation* and *creation* operators, for reasons we shall soon see:

$$\hat{a} := (2m\hbar\omega)^{-1/2}(xm\omega + ip)$$

$$\hat{a}^\dagger := (2m\hbar\omega)^{-1/2}(xm\omega - ip)$$

It should be clear that $\hat{a}^\dagger$ is defined as the conjugate transpose of $\hat{a}$, and moreover, their product is related to the Hamiltonian for a harmonic oscillator

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2$$

as follows

$$\begin{aligned}
\hat{a}\hat{a}^\dagger &= (2m\hbar\omega)^{-1}(xm\omega + ip)(xm\omega - ip) \\
&= (2m\hbar\omega)^{-1}[x^2m^2\omega^2 + p^2 + im\omega(px - xp)] \\
&= (\hbar\omega)^{-1}(H + \frac{i\omega}{2}[p, x]) \\
&= (\hbar\omega)^{-1}(H + \frac{\hbar\omega}{2})
\end{aligned}$$

And so we can now express the Hamiltonian of this system as

$$H = \hbar\omega(\hat{a}\hat{a}^\dagger - \frac{1}{2})$$

Similarly, the reader can verify that

$$\hat{a}^\dagger\hat{a} = (\hbar\omega)^{-1}\left(H - \frac{\hbar\omega}{2}\right)$$

and thus

$$H = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2})$$

with the consequence that

$$[\hat{a}, \hat{a}^\dagger] = \hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} = I$$

We also leave it to the reader to verify that

$$[H, \hat{a}] = -\hbar\omega\hat{a}$$

and

$$[H, \hat{a}^\dagger] = \hbar\omega\hat{a}^\dagger$$

Now, physically speaking, a simple harmonic oscillator can only have non-negative real energy values. So, when considering this Hamiltonian, we should find that there are only nonnegative real corresponding eigenvalues.

Assuming this result, let E_0 denote the lowest energy level allowed by the system, and let ψ_0 be the corresponding eigenvector. Then

$$H\psi_0 = E_0\psi_0$$

and thus

$$\hat{a}H\psi_0 = E_0\hat{a}\psi_0.$$

Now, by the commutation relations we established above for H and a, we find

$$\hat{a}H\psi_0 = (H\hat{a} + \hbar\omega\hat{a})\psi_0 = E_0 a\psi_0 \Rightarrow H\hat{a}\psi_0 = (E_0 - \hbar\omega)\hat{a}\psi_0$$

By our assumption that E_0 is the lowest allowed eigenvalue, it follows that this must be a trivial solution, and so $\hat{a}\psi_0 \equiv 0$. Now let's apply $a^\dagger$ to this. Now we find

$$\hat{a}^\dagger\hat{a}\psi_0 = (\hbar\omega)^{-1}\left(H - \frac{\hbar\omega}{2}\right)\psi_0 = 0 \Rightarrow H\psi_0 = \frac{\hbar\omega}{2}\psi_0 \Rightarrow E_0 = \frac{\hbar\omega}{2}$$

Now we have finally seen our first quantized, energy value. We see that the ground state of this system must be an energy value of $\frac{\hbar\omega}{2}$.

Additional application of the $\hat{a}^\dagger$ operator gives us a larger energy value, as seen by

$$\hat{a}^\dagger H\psi_0 = (H\hat{a}^\dagger - \hbar\omega\hat{a}^\dagger)\psi_0 = E_0\hat{a}^\dagger\psi_0 \Rightarrow H(\hat{a}^\dagger\psi_0) = (E_0 + \hbar\omega)\hat{a}^\dagger\psi_0$$

and so $E_0 + \hbar\omega = \frac{3\hbar\omega}{2}$ is the corresponding energy level for $\hat{a}^\dagger\psi_0$. This is why $\hat{a}^\dagger$ is called the creation operator.

We leave it to the reader to figure out why a is called the annihilation operator, and moreover, to verify that

$$E_n = (n + \frac{1}{2})\hbar\omega$$

and

$$\psi_n = c_n(\hat{a}^\dagger)^n\psi_0$$

where c_n is a constant that normalizes ψ_n. (We leave it to the reader to verify that for $\hat{a}^\dagger\psi_n = \sqrt{n+1}\psi_{n+1}$ and that $\hat{a}\psi_n = \sqrt{n}\psi_{n-1}$

2.4 Schrödinger's Wave Mechanics

While Heisenberg, Born and Jordan were the first physicists to publish a logically consistent model of quantum mechanics that corresponded to the experimental work in atomic spectroscopy, their formulation was highly controversial. Matrices were viewed as abstract objects belonging to pure mathematicians by the physics community at large, and the focus

on discrete state jumping and state equation collapse upon observation rather than the wave-particle duality of the Einstein-Planck mold suggested a highly bizarre world, too strange to be a serious candidate for quantum mechanics.

Unfortunately for those physicists, desperate to hold onto the determinism of classical mechanics, the world really is bizarre. The task of resolving the strangeness of the world was not helped by Erwin Schrödinger, who further refined the wave-particle duality of subatomic particles.

Building off the description of the momentum of a photon

$$p = \hbar k$$

researchers such as de Broglie showed that assumed that matter waves propagate along with their particle counterparts, the subatomic particles would form standing waves, where only certain discrete rotational frequencies about the nucleus would be allowed. These quantized orbits corresponded with the observed discrete energy levels. What remained to be discovered was a wave equation describing the particle behaviour.

Recall that wave equations are hyperbolic partial differential equations whose solutions are functions. They are common throughout mathematical physics, and are often derived from other, earlier physical laws. This was not to be the case for Schrödinger's equation, which can be simply expressed in the time-independent case as

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi = E\Psi \tag{2.21}$$

where ψ is a wave equation satisfying a Hamiltonian corresponding to the observable in question.

The basis for the equation is the observed energy of a particle followed from Einstein's work on the photo-electric effect, where energy is a function of angular frequency

$$E = \hbar\omega,$$

the assumed principle of energy conservation, and de Broglie's doctoral work which hypothesized that any particle could be associated with a wave, and that the momentum corresponded to the wavelength of a wave, i.e. the familiar

$$\vec{p} = \hbar\vec{k}.$$

Realizing that $\hbar$ could be a natural unit, Schrödinger saw the underlying identities in the earlier equations, wherein energy and time are related, as are space and momentum. Then, in an insight without precedence in the physical research, Schrödinger saw that using these identities, one could describe the phase of a plane wave as a complex phase factor, i.e. in the

one-dimensional case with normalizing constant A,

$$\Psi = Ae^{i(\vec{k}\cdot\vec{q}-\omega t)} = Ae^{i\frac{(\vec{p}\cdot\vec{q}-Et)}{\hbar}} \tag{2.22}$$

Then, in considering the first order partial derivatives of this wave, with respect to time

$$\frac{\partial\Psi}{\partial t} = -\frac{iE}{\hbar}Ae^{i\frac{(\vec{p}\cdot\vec{q}-Et)}{\hbar}} = -\frac{iE}{\hbar}\Psi$$

and with respect to space

$$\nabla\psi = \frac{i}{\hbar}\vec{p}Ae^{i(\frac{(\vec{p}\cdot\vec{q}-Et)}{\hbar}} = i\vec{p}\psi$$

Schrödinger found

$$-i\hbar\nabla\Psi = \vec{p}\psi \Rightarrow -\frac{\hbar^2}{2m}\nabla^2\psi = \frac{1}{2m}\vec{p}\cdot\vec{p}\Psi$$

and

$$\frac{\partial}{\partial t}\Psi = -\frac{iE}{\hbar}\Psi \Rightarrow i\hbar\frac{\partial\Psi}{\partial t} = E\Psi.$$

Thus, in the time-independent case of a single, non-relativistic particle, the Hamiltonian can be given as

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{q})$$

To illustrate the wave equation, consider the following two examples:

Example. First, consider the Hamiltonian for a free, non-relativistic particle with no potential energy in one dimension. Here, the Hamiltonian will be given as

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}$$

Then we are asked to solve

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\Psi = E\Psi$$

We do so by considering the

$$\Psi(x,t) = Ae^{\alpha x}e^{\beta t}$$

and so we find

$$\begin{aligned}
\frac{\partial}{\partial x}\Psi &= \frac{\partial}{\partial x}(Ae^{(\alpha x+\beta t)}) \\
&= \alpha(Ae^{(\alpha x+\beta t)}) \\
&= \alpha\Psi \\
&\Rightarrow \\
\frac{\partial^2}{\partial x^2}\Psi &= \frac{\partial}{\partial x}\alpha(Ae^{(\alpha x+\beta t)}) \\
&= \alpha^2\Psi
\end{aligned}$$

Hence

$$\hat{H}\Psi = -\frac{\hbar^2}{2m}\alpha^2\Psi = E\Psi \Rightarrow \alpha^2 = \frac{-2mE}{\hbar^2} \Rightarrow \alpha = i\frac{\sqrt{2mE}}{\hbar}$$

In the particle view of the system, we have

$$E = \frac{m}{2}v^2 = \frac{p^2}{2m} \Rightarrow \alpha = i\frac{p}{\hbar}$$

and using the deBroglie hypothesis, generalizing the photo-electric effect, we find that

$$\alpha = i\frac{p}{\hbar} = i\frac{\hbar k}{\hbar} = ik$$

$$\begin{aligned} \frac{\partial}{\partial t}\Psi &= \frac{\partial}{\partial t}(Ae^{ikx+\beta t}) \\ &= \beta\Psi \end{aligned}$$

Hence

$$i\hbar\frac{\partial}{\partial t}\Psi = i\hbar\beta\Psi \Rightarrow E = i\hbar\beta \Rightarrow \beta = \frac{-iE}{\hbar} = -\omega$$

and thus we find that in the non-relativistic free particle case that the corresponding wave equation is

$$\Psi = Ae^{i(kx-\omega t)}$$

solves the wave equation. What's more, we see that in a 'position' basis, momentum in this formulation can be described as

$$\vec{p} = -i\hbar\nabla \tag{2.23}$$

When it comes to finding the permissible energy level's the free system, we note that

$$H\psi = E\psi$$

allows us to derive

$$E = \frac{(\hbar k)^2}{2m} = \hbar\omega$$

However, there is one MAJOR problem: the wave function ψ is NOT normalizable, and so the theory tells us that a free particle with definite energy does not exist. To see this, notice that

$$\int_{\mathbb{R}} \psi^*\psi = \int_{\mathbb{R}} |\psi|^2 = |A|^2 \int_{\mathbb{R}} dx = |A|^2(\infty)$$

The failure of a free particle to have physical meaning does not mean it lacks use in the study of quantum mechanics. Physicists use the separable solutions to the time independent case when studying general solutions to the time-dependent Schrödinger's equation. In practice, when it is possible to use a discrete index, this can be found by considering the plane waves, $\psi(x) \equiv \Psi(x, 0)$. Then, if a discrete energy level is allowed $\{E_n\}$ the Schrödinger wave equation

implies that

$$\Psi(x,t) = \sum_n A_n \psi_n(x) e^{\frac{-iE_n t}{\hbar}} \tag{2.24}$$

Moreover, in practice, when dealing with continuous spectra, which observationally fall into countable partitions, this can be described by taking the limit of a linear combination over a discrete index over arbitrarily small intervals. However, rather than consider a linear combination over a discrete index, the general solution is found as an integral over a continuous variable k, called the *wave packet*. The *wave packet* is given the form

$$\psi(x,t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \phi(k)^{i(kx - \frac{\hbar k^2}{2m} t)}\, dk \tag{2.25}$$

The wave function can be normalized for an appropriate function $\phi(k)$, which can be found by considering the time-independent free particle case,

$$\psi(x,0) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \phi(k) e^{ikx}\, dk$$

and applying Plancheral's formula (i.e., consider a Fourier transform)

Now onto a more sophisticated example.

Example. The solution to the equation

$$F = -kx = m\frac{d^2}{dx^2}$$

can be given by the generalized solution

$$x(t) = A\sin(\omega t) + B\cos(\omega t)$$

where ω is the angular frequency of oscillation, defined here by $\omega = \sqrt{\frac{k}{m}}$. The reader can verify this.

Moreover, the potential energy is given as

$$V(x) = \frac{1}{2}kx^2$$

in the classical case, while in the quantum case, the quantum potential of a harmonic oscillator is given as

$$V(x) = \frac{1}{2}m\omega^2 x^2.$$[10]

Solving this system via Schrödinger's equation means solving

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

[10]Hopefully the reader sees both the similarity to the classical case and understands the reasoning for why we describe the potential as a function of angular frequency.

We can simplify this solution method by introducing the dimensionless variable

$$\xi = \sqrt{\frac{m\omega x^2}{\hbar}}$$

and by expressing energy in terms of $\frac{\hbar\omega}{2}$ units, so that

$$E' = \frac{2E}{\hbar\omega}.$$

Then Schrödinger's equation reads as

$$\frac{d^2\psi}{d\xi^2} = (\xi^2 - E')\psi$$

We leave deriving the allowable energy states to the reader (Hint: $E_N = (N + \frac{1}{2})\hbar\omega$. First consider ψ as $\xi \to \pm\infty$. Can you find a recursion relationship?).

Even though there were now two competing models, it was clear: particles no longer had determined properties as described by classical mechanics, where position and momentum were known to follow Newton's laws. Now, measuring a particle meant observing a result drawn from a probability distribution. Even if we knew the wave function governing a particle body, the result of a measurement would always remain uncertain. The world became uncertain.

It is worth taking the time to point out that with this wave formulation of an electron path, physicists first believed that the wave equation was describing the electron as being a spread-out entity. That is, the wave equation described an electron smearing out in space by a *wave density* function. However, this is incorrect. Experiments and further refinements to the theory of quantum mechanics have verified that the electron is a point-like particle. It was Max Born who resolved this conundrum by describing a wave satisfying Schrödinger's equation as a probability wave which describes where the particle is spatially expected to be, as well as implicitly describing the expected velocity of the particle. With this interpretation of the wave function ψ as a *probability density function*, the greater the amplitude of a region of the wave, the greater the probability that a particle would be within that bounded region of space. Consequently, classical mechanics is simply verboten at the quantum level of Planck lengths since electrons are no longer seen to follow fixed trajectories described by the Newtonian position function $x(t)$. Instead, electrons and other quantum bodies, are part of a dynamical probability wave.

2.5 Dirac's Quantum Mechanics

Schrödinger's equation had numerous startling consequences, as it struck the death blow to classical mechanics once and for all. What's more, because of its linearity, solutions could be

combined to derive new solutions, i.e.

$$\{\psi_{k_j} = e^{i(k_j x - \omega_j t)}\} \mapsto \sum_{k_j} c_j e^{i(k_j x - \omega_j t)}$$

Now, if given a wave function $\psi \in L_2(\Omega)$, there exists another function $\phi(k)$ which would solve the equation such that

$$\psi(x,t) = \frac{1}{\sqrt{2\pi}} \int_\Omega \phi(k) e^{ikx}\, dk$$

and

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_\Omega \psi(x,t) e^{-ikx}\, dx.$$

This can be recognized as the Fourier transform. What is quite astonishing about this is that $\phi(k)$ has a physical meaning: $\phi(k)$ yields the probability that a particle has momentum $p = \hbar k$! The abstract spaces representing position and momentum are respectively linked via a Fourier transform. In fact, this would entail that not only would a particle have more than one possible position, but that at each position it has more than one momentum.[11]

By 1927, quantum mechanics was faced with startling theoretical difficulties. There were two formulations which yielded the same results, were logically self-consistent, amazingly successfully at describing and predicting the world, and yet were seemingly at odds. There were slippery problems that needed to be addressed. Heisenberg's formulation was a theory of observables, a theory for the experimentalists; Schrödinger's was a theory of quantum states, a theory for the theorists. Schrödinger recognized that Heisenberg's matrices were matrix elements of a position operator $\hat{Q}$ with respect to an orthonormal basis in $L_2(\mathbb{R}^3)$ given the general Hamiltonian $H = \dfrac{\vec{p}^2}{2m} + V(Q)$, while the vectors in ℓ_2 on which the Heisenberg matrices corresponded to quantum states. What he lacked was an articulation of abstract Hilbert spaces, and so he was unable to prove the two formulations were equivalent (that would take von Neumann). Furthermore, while some solutions to Schrödinger's equation were normalizable and were labeled by discrete indices (like the simple harmonic oscillator), other solutions to Schrödinger's equation were effectively non-normalizable and labeled by a continuous variable (such as the free particle with states k).

Paul Dirac, a theoretical physicist who wrote the first textbook on quantum mechanics[12] based off of his Cambridge lectures, made an attempt to prove the equivalence of Heisenberg's matrix mechanics with Schrödinger's wave mechanics. His proposed solution to enforce

[11] The question we need to ask ourselves is why do we not notice quantum effects (except in how they're used in day to day life)? Why is it that we don't see objects popping in and out of existence, appearing and disappearing? The answer is: mass. Consider a cat weighing about 2.5 kilograms, and consider our error measurement of this cat's location to be $\Delta x = .02mm$. We can now derive the intrinsic speed of the cat due to quantum effects by

$$\Delta v = \frac{\hbar}{2m\Delta x} \approx \frac{10^{-34} J \cdot s}{4(2.5kg)(10^{-5}m)} \approx 5 \times 10^{-30} m/s$$

This is an incredibly imperceptible evolution, and it becomes moreso as we increase the mass of our observed object. In this way, we were able to 'recover' classical mechanics at the scale of everyday observation.

[12] Principles of Quantum Mechanics, 1930

orthonormality and unify the two formulations was a function which he dubbed the *Dirac-Delta function*, which was defined as having a value of 0 except at a single specified point a, where it would have a value of ∞, while the area under the 'curve' would be 1. That is

$$\delta(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases}$$

$$\int_{\mathbb{R}} \delta(x-a)\,dx = 1$$

moreover, for $\epsilon > 0$

$$\int_{a-\epsilon}^{a+\epsilon} \delta(x-a)\,dx = 1$$

and

$$\int_{\mathbb{R}} F(x)\delta(x-a)\,dx = F(a).$$

Dirac placed several rules on this function, as it was only intended to simplify the mathematical expression he was trying to capture as a bit of notation. In practice, this formulation was intended to behave as a collapsed Gaussian distribution

$$\delta(x-y) = \lim_{\varepsilon \to 0} \frac{1}{\sqrt{2\pi}\varepsilon} e^{-\frac{(x-y)^2}{2\varepsilon^2}}$$

To see how it can be 'derived', start with the Fourier transforms

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \phi(k) e^{ikx}\,dk$$

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \psi(y) e^{-iky}\,dy$$

and then notice that

$$\psi(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \psi(y) e^{ik(x-y)}\,dk\,dy$$

where $\delta(x-y) = \dfrac{1}{2\pi} \int_{\mathbb{R}} e^{ik(x-y)}\,dk$ and hence

$$\psi(x) = \int_{\mathbb{R}} \psi(y)\delta(x-y)\,dy.$$

The physical interpretation of this δ function is to suggest that the physical location of the particle is completely determined. This would entail that $\phi(k) = 1$ for all k.

Example. When considering Schrödinger's equation with a Dirac-delta potential

$$V(x) = -\alpha\delta(x)$$

we are to solve for

$$-\frac{\hbar}{2m}\frac{d^2\psi}{dx^2} - \alpha\delta(x)\psi = E\psi$$

Before proceeding to solve this equation, there are three conditions that need to be stated:

1. ψ is continuous on Ω, and here $\Omega = \mathbb{R}$.
2. $\frac{d\psi}{dx}$ is continuous almost everywhere on Ω.
3. ψ vanishes at the boundaries of Ω, which means here that $\lim_{|x|\to\infty} \psi(x) = 0$

Now there are two possible states to consider: *bound states* where $E < 0$ and *scattering states* where $E > 0$. First consider bound states:

- (Bound States) In region $x < 0$, the potential $V(x) = 0$ by definition and so

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2} = \kappa^2\psi$$

where $\kappa = \dfrac{\sqrt{-2mE}}{\hbar} > 0$ since $E < 0$. Then the general solution is given by

$$\psi(x) = Ae^{-\kappa x} + Be^{\kappa x}$$

and since $e^{-\kappa x} \overset{x\to-\infty}{\to} +\infty$, it follows that $A = 0$ in order for this to be a real valued solution, hence

$$\psi(x) = Be^{\kappa x}$$

for $x < 0$. Similarly, for $x > 0$, the general form taken is

$$\psi(x) = Ce^{-\kappa x} + De^{\kappa x}$$

and since $e^{\kappa x} \overset{x\to\infty}{\to} +\infty$, it follows that $D = 0$ and so

$$\psi(x) = Ce^{-\kappa x}$$

for $x > 0$. Now considering the boundary conditions at $x = 0$ such that ψ is continuous and $\dfrac{d\psi}{dx}$ is continuous a.e. (expect where the potential is infinite). Then, without loss of generality, let $B = C$

$$\psi(x) = \begin{cases} Be^{\kappa x} & x \le 0 \\ Be^{-\kappa x} & x \ge 0 \end{cases}$$

Now we integrate the Schrödinger equation over an infinitesimal neighbourhood about the origin, i.e

$$-\frac{\hbar^2}{2m}\int_{-\varepsilon}^{\varepsilon} \frac{d^2\psi}{dx^2}\,dx + \int_{-\varepsilon}^{\varepsilon} V(x)\psi(x)\,dx = E\int_{-\varepsilon}^{\varepsilon} \psi(x)\,dx$$

which integrates to

$$-\frac{\hbar^2}{2m}\left(\left.\frac{d\psi}{dx}\right|_{-\varepsilon}^{\varepsilon} = +\alpha\int_{-\varepsilon}^{\varepsilon} \delta(x)\psi(x)\,dx = \alpha\psi(0)\right.$$

We find as $\varepsilon \to 0$ that

$$\lim_{\varepsilon \to 0} -B\kappa e^{-\kappa\varepsilon} - B\kappa e^{\kappa\varepsilon} = -\frac{2m\alpha}{\hbar^2}\psi(0) \Rightarrow -2B\kappa = -\frac{2m\alpha}{\hbar^2}B \Rightarrow \kappa = \frac{m\alpha}{\hbar^2}$$

therefore

$$E = -\frac{(\hbar\kappa)^2}{2m} = -\frac{m\alpha^2}{2\hbar^2}$$

and we can normalize ψ by setting $B = \sqrt{\kappa}$. In this case, there is exactly one bound state given by

$$\psi(x) = \frac{\sqrt{m\alpha}}{\hbar} e^{-\frac{m\alpha|x|}{\hbar^2}}$$

with corresponding energy

$$E = -\frac{m\alpha^2}{2\hbar^2}$$

- (Scattering States) When $E > 0$, we consider $x < 0$ to find

$$\frac{d\psi^2}{dx^2} = -\frac{2mE}{\hbar^2}\psi = -k^2\psi$$

with $k = \dfrac{\sqrt{2mE}}{\hbar}$, which is real and positive. The general solution then can be given by

$$\psi(x) = Ae^{ikx} + Be^{-ikx}$$

and similarly for $x > 0$, we have

$$\psi(x) = Ce^{ikx} + De^{-ikx}$$

and our boundary conditions require that

$$A + B = C + D$$

and

$$ik(A - B) = ik(C - D)$$

When integrating Schrödinger's equation over an infinitesimal neighbourhood about the origin we find that the second boundary condition requires

$$ik(C - D - A + B) = -\frac{2m\alpha}{\hbar^2}(A + B) \Rightarrow C - D = A\left(1 + 2i\frac{m\alpha}{\hbar^2 k}\right) - B\left(1 - 2i\frac{m\alpha}{\hbar^2 k}\right)$$

We immediately see that we have two equations with five unknowns, which cannot be solved, and moreover, that this state ψ is not normalizable. The theory would have it that this state does not exist. However, as physicists like to study the scattering phenomenon, they have resolved this difficulty by constructing normalizable linear combinations of the stationary states, and then look at the relative probabilities that a particle will push through the potential barrier. These problems form the theoretical bedrock of quantum

tunneling.

In subsequent editions of his textbook, Dirac unveiled the "bra-ket" notation, which is still standard in the literature (and much preferred to von Neumann's own notation), when dealing with the arbitrary vectors that appear in the competing quantum formalisms.

Definition. A *ket* can be thought of as a vector, denoted $|x\rangle$, existing in an arbitrary inner product space χ.

Definition. A *bra* can be thought of as the dual of $|x\rangle$, denoted $\langle x|$, effectively it is a linear transformation acting on $|x\rangle$ such that

1. $$\langle y|x\rangle \in \mathbb{C}$$
2. $$\langle y|(|x_1\rangle + |x_2\rangle) = \langle y|x_1\rangle + \langle y|x_2\rangle$$
3. $$(\langle y_1| + \langle y_2|)|x\rangle = \langle y_1|x\rangle + \langle y_2|x\rangle$$
4. $$\langle y|(c|x\rangle) = (c\langle y|)|x\rangle = c\langle y|x\rangle$$
5. $$\langle y|cx\rangle = \langle \bar{c}y|x\rangle = c\langle y|x\rangle$$
6. $$\langle y|x\rangle = \overline{\langle x|y\rangle}$$
7. and
$$\langle x|x\rangle > 0 \qquad (|x\rangle \neq 0)$$

Ket vectors may be multiplied by complex numbers and added to one another to form another

$$c_1|A\rangle + c_2|B\rangle = |C\rangle$$

Ket vectors are also be integrable, as they can be taken as the sum of a parameter $|x\rangle$ over a certain range to get another ket vector, i.e.

$$\int |x\rangle \, dx = |Q\rangle$$

Dirac established a one-to-one correspondence between bras and kets, such that the bra corresponding to $|x\rangle$ is $\langle x|$. The bra-ket notation addressed the inadequacy of using ordinary vectors, which were not sufficiently general for most dynamical systems in quantum mechanics, such that any state of the dynamical system at a particular time would be specified by both the bra and ket vectors.

Preserving the Heisenberg formulation, an observable A would have a corresponding eigen-

value λ and an 'eigenket' $|\psi\rangle$ satisfying

$$A|\psi\rangle = \lambda|\psi\rangle. \tag{2.26}$$

For a basis $\{e_i\}_{i=1}^n$, the corresponding ket's are $|e_1\rangle, \ldots, |e_n\rangle$. In general, for $|v\rangle \in \chi$,

$$|v\rangle = \sum_{i=1}^{n} \alpha_i |e_i\rangle \tag{2.27}$$

Now, when put over an orthonormal basis

$$\langle e_j|e_i\rangle = \langle e_j, e_i\rangle = \delta_{ij}$$

Dirac proposed this notation to reconcile the matrix formalism with the wave-mechanical formalism, by noting that with proper normalizing constants, the wave-function ψ corresponded to a state function $|\psi\rangle$ such that *Born's* observation that Schrödinger's wave equation described a probability-wave holds, i.e.

$$\langle\psi|\psi\rangle = \int_{\mathbb{R}} \psi^*\psi = \int_{\mathbb{R}} |\psi|^2 = 1 \tag{2.28}$$

while for any observable A, with operator $\hat{A}$

$$\langle\psi|A|\psi\rangle = \int_{\mathbb{R}} \psi^* \hat{A}\psi$$

corresponds to the expected value of that observable. It should not strike the reader as coincidental that we've used $\langle,\rangle$ to represent the inner-product. This in fact was deliberate, and so we should note that the bra-ket notation possess all the inner-product space properties: in particular, linearity.

Now, let's demonstrate how change of bases are represented in this notation. Consider we have an orthonormal basis given by $\{f_k\}$, where

$$|e_i\rangle = \sum_k A_{ik}|f_k\rangle$$

for $A_{ik} \in \mathbb{C}$. If both bases $\{e_i\}$ and $\{f_k\}$ are orthonormal, then

$$\langle f_j|e_i\rangle = \langle f_j| \sum_k A_{ik}|f_k\rangle = \sum_k A_{ik}\langle f_j|f_k\rangle = A_{ij}$$

This yields the change of basis formula:

$$\begin{aligned} |e_i\rangle &= \sum_k \langle f_k|e_i\rangle |f_k\rangle \\ &= \sum_k A_{ik}|f_k\rangle \\ &= \sum_k |f_k\rangle A_{ik} \\ &= \sum_k |f_k\rangle\langle f_k|e_i\rangle \end{aligned}$$

which implies that $\mathbb{I} = \sum_k |f_k\rangle\langle f_k|$.[1] Thus, for an arbitrary $|v\rangle \in \chi$

$$\begin{aligned}
|v\rangle &= \sum_i v_i |e_i\rangle \\
&= \sum_i v_i \left(\sum_k |f_k\rangle\langle f_k| \right) |e_i\rangle \\
&= \sum_i v_i \sum_k |f_k\rangle A_{ik} \\
&= \sum_i \sum_k v_i A_{ik} |f_k\rangle \\
&= \sum_i \sum_k v_i (\langle f_k|e_i\rangle) |f_k\rangle \\
&= \sum_k \sum_i v_i (\langle f_k|e_i\rangle) |f_k\rangle \\
&= \sum_k v'_k |f_k\rangle
\end{aligned}$$

where $v'_k = \sum_i \langle f_k|e_i\rangle v_i = \sum_i A_{ik} v_i$.

2.6 Von Neumann's Synthesis

It was while he was working as David Hilbert's assistant that von Neumann found himself involved in the revolutions going on in physics. Although he applauded the formal framework Dirac proposed for its elegance and power in applications, von Neumann loathed the Dirac delta function's central importance in that framework. Considering it to be an improper function, with self-contradictory principles[13], von Neumann went about reformulating quantum mechanics within a radically different framework based on Hilbert's theory of operators.

It was von Neumann who recognized the mathematical structure which had eluded the physicists. Although some Hilbert spaces were known, von Neumann recognized that Schrödinger's wave functions corresponded to unit vectors in a Hilbert space $\mathcal{H}$, that Heisenberg's observables were linear operators in ℓ_2, and that the Riesz-Fishcher theorem implied the two were equivalent. His formulation saw that states of the physical system $|\psi\rangle$ are described by Hilbert space vectors and the measurable quantities, the *observables* correspond to Hermitian operators acting upon the state vectors. In short order, von Neumann's brilliant synthesis of the matrix formulation of Heisenberg and the wave mechanical formulation of Schrödinger dispensed with Dirac's delta functions by way of a recognition of the isomorphism between the sequence space of ℓ_2 and the function space of L_2. This isomorphism led him to formulate the notion of an abstract Hilbert space and develop his mathematical theory for quantum mechanics, yielding a mathematically rigourous theory which was as powerful as the heuristics developed by the physicists whose work he synthesized.[6]

von Neumann recognized that the mathematical framework of matrix mechanics required self-

[13] von Neumann was correct about this: the Dirac delta function is actually a *generalized function*, or *distribution*. A distribution is a linear functional F on an open set U such that $\lim_{n\to\infty} F(\varphi_n) = F(\lim_{n\to\infty} \varphi_n)$ for any convergent sequence $\{\varphi_n\} \subset U$.

adjoint (and possibly unbounded) linear operators $\hat{A}$ acting on a possibly infinite dimensional, separable Hilbert space (depending on the nature of the observable). This formulation had the observable identifying pure states of the system associated with the unit vectors satisfying Schrödinger's equation. This formulation naturally buttressed the probability[14] interpretation of the wave function proposed by Born, himself a student of Hilbert. Since von Neumann started with the space of square-integrable functions on the real line, specifying that integration must be defined by the Lebesgue integral, and that the inner-product operation was defined on the equivalence classes of square integrable functions differing on sets of measure zero, a formal axiomatic framework for elementary quantum mechanics linking the two competing formalisms had been found.

In effect, since the quantum state is a square integrable function $\psi \in L_2(\mathbb{R})$, this state can be normalized by dividing by the constant $C = \int_{\mathbb{R}} |\psi|^2$. For a normalized state ψ, the expectation value of an observable A could be derived from $\langle \psi, A\psi \rangle$, and more notably, the transition probability from states ψ, φ could be described by $|\langle \varphi, \psi \rangle|^2$. In this way, von Neumann recognized that the geometry afforded by Hilbert space had a direct physical meaning in quantum mechanics. To see this, notice that $|\langle \varphi, \psi \rangle|^2 = \cos^2(\theta)$, since θ is the angle between unit vectors ψ, φ. This geometry is what allows us to mathematically formalize the uncertainty principle![6]

2.7 The Axioms of Quantum Mechanics

Given in Dirac's bra-ket notation, the axioms of canonical quantization which reconciled Heisenberg's formulations with Schrödinger's can be stated in various ways. The following axioms describe canonical quantization for a single non-relativistic particle[1][4][5][8]:

1. For a quantum system, there exists a Hilbert space $\mathcal{H}$ such that the state of the system is described by a state 'vector' $|\Psi\rangle \in \mathcal{H}$ such that

$$\langle \Psi | \Psi \rangle = 1 \tag{2.29}$$

   That is, we recognize that for each state $|\Psi\rangle \in \mathcal{H}$, there is a corresponding dual $\langle \Psi| \in \mathcal{H}^*$ such that

$$\langle \Psi | \Psi \rangle := \int \Psi^* \Psi = 1 \tag{2.30}$$

   In practical terms, solutions Ψ to Schrödinger's equation are normalized square integrable functions which satisfy

$$\int_\Omega \Psi^* \Psi = \int_\Omega |\Psi|^2 = 1 \tag{2.31}$$

2. The physical quantities A of classical mechanics are replaced by physical observables corresponding to Hermitian linear operators $\hat{A}$ acting on $\mathcal{H}$. The result obtained when A is measured is an eigenvalue of $\hat{A}$ which is guaranteed to be real by the hermiticity of $\hat{A}$. This is where the spectral theorem, Theorem 13, comes into play. These eigenvectors are

[14] The reader should be aware that in the historical context, Kolmogorov had only recently axiomatized probability.

orthogonal by the self-adjointness of the operator, i.e.

$$\langle x|y\rangle = \delta(x-y) \tag{2.32}$$

in the position basis and

$$\langle p|q\rangle = \delta(p-q) \tag{2.33}$$

in the momentum basis. Note that there is no given means for finding the operator corresponding to an observable.[15]

Finally, this suggests that the probability of measuring λ_j, an eigenvalue of observable $\hat{A}$, is given by

$$|\langle\psi_j|\Psi\rangle|^2 \tag{2.34}$$

where $|\psi_j\rangle$ is the associated eigenvector.

3. Given a system whose state vector $|\Psi\rangle \in \mathcal{H}$, there exists a Hermitian linear operator for which $|\Psi\rangle$ is an eigenstate. Furthermore, supposing one prepares many systems in this state, the observation of A in these systems at time t is generally a random variable with an expectation value given by

$$\langle\hat{A}\rangle_t = \langle\Psi|\hat{A}(t)|\Psi\rangle \tag{2.35}$$

4. The operators for position, $\hat{x}$, and momentum, $\hat{p}$, satisfy the following commutation relation[16]

$$[\hat{x},\hat{p}] = i\hbar \tag{2.36}$$

and

$$[\hat{p},\hat{x}] = -i\hbar \tag{2.37}$$

and furthermore, the Poisson bracket of classical mechanics is replaced with the commutator

$$-\frac{i}{\hbar}[\hat{A},\hat{B}]$$

5. The time evolution of a vector space $|\psi_j\rangle$ is given by

$$i\hbar\frac{d}{dt}|\psi_j\rangle = \hat{H}|\psi_j\rangle \tag{2.38}$$

We recognize this as being Schrödinger's equation

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \tag{2.39}$$

The link to what came before and what comes next is that $\mathcal{H}$ can be identified with $L_2(\Omega)$ for some suitable domain Ω, such that $\psi \in L_2(\Omega)$ is a wave function. Then for instance $|\psi(x)|^2$ gives us the probability that a particle is at position x, moreover, by being associated

[15] This is quite important in theoretical quantum mechanics. Everett's Many World's interpretation of quantum mechanics implies that the observed probability of a particle is an observable. However, it is not well-established that there is a way to gauge the 'true' probability distribution of any given function. For a clearer insight into what this debate really entails, the reader is encouraged to study the difference between Bayesian and frequentist approaches to probability and statistics.

[16] A detailed discussion of the derivation of these relations is discussed at length in the companion paper

with a vector space and having observables be strictly limited to Hermitian operators, we can associated to each operator an orthonormal basis of eigenvectors $\{n\}$ so that the associated state $|\psi\rangle$ of an observable can be expressed

$$|\psi\rangle = \sum \lambda_n |\psi_n\rangle$$

As mentioned earlier, we notice that this is intimately tied to the Fourier transform.

Finally, as one concrete example, we see this linear algebra at work for

$$\hat{x}|x\rangle = x|x\rangle$$

where x is an eigenvalue, and $|x\rangle$ is the associated eigenvector.

3 Determining The Probability Interpretation

"*I think I can safely say that nobody understands quantum mechanics.*"- Richard Feynman, Ch. 6, "Probability and Uncertainty"

3.1 The State of our Expectations

The establishment of the equivalence of the Heisenberg and Schrödinger mechanics by an isomorphism between Hilbert spaces has resulted in a spectacularly successful physical theory. Although fundamentally probabilistic in practice, it has been remarkably accurate in terms of its predictive power. Among the most crucial physical implications of these axioms is that the eigenfunctions of the hermitian operators are the *determinate states* of observables. That is to say, where a determinate state is a state in which an observable quantity has a definite, single, measurable value, there is a corresponding eigenfunction.

In this sense the spectrum of our operators describe what we can observe; that our operators must be Hermitian is simply a consequence of what it means to observe by measurement. Our values need to be real, as we cannot take complex measurements simultaneously. Now to get a sense of how the probabilities work, let's first consider an operator $\hat{A}$ with a discrete set of eigenvalues $\{\lambda_n\}$ corresponding with normalized eigenvectors $\{|\psi_n\rangle\}$ (again, this is possible since $\hat{A}$ is Hermitian, and thus, can always be given an orthonormal basis). Then

$$\hat{A}|\psi_n\rangle = \lambda_n|\psi_n\rangle; \langle\psi_n|\psi_n\rangle = 1$$

Thus the probability of $|\psi\rangle$ being in state $|\psi_n\rangle$ at any given time is

$$|c_n|^2 = |\langle\psi_n|\psi\rangle|^2.$$

Definition. The coefficient $\psi(\lambda) \in \mathbb{C}$ is called a *wave function* and a wave function

is precisely the *probability amplitude* of finding an observable at λ in the state $|\psi\rangle$. Formally speaking

$$\psi(\lambda) := \langle\lambda|\psi\rangle$$

In a discrete sum, it can be seen that $\langle\psi_n|\psi\rangle$ is the weight of $|\psi_n\rangle$ in the quantum state $|\psi\rangle$.

However, some observables do not have discrete spectra. In this case, the set of eigenstates are not technically in a Hilbert space, something which both David Hilbert and Dirac endeavoured to resolve, albeit in different forms. Amongst physicists, the solution has been to adopt a *rigged Hilbert space*, which we will define in section 4. The rigged Hilbert space is a construction which links the eigenstates to a continuous spectrum. As a consequence of the spectral theory mentioned at the beginning of this paper, if $\hat{A}$ has a continuous spectrum $\{\lambda\}$, then the state $|\psi\rangle$ is described by

$$|\psi\rangle = \int \psi(\lambda)|\lambda\rangle \, d\lambda$$

from which completeness follows as

$$I := \int \, |\lambda\rangle\langle\lambda| \, d\lambda \tag{3.1}$$

This definition of the identity operator gives us

$$\int |\lambda'\rangle\langle\lambda'|\lambda\rangle \, d\lambda' = |\lambda\rangle$$

which recovers the Dirac delta function under normalization as

$$\langle\lambda'|\lambda\rangle = \delta(\lambda' - \lambda)$$

The coefficient expansion of ψ can be obtained by this normalization condition as

$$\psi(\lambda) = \langle\lambda|\psi\rangle$$

and thus, since $|\psi\rangle$ is assumed to be a unit vector

$$\iint \psi^*(\lambda)\psi(\lambda')\langle\lambda|\lambda'\rangle \, d\lambda \, d\lambda' = \int |\psi(\lambda)|^2 \, d\lambda = 1$$

It follows from this relation that

$$\langle\psi|\hat{A}|\psi\rangle = \int \lambda|\psi(\lambda)|^2 \, d\lambda$$

from which we can get the probability interpretation of the measurement of A. That is, the probability of finding $A \in [\lambda, \lambda + d\lambda]$ is given by $|\psi(\lambda)|^2 \, d\lambda$ and from this, we can derive the probability density

$$\rho(\lambda) = |\langle\lambda|\psi\rangle|^2$$

Concretely, consider the position 'basis', $|x\rangle$, and so the probability density for position is described by

$$\rho(x) = |\langle x|\psi\rangle|^2$$

Hence, we can provide our familiar probability definitions:

Definition. The *expectation* of an observable in state ψ is given as

$$\langle A\rangle = \langle\psi|\hat{A}|\psi\rangle = \int \lambda|\psi(\lambda)|^2\, d\lambda$$

Definition. The *variance* of an observable in state ψ is given as

$$\sigma_A^2 = \left\langle (\hat{A} - \langle A\rangle)\psi | (\hat{A} - \langle A\rangle)\psi \right\rangle = \langle A^2\rangle - \langle A\rangle^2$$

In general, the inner product of two states is given by the wavefunctions as follows:

$$\langle\psi|\varphi\rangle = \int \langle\psi|\lambda\rangle\langle\lambda|\varphi\rangle\, d\lambda \Rightarrow \int \psi^*(x)\varphi(x)\, dx = \int \psi^*(p)\varphi(p)\, dp$$

In this way, we've identified the abstract ket vectors with the definite wavefunctions from before.

Let's consider the following examples to illustrate the power of this formulation:

Example. Momentum In the one dimensional case, let our momentum operator $\hat{p} := -i\hbar\frac{d}{dx}$. Let's first verify this is in fact a Hermitian operator

$$\begin{aligned}
\langle\psi|\hat{p}\psi\rangle &= \int \psi^*(-i\hbar\frac{d}{dx})\psi\, dx \\
&= -i\hbar\left(\psi^*\psi - \int (\frac{d}{dx}\psi^*)\psi\, dx\right) \text{ (Integration by parts)} \\
&= \int (-i\hbar\frac{d}{dx}\psi)^*\psi\, dx (\psi^*\psi \text{ vanishes as } \psi \in L_2(\Omega)) \\
&= \langle\hat{p}\psi|\psi\rangle
\end{aligned}$$

Now given that $\hat{p}$ is a Hermitian operator, we note that $\hbar$, as a real valued had no bearing on the Hermiticity of the operator, and so for simplification, we 'drop' it from further analysis. Finally we see that

$$\hat{p}|\psi\rangle = p|\psi\rangle \Rightarrow -i\frac{d\psi}{dx} = p\psi \Rightarrow \psi(x) = e^{-ipx}$$

We can have a basis given by $|p\rangle$, so that we will have a momentum space representation of our operators. It should be clear from the quantum mechanics axioms that the operator $\hat{p}$ will take out the corresponding p in the $|p\rangle$ basis, i.e.

$$\hat{p}|p\rangle = p|p\rangle$$

and similarly

$$\hat{x}|x\rangle = x|x\rangle$$

It will be useful to show that

- $$\hat{x}|p\rangle = -i\frac{d}{dp}|p\rangle \tag{3.2}$$
- $$\langle p\hat{x}|\psi\rangle = i\frac{d}{dp}\langle p|\psi\rangle = i\frac{d}{dp}\psi(p) \tag{3.3}$$
- $$\langle p|\hat{p}|\psi\rangle = p\langle p|\psi\rangle = p\psi(p) \tag{3.4}$$

These equations follow as:

- (3.1)

  Proof. First, let's define the following unitary operator

  $$\hat{U}(q) = e^{iq\hat{x}}$$

  and then note that from axiom 4 $[\hat{p}, \hat{x}] = -i\hbar$, here simplified to

  $$[\hat{p}, \hat{x}] = -i$$

  which implies

  $$[\hat{p}, \hat{x}^n] = -in\hat{x}^{n-1}$$

  From this we find

  $$[\hat{p}, \hat{U}(q)] = \left[\hat{p}, \sum_{n=0}^{\infty}\frac{(iq)^n}{n!}\hat{x}^n\right] = \sum_{n=0}^{\infty}\frac{(iq)^n}{n!}[\hat{p}, \hat{x}^n] = \sum_{n=0}^{\infty}\frac{(iq)^n}{n!} - in\hat{x}^{n-1} = q\hat{U}(q)$$

  Hence

  $$\hat{p}\hat{U}(q) - \hat{U}(q)\hat{p} = q\hat{U}(q) \Rightarrow \hat{p}\hat{U}(q) = \hat{U}(q)\,(\hat{p} + q)$$

  And thus

  $$\hat{p}\hat{U}(q)|p\rangle = \hat{U}(q)(\hat{p} + q)|p\rangle = \hat{U}(q)(p + q)|p\rangle = (p + q)\hat{U}(q)|p\rangle$$

  and since U is unitary, and thus preserves the norm, it follows that

  $$\hat{U}(q)|p\rangle = |p + q\rangle$$

  Now for infinitesimal q, it follows that

  $$\hat{U}(\varepsilon)|p\rangle = |p + \varepsilon\rangle \approx (1 + i\varepsilon\hat{x})|p\rangle$$

from which we find that

$$\hat{x}|p\rangle = \frac{|p+\varepsilon\rangle - |p\rangle}{i\varepsilon} \overset{\varepsilon\to 0}{\to} -i\frac{d}{dp}|p\rangle$$

Hence

$$\hat{x}|p\rangle = -i\frac{d}{dp}|p\rangle$$

and

$$\langle p|\hat{x} = i\frac{d}{dp}\langle p|$$

(as this is the dual to $|p\rangle$.) Similarly we can show

$$\hat{p}|x\rangle = i\frac{d}{dx}|x\rangle$$

□

- (3.2)

  Proof. $\langle p|\hat{x}|\psi\rangle = i\frac{d}{dp}\langle p|\psi\rangle = i\frac{d}{dp}\psi(p)$ follows from (3.2) as the dual of $\hat{x}|p\rangle$, so

  $$\langle p|\hat{x}|\psi\rangle = (-i\frac{d}{dp}|p\rangle)^*|\psi\rangle = i\frac{d}{dp}\langle p|\psi\rangle = i\frac{d}{dp}\psi(p)$$

  □

- (3.3)

  Proof.

  $$\begin{aligned}\langle p|\hat{p}|\psi\rangle &= \langle p|p|\psi\rangle \text{ (follows from } (\hat{p}|p\rangle)^* = \langle p|\hat{p} \text{)} \\ &= p\langle p|\psi\rangle \\ &= p\psi(p)\end{aligned}$$

  □

Theorem 14. *The probability amplitudes $\langle p|\psi\rangle$ and $\langle x|\psi\rangle$ are related by a Fourier transform.*

Proof. First recall our wavefunction-probability amplitude equivalence such that

$$\psi(p) = \langle p|\psi\rangle$$

$$\psi(x) = \langle x|\psi\rangle$$

Take $|\psi\rangle = |x\rangle$ in the relation

$$(\hat{x}\psi)(p) = \langle p|\hat{x}|\psi\rangle = -\frac{d}{dp}\psi(p)$$

and thus

$$x\langle p|x\rangle = \langle p|\hat{x}|x\rangle = i\frac{d}{dp}\langle p|x\rangle$$

Now, take $|\psi\rangle = |p\rangle$, and thus

$$(\hat{p}\psi)(x) = \langle x|\hat{p}|\psi\rangle = -i\frac{d}{dx}\psi(x)$$

from which we can find

$$p\langle x|p\rangle = \langle x|\hat{p}|p\rangle = -i\frac{d}{dx}\langle x|p\rangle$$

Next it follows that

$$\langle x|p\rangle = Ce^{ipx}$$

as

$$\frac{d}{dx}\langle x|p\rangle = \frac{d}{dx}Ce^{ipx} = ipCe^{ipx} = ip\langle x|p\rangle$$

will satisfy this equation. Moreover, since this is an inner product space, $\langle x|p\rangle$ is conjugate to $\langle p|x\rangle$ and hence

$$\langle p|x\rangle = C^*e^{-ipx}.$$

Now, by our normalization condition

$$\delta(x-y) = \langle x|y\rangle\langle x|\int |p\rangle\langle p|\,dp|y\rangle = \int \langle x|p\rangle\langle p|y\rangle\,dp = |C|^2\int e^{ip(x-y)}\,dp = |C|^2 2\pi\delta(x-y)$$

implies that $|C|^2 = \dfrac{1}{2\pi} \Rightarrow |C| = \dfrac{1}{\sqrt{2\pi}}$ and hence when C is real, $C = \frac{1}{\sqrt{2\pi}}$ and hence

$$\psi(p) = \langle p|\psi\rangle = \int \langle p|x\rangle\langle x|\psi\rangle\,dx = \frac{1}{\sqrt{2\pi}}\int \langle p|x\rangle\psi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int \psi(x)e^{-ipx}\,dx$$

which we recognize as the Fourier transform of $\psi(x)$. □

Crucially, we notice that we can change the basis representing a state as follows

$$|\Psi\rangle = \int \psi(x,t)|x\rangle\,dx = \int \psi(p,t)|p\rangle$$

because

$$\begin{aligned}
\int \psi(p,t)|p\rangle\,dp &= \int \psi(p,t)I|p\rangle\,dp = \iint \psi(p,t)|x\rangle\langle x|p\rangle\,dx\,dp \\
&= \frac{1}{\sqrt{2\pi}}\iint \psi(p,t)e^{ipx}|\rangle\,dp\,dx \\
&= \int\left(\frac{1}{2\pi}\int \psi(p,t)e^{ipx}\,dp\right)|x\rangle\,dx \\
&= \int \psi(x,t)|x\rangle\,dx
\end{aligned}$$

That is, the change of 'basis' is accomplished by a Fourier transform where $\psi(x,t) =$

$\frac{1}{\sqrt{2\pi}}\int \psi(p,t)e^{ipx}\,dp$. This has somewhat profound physical implications, as alluded to throughout this paper: space and time are related by a Fourier transform at the quantum level, and as we shall soon see this fact and the geometry of Hilbert space will give us the uncertainty principle. Now let's look at the eigenvalues and eigenfunctions associated with some familiar operators

Example. Consider the position operator $\hat{x}$, and let $\psi_y(x)$ be an eigenfunction and y be the eigenvalue, i.e.

$$\hat{x}|\psi_y(x)\rangle = y|\psi_y(x)\rangle$$

or equivalently

$$x\psi_y(x) = y\psi_y(x)$$

Since y is a fixed value, while x is a continuous variable, we must consider a function that has the property that

$$x\psi_y(x) = y\psi_y(x)$$

This would imply that $\psi_y(x) = 0$ when $x \neq y$, hence

$$\psi_y(x) = \delta(x-y)$$

and so clearly, the position operator does not have a square-integrable eigenfunctions. Nonetheless, these eigenfunctions are 'complete' in the sense that

$$\psi(x) = \int_{\mathbb{R}} c(y)\psi_y(x)\,dy = \int_{\mathbb{R}} c(y)\delta(x-y)\,dy$$

and we trivially have 'coefficients'

$$c(y) = \psi(y)$$

This is permissible as these eigenfunctions satisfy Dirac orthonormality (i.e.)

$$\int_{\mathbb{R}} \psi_z^*(x)\psi_y(x)\,dx = |A|^2 \int_{\mathbb{R}} \delta(x-z)\delta(x-y)\,dx = |A|^2\delta(y-z)$$

which entails that $\psi_y(x) = \delta(x-y)$ and $\langle\psi_z|\psi_y\rangle = \delta(y-z)$.

Example. Consider $\hat{p}$, and let $\psi_p(x)$ be the eigenfunction corresponding to eigenvalue p. The general solution is

$$\psi_p(x) = Ae^{ipx}$$

which is not square-integrable, and thus momentum has no eigenfunctions in Hilbert space. This worried von Neumann, but physicists are fine looking only at real eigenvalues, and thus in the one dimensional case, are fine considering

$$\int_{\mathbb{R}} \psi_q^*(x)\psi_p\,dx = |A|^2 \int_{\mathbb{R}} e^{\frac{i(p-q)x}{\hbar}}\,dx = |A|^2 2\pi\hbar\delta(p-q)$$

which entails that $|A| = \frac{1}{\sqrt{2\pi}}$.

We now see the problem which disturbed von Neumann. When an observable has a continuous spectrum, the eigenfunctions are not normalizable, nor are they in readily in a Hilbert space, nor can they represent a physical state. However, if our consideration is restricted to the eigenfunctions with real values, we find that these eigenfunctions are *Dirac* ortho-normalizable and 'complete' over the integral given the Dirac delta function, i.e. for $\psi_p(x) = \frac{1}{\sqrt{2\pi}} e^{ipx}$

$$\langle \psi_q(x) | \psi_p(x) \rangle = \delta(p - q)$$

and

$$\psi(x) = \int_{\mathbb{R}} c(p)\psi_p(x)\, dp = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} c(p) e^{ipx}\, dp$$

where $c(p) = \langle \psi_q | \psi \rangle$. This would entail that we are working within a *rigged Hilbert space*, which we will define in section 4.

This is meant to illustrate the important, albeit VERY subtle point regarding the axioms of quantum mechanics mentioned at the beginning of section 3: chiefly the difference between states and observables. States ψ must be square-integrable functions, and thus exist in a Hilbert space. The physical meaning of an abstract state under the Copenhagen interpretation of quantum mechanics has the state describing a probability distribution for an outcome. An eigenstate, corresponding to an eigenvector, is called *determinate* if it is normalizable. The term *determinate* is used as it would entail that performing the same observation will yield the same result. One concrete example are the stationary states of the Hamiltonian. Physically speaking, performing a measurement $\hat{H}$ on a particle in state ψ_n certainly yields the corresponding allowed energy E_n (consider the simple harmonic oscillator). On the other hand, observables, the actual 'stuff' of the world, may not have a corresponding normalizable eigenfunctions, and so there may not be a corresponding *determinate* state. In fact, this precisely occurs when we are dealing with observables that have continuous spectra, i.e., with unbounded self-adjoint operators. As we will discuss in section 4, two different approaches were taken to smooth over this difficulty: von Neumann's spectral theory for unbounded operators, and Schwartz's theory of distributions to rigourously define the Dirac delta function.

For now, we direct the reader to consider two familiar examples of this probability formalism:

Example. First consider the case of the simple harmonic oscillator, and find the expected values for the position, momentum and energy operators. Recall our creation and annihilation operators $\hat{a}^\dagger, \hat{a}$ respectively and note that we can use their definition

$$\hat{a}_\pm = \frac{1}{\sqrt{2\hbar m\omega}} (m\omega\hat{x} \mp i\hat{p})$$

to express $\hat{x} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a} + \hat{a}^\dagger)$ and $\hat{p} = -i\sqrt{\frac{m\hbar\omega}{2}}(\hat{a} - \hat{a}^\dagger)$. For simplicity, take $\hbar$ as a natural unit, and so you can simplify the expressions to

$$\hat{x} = \sqrt{\frac{1}{2m\omega}}(\hat{a} + \hat{a}^\dagger)$$

and

$$\hat{p} = -i\sqrt{\frac{m\omega}{2}}(\hat{a} - \hat{a}^\dagger)$$

Now, as a 'hint', I will solve a simpler example where we consider the superposition of states ψ_0 and ψ_1 corresponding to the first two energy levels E_0 and E_1. In this case the initial superposition of ψ_0 and ψ_1 can be given as

$$\Psi(x, 0) = \frac{1}{\sqrt{2}}(\psi_0(x) + \psi_1(x))$$

We make this into a time-dependent equation as follows:

$$\Psi(x, t) = \frac{1}{\sqrt{2}}(\psi_0(x)e^{-iE_0t}) + \psi_1(x)e^{-iE_1t}$$

Now notice that

$$\hat{a}\Psi(x, t) = \frac{1}{\sqrt{2}}\psi_0(x)e^{-iE_1t}$$

since $\hat{a}\psi_0 = 0$ by definition. Similarly

$$\hat{a}^\dagger\Psi(x, t) = \frac{1}{\sqrt{2}}((\hat{a}^\dagger\psi_0)e^{-iE_0t} + (\hat{a}^\dagger\psi_1)e^{-iE_1t}) = \frac{1}{\sqrt{2}}(\psi_1(x)e^{-iE_0t} + \sqrt{2}\psi_2 e^{-iE_1t})$$

- **Position**

$$\begin{aligned}
\langle\Psi|\hat{x}|\Psi\rangle &= \int \Psi^*(x, t)\hat{x}\Psi(x, t)\, dx \\
&= \sqrt{\frac{1}{2m\omega}}\int \Psi^*(x, t)\hat{x}\Psi(x, t)\, dx \\
&= \sqrt{\frac{1}{2m\omega}}\int \Psi^*(x, t)(\hat{a} + \hat{a}^\dagger)\Psi(x, t)\, dx \\
&= \sqrt{\frac{1}{2m\omega}}\left(\int \Psi^*(x, t)\hat{a}\Psi(x, t)\, dx + \int \hat{a}^\dagger\Psi(x, t)\, dx\right) = \\
&= \sqrt{\frac{1}{4m\omega}}\left(\int \Psi^*\left(\psi_1 e^{-iE_0t} + (\psi_0 + \sqrt{2}\psi_2)e^{-iE_1t}\right)\right) \\
&= \frac{1}{2\sqrt{2m\omega}}\int (\psi_0^* e^{iE_0t} + \psi_1^* e^{iE_1t})\left(\psi_1 e^{-iE_0t} + (\psi_0 + \sqrt{2}\psi_2)e^{-iE_1t}\right)
\end{aligned}$$

and by orthonormality of states ψ_n (as they are eigenstates with discrete energy

spectra)

$$= \frac{1}{2\sqrt{2m\omega}}\left(e^{i(E_0-E_1)t} + e^{-i(E_0-E_1)t}\right) = \frac{1}{2\sqrt{2m\omega}}\cos((E_1 - E_0)t)$$

and since $E_1 = E_0 + \hbar\omega$, we have by setting $\hbar = 1$

$$\langle x\rangle = \frac{1}{2m\omega}\cos(\omega t)$$

- **Momentum** Similarly, we leave it to the reader to derive this in full, but for the simple case of Ψ given above

$$\langle p\rangle = \langle\Psi|\hat{p}|\Psi\rangle = -\sqrt{\frac{m\omega}{2}}\sin(\omega t)$$

What is extraordinary about this result is that the expectation values of position and momentum are not stable, but *dynamic*, that is, the evolve over time. Moreover,

$$\langle p\rangle = m\frac{d\langle x\rangle}{dt}$$

which is the result we would expect[17] in classical mechanics.

- **Energy** In our simplified Ψ state, it helps to recall that $\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2})$

$$\langle E\rangle = \langle\Psi|\hat{H}|\Psi\rangle = \langle\Psi|(\hat{a}^\dagger\hat{a}+\frac{\hbar\omega}{2})|\Psi\rangle = \langle\Psi|(\hat{a}^\dagger\hat{a})|\Psi\rangle+\langle\Psi|\frac{\hbar\omega}{2}|\Psi\rangle = \langle\Psi|(\hat{a}^\dagger\hat{a})|\Psi\rangle+\frac{\hbar\omega}{2}$$

and so

$$\hat{a}^\dagger(\hat{a}\Psi) = \hat{a}^\dagger\left(\frac{1}{\sqrt{2}}\psi_0 e^{-iE_1t}\right) = \frac{1}{\sqrt{2}}\psi_1(x)e^{-iE_1t}$$

and so by orthonormality of the eigenvectors ψ_n, we have

$$\langle\Psi|\hat{a}^\dagger\hat{a}|\Psi\rangle = \frac{\hbar\omega}{2}$$

and thus

$$\langle E\rangle = \frac{\hbar\omega + \hbar\omega}{2}$$

or simplified to

$$\langle E\rangle = \hbar\omega$$

which is precisely

$$\langle E\rangle = \frac{E_0 + E_1}{2}$$

Example. We leave it to the reader to consider the case of the free particle, and find the expected values for the position, momentum and energy operators. Recall that we can think of $\Psi(x, 0)$, as a solution to the time-independent Schrödinger equation.

[17] This is more than just a pun

3.2 Clarifying the Uncertainty Principle

After laying out all of the mathematical and physical formalisms necessary for this remarkable result, let us first consider two abstract Hermitian operators $\hat{A}$ and $\hat{B}$, both of which correspond to physical observables acting on a quantum system represented by Ψ. Now let

$$|f\rangle = (\hat{A} - \langle A\rangle)|\Psi\rangle$$

and

$$|g\rangle = (\hat{B} - \langle B\rangle)|\Psi\rangle$$

We notice that $|f\rangle$ and $|g\rangle \in \mathcal{H}$, as they are defined as a linear combination of an eigenfunction and a scalar multiple of the underlying state $|\Psi\rangle$. Thus, by definition,

$$\sigma_A^2 = \langle(\hat{A} - \langle A\rangle)\Psi|(\hat{A} - \langle A\rangle)\Psi\rangle = \langle f|f\rangle = \|f\|^2$$

and

$$\sigma_B^2 = \langle(\hat{B} - \langle B\rangle)\Psi|(\hat{B} - \langle B\rangle)\Psi\rangle = \langle g|g\rangle = \|g\|^2$$

And from this we find that

$$\sigma_A^2\sigma_B^2 = \langle f|f\rangle\langle g|g\rangle = \|f\|^2\|g\|^2 \tag{3.5}$$

As a consequence of being in a Hilbert space, we can apply Cauchy-Schwarz and see that

$$|\langle f, g\rangle|^2 \leq \|f\|^2\|g\|^2 = \sigma_A^2\sigma_B^2 \tag{3.6}$$

It is clear that $\langle f|g\rangle \in \mathbb{C}$, and so we remark that

$$|\langle f|g\rangle|^2 = (\mathrm{Re}(\langle f|g\rangle))^2 + (\mathrm{Im}(\langle f|g\rangle))^2 \geq (\mathrm{Im}(\langle f|g\rangle))^2 = \left(\frac{1}{2i}(\langle f|g\rangle - \langle g|f\rangle)\right)^2$$

since for any $z \in \mathbb{C}$

$$|z|^2 \geq (\mathrm{Im}(z))^2 = (\frac{1}{2i}(z - z^*))^2$$

and so we can now refine our lower bound further to

$$\sigma_A^2\sigma_B^2 \geq \left(\frac{1}{2i}(\langle f|g\rangle - \langle g|f\rangle)\right)^2 \tag{3.7}$$

Now let us consider what $\langle f|g\rangle$ corresponds to in terms of our observables. We find that

$$\begin{aligned}
\langle f|g\rangle &= \langle(\hat{A}-\langle A\rangle)\Psi|(\hat{B}-\langle B\rangle)\Psi\rangle \\
&= \langle\Psi|(\hat{A}-\langle A\rangle)(\hat{B}-\langle B\rangle)\Psi\rangle \\
&= \langle\Psi|(\hat{A}\hat{B}-\hat{A}\langle B\rangle-\langle A\rangle\hat{B}+\langle A\rangle\langle B\rangle)\Psi\rangle \\
&= \langle\Psi|\hat{A}\hat{B}|\Psi\rangle-\langle B\rangle\langle\Psi|\hat{A}|\Psi\rangle-\langle A\rangle\langle\Psi|\hat{B}|\Psi\rangle+\langle A\rangle\langle B\rangle\langle\Psi|\Psi\rangle \\
&= \langle AB\rangle-\langle A\rangle\langle B\rangle
\end{aligned}$$

and similarly

$$\langle g|f\rangle = \langle BA\rangle - \langle A\rangle\langle B\rangle$$

and so

$$\langle f|g\rangle + \langle g|f\rangle = \langle AB\rangle + \langle BA\rangle - 2\langle A\rangle\langle B\rangle$$

while

$$\langle f|g\rangle - \langle g|f\rangle = \langle AB\rangle - \langle BA\rangle = \langle[\hat{A},\hat{B}]\rangle \tag{3.8}$$

and so, by substituting (3.8) into (3.7), our general uncertainty principle can be written

$$\sigma_{\hat{A}}^2\sigma_{\hat{B}}^2 \geq \left(\frac{1}{2i}\langle[\hat{A},\hat{B}]\rangle\right)^2 \tag{3.9}$$

and so we see that our uncertainty about a system has a lower bound related to the degree that it fails to commute! In the special case where $\hat{A}=\hat{x}$ and $\hat{B}=\hat{p}$, we find by Axiom IV,

$$\sigma_{\hat{x}}^2\sigma_{\hat{p}}^2 \geq \left(\frac{1}{2i}\langle[\hat{x},\hat{p}]\rangle\right)^2 = \left(\frac{1}{2i}\langle\Psi|i\hbar|\Psi\rangle\right)^2 = \left(\frac{1}{2i}i\hbar\langle\Psi|\Psi\rangle\right)^2 = \left(\frac{\hbar}{2}\right)^2 = \frac{\hbar^2}{4}. \tag{3.10}$$

We can then take the square-root of (3.9) to find

$$\sigma_{\hat{x}}\sigma_{\hat{p}} \geq \frac{\hbar}{2} \tag{3.11}$$

This can be thought of probabilistically as saying *the uncertainty in position multiplied by the uncertainty in momentum is at least one half of a planck constant.* Physically speaking, there is a lower bound to how finely we can know the universe in terms of position and momentum.

4 Nothing Is Perfect: A Tale of Two Competing Formalisms

"*It seems clear that the present quantum mechanics is not in its final form.*"- Paul Dirac, "The Early Years of Relativity" in *Albert Einstein: Historical and*

4.1 Just Because It Works, Doesn't Mean It's Right: von Neumann's Obsession with Rigour

As mentioned earlier in the paper, von Neumann had some serious doubts about the formalism which he had proposed. At first, von Neumann worked to find a mathematical structure of continuous geometries to supersede the Hilbert space formulation. However the framework he pursued could not generalize the separable Hilbert space framework. Instead, it strengthened the Hilbert space formulation, while giving physicists and mathematicians useful tools to study in their own right. Chief among them were his research into rings of operators. Without extensively covering this research, we will provide a few definitions and proceed to note that this led to the theory of C^* algebras.

4.1.1 von Neumann Algebras

Definition. A $*-$ *ring* is an associative ring A with map $* : A \to A$ which is both an antiautomorphism and an involution, i.e. for all $x, y \in A$

1. $1^* = 1$
2. $(x^*)^* = x$
3. $(x + y)^* = x^* + y^*$
4. $(xy)^* = y^* x^*$

Definition. A $*-$algebra A is a $*-$ring with involution $*$ that is an associative algebra over a commutative $*-$ring R with involution $\dagger$ such that for all $r \in R, x \in A$

$$(rx)^* = r^\dagger x^*$$

Definition. A *von Neumann algebra* or W^*-algebra is a $*-$algebra of bounded operators on a Hilbert space that is closed in the weak operator topology and contains the identity operator.

Definition. A von Neumann algebra W whose centre consists only of the identity operator is called a *factor*. von Neumann showed that every such algebra W on a separable Hilbert space is isomorphic to a direct integral of factors whose decomposition is essentially unique. This led to the problem of classifying isomorphism classes of von Neumann algebra's on separable Hilbert spaces, which he and Murray were able to break into three types.

1. A factor is type I if there is a minimal projection operator $E \neq 0$ such that there is no other projection F with $0 < F < E$. A subtype of type I operators are the type I_n operators , which correspond to finite dimensional Hilbert spaces, while type I_∞ corresponds to an infinite dimensional, separable Hilbert space. The bounded operators on a separable, infinite dimensional Hilbert

space are type I_∞.

2. A factor is type II if there are no minimal projections, but there are non-zero, finite projections. A type II_1 factor F is the unique, smallest infinite dimensional factor contained in other infinite dimensional factors such that any infinite dimensional factor contained inside F is isomorphic to F. It is worth noting that while von Neumann came to regard Type II_1 factors as the proper generalization of type I_n factors, rather than the type I_∞ factor, this approach has not been as preferred by physicists as the use of *distributions* and *generalized functions*.
3. A factor is type III if it does not contain any nonzero, finite projections. Although Murray and von Neumann were not able to distinguish subtypes for type III factors, and in fact, questioned if they even existed, subtypes have been found since 1936, when they were first introduced.

4.2 Formalizing the Pragmatic: The Theory of Distributions

While von Neumann worried about the formalism he proposed, in part because he insisted that it was necessary for every experimental observer to have an a priori probability distribution, Paul Dirac got along quite well with this framework, and by the 1940s found that his Delta function was in fact a real mathematical object, a *generalized function* whose emerging theoretical background, the *theory of distributions*, strengthened the Dirac orthonormalizable formulation of quantum mechanics by placing it inside a *rigged Hilbert space*.

Definition. A vector space V over k is locally convex if it is defined in terms of semi-norms. Recall a semi-norm on V is a map $\rho : V \to \mathbb{R}$ such that

1. (Non-negative) $\rho(x) \geq 0$
2. (Homogeneous) $\rho(cx) = |c|\rho(x)$
3. (Subadditive) $\rho(x + y) \leq \rho(x) + \rho(y)$

Definition. Ω is a paracompact, smooth manifold if

- Every open cover of Ω has an open refinement which is locally finite.
- All transition maps on Ω are smooth.

Definition. For an open subset $U \subset \Omega$, where Ω is a paracompact, smooth manifold. let $\{U_i\}$ be a countable, nested family of open subsets of U with compact closures given as $C_i = \bar{U}_i$.
Let S_i be the family of smooth functions on U with support lying in C_i.
Now let $D(U) = \bigcup_{i \in \mathbb{N}} S_i$. We call $D(U)$, the domain of a distribution, and it consists smooth maps. It can be given the following limit-topology, where a family of functions $\{\varphi_n\} \subset D(U)$ converges to φ provided:

- If there is compact support for all $\{\varphi_n\}$, so that $C \subset U$ contains the support for all $\{\varphi_n\}$, i.e.

$$\bigcup_{n\in\mathbb{N}} \text{supp}(\varphi_n) \subset C$$

- For each multi-index α, the sequence of partial derivatives $D^\alpha\varphi_n$ tends uniformly to $D^\alpha\varphi$

This allows us to consider $D(U)$ to be a complete locally convex topological vector space with the Heine-Borel property (i.e., it is closed and bounded, and compact).

Notation. We denote the space of all distributions on U by $D'(U)$.

Definition. *A distribution* or *generalized function* on U is a linear functional $F : D(U) \to k$, where $k = \mathbb{R}$ or $\mathbb{C}$ such that

$$\lim_{n\to\infty} F(\varphi_n) = F(\lim_{n\to\infty} \varphi_n)$$

for any convergent sequence of $\{\varphi_n\} \subset D(U)$. So, a generalized function belongs to a class of linear functionals, denoted d, which map the smooth test functions into $\mathbb{R}$ or $\mathbb{C}$. Instead of writing $F(\varphi)$, we write $\langle F, \varphi\rangle$.

With the space of distributions on an open subset defined, we can now define a *rigged Hilbert space.*

Definition. A *rigged Hilbert space* is a construction to link the distribution and square-integrable aspects of functional analysis. First introduced to study spectral theory, they are used to bring together a bound state (eigenvectors) and a continuous spectrum. The formal defintion comes courtesy of Israel Gelfand, and is given as a pair $(\mathcal{H}, \mathcal{X})$, where $\overline{\mathcal{X}} = \mathcal{H}$, and $\mathcal{X}$ is given a topological vector space structure for which the inclusion map ι is continuous.
Then, when identifying $\mathcal{H}$ with $\mathcal{H}^*$, the adjoint to ι is the map

$$\iota^* : \mathcal{H}^* \to \mathcal{X}^*$$

The duality pairing of $\mathcal{X}, \mathcal{X}^*$ must be compatible with the inner product of $\mathcal{H}$, i.e.

$$\langle u, v\rangle_{\mathcal{X}\times\mathcal{X}^*} = \langle u, v\rangle_{\mathcal{H}}$$

whenever $u \in \mathcal{X} \subset \mathcal{H}$ and $v \in \mathcal{H} = \mathcal{H}^* \subset \mathcal{X}^*$. The term *rigged* was chosen because of the inclusion relation

$$\chi \subset \mathcal{H} \subset \chi^*$$

The study of rigged Hilbert spaces emerged out of abstract functional analysis, where $\chi \subset \mathcal{H}$ was a subspace of $\mathcal{H}$ with a finer topological structure.

Definition. The dual χ^* is realized as a *space of distributions and the linear*

functionals on χ such that for $x \in \mathcal{H}$

$$\varphi \mapsto \langle x, \varphi \rangle$$

are faithfully represented as distributions by the requirement that χ is dense in $\mathcal{H}$.

Finally, it's worth mentioning that $\mathcal{H}$ is self-dual as a consequence of the Riesz representation theorem. This formalization is what makes Dirac orthonormality mathematically sound, and is the preferred basis to von Neumann's approach.

5 Afterword: Comments on the sources, and the Axioms of Canonical Quantization

The vast bulk of this paper was drawn from 4 primary sources. A vast majority of the working examples, as well as the historical context was drawn from my course notes and the assignments from Brian Greene's year long course on the Mathematics of Quantum Mechanics[1], as well as from Byron-Fuller's unabridged *Mathematics of Classical and Quantum Physics*[4]. Furthermore, some examples were drawn from David Griffiths *Introduction To Quantum Mechanics*[5], which takes a very applications based approach to the material. In some instances, his exposition was the cleanest regarding how to solve the wave equation. Finally, some of the discussion regarding the Lagrangian and Hamiltonian formulation were best summarized by Nakahara's book[8]. What's more intriguing is that all four sources gave completely different accounts of these fundamental axioms, or rather, different wordings and orderings of the axioms. Through the remaining sources were read in preparation for this paper, I wrote out my own version of these axioms. Although the axioms as presented are closest in ordering to the axioms as given in Brian Greene's course, the language and notation is closer to Nakahara's[8]and Dirac's[3]. For those interested in von Neumann's notation, Byron and Fuller may be the most comprehensive source.

References

[1] Alexander Berenbeim from *Brian Greene's Mathematics of Quantum Mechanics course*, Columbia University, 2011-2012 Departments of Mathematics and Physics.

[2] Bruckner, Bruckner, and Thompson *Real Analysis* 1997, Prentice-Hall

[3] Paul Dirac *The Principles of Quantum Mechanics*, Fourth Edition, 1958, Oxford

[4] F.W. Byron, Jr. and Robert W. Fuller *Mathematics of Classical and Quantum Physics, Vol 1 & Vol 2* Dover Publications, Inc. New York

[5] David J Griffiths *Introduction to Quantum Mechanics* Second Edition, 2005, Pearson, Prentice Hall

[6] Léon van Hove *von Neumann's Contributions to Quantum Theory*, 1992 Bulletin of the American Mathematical Society, **64** ,3, Part 2,(1958), 95-99

[7] N.P. Landsman *2006 Lecture Notes on Hilbert Spaces and Quantum Mechanics*, Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen Toernooiveld 1, http://www.ru.nl/∼landsman/HSQM.html

[8] Mikio Nakahara *Geometry, Topology and Physics*, Second Edition, 2003, Taylor & Francis Group

[9] C.E. Rickart *General Theory of Banach Algebras* 1960 D. Van Nostrand Company, Inc.

[10] Elias M. Stein & Rami Shakarchi *Fourier Analysis: An Introduction* Princeton Lectures in Analysis I, Princeton University Press

[11] Peter Woit, *Quantum Mechanics for Mathematicians: 2012 Course Notes*, Department of Mathematics, woit@math.columbia.edu, 2012