Hilbert Space Theory and Applications in Basic Quantum Mechanics

by

Matthew R. Gagne

Mathematics Department
California Polytechnic State University
San Luis Obispo
2013

Abstract

We explore the basic mathematical physics of quantum mechanics. Our primary focus will be on Hilbert space theory and applications as well as the theory of linear operators on Hilbert space. We show how Hermitian operators are used to represent quantum observables and investigate the spectrum of various linear operators. We discuss deviation and uncertainty and briefly suggest how symmetry and representations are involved in quantum theory.

APPROVAL PAGE

TITLE: Hilbert Space Theory and Applications in Basic Quantum Mechanics
AUTHOR: Matthew R. Gagne
DATE SUBMITTED: June 2013

_____	_____
Senior Project Advisor	Senior Project Advisor
(Professor Jonathan Shapiro)	(Professor Matthew Moelter)

_____	_____
Mathematics Department Chair	Physics Department Chair
(Professor Joe Borzellino)	(Professor Nilgun Sungar)

Contents

Chapter 1

Introduction and History

The development of Hilbert space, and its subsequent popularity, were a result of both mathematical and physical necessity. The historical events and individuals responsible for the topics we will cover make up an interesting story. It can be told as a part of the history of mathematics or as an integral stage in the development of modern physics. Here we will briefly summarize the tale from both perspectives. The majority of what is to come is told according to the historical note at the beginning of [LB&PM] and also summarizes some of the events covered in [CMMC]. At the end the history there will be a brief note of these and other references for the material that will eventually be covered.

1.1 Physics

The theory of physical law developed by Newton in the 1700's had carried the sciences through centuries of progress and successes, but by the early 1900's it had become apparent that Newton's laws, in fact, would not stand the test of time. With the development of quantum mechanics came the need for a new "setting" or mathematical structure for the description of physical systems. Hilbert space would turn out to satisfy this need and the resulting interest in this new field of mathematical physics has led to nearly a century of increasingly successful theory and experimentation.

The progression of experimental realizations and theoretical advancements that led to our modern formulation of quantum mechanics can be broken into several key stages.

The origin of the theory is often considered to be a paper presented by Max Planck on December 14, 1900 entitled "On the Law of Distribution of Energy in the Normal Spectrum." In this paper Planck presented his work on the law of blackbody radiation. His conclusion was that energy could no longer be treated as continuous variable. Planck observed that the energy of a vibrating system could not change continuously, but was limited to an integral multiple of so

called energy quanta. He coined the term *quantum discontinuity* and defined the proportionality constant in his energy expression as a new universal constant h, referred to now as Planck's constant.

This led Einstein to develop his well known theory of the photoelectric effect around 1907 which describes the process by which an oscillating particle absorbs and emits quanta of radiation or photons. This theory incorporated the newly discovered Planck's constant in the energy, E, for light quanta as $E = hf$ where f is the frequency of the photon being considered. Einstein's theory of the photoelectric effect was also supported by previous experiments done by Millikian's and was taken as a sign that the mystery of atom was, and likely would continue, to gradually diminish.

The work of Planck had created an intense excitement in the physics community, but with so many fundamental notions beings challenged and century old theories becoming obsolete overnight, there was also substantial confusion. The structure of the atom was still a mystery, as were many of its other characteristics. It was at least known that atoms had a neutral charge so with the discovery of the electron, due to J.J. Thompson, the "plum-pudding" model of the atom was proposed. According to this tentative model, the negatively charged electrons were stuck in some positively charged distribution, resulting in an overall neural charge. This model was quickly improved when, in 1911, E. Rutherford demonstrated through experiment that the positive charge of an atom is not evenly distributed, but rather densely concentrated at the core of the atom in what is now called the nucleus.

Even though the familiar "planetary system" model was now realized there was still no explanation for the discrete atomic spectrum. With this model, according to classical electrodynamics, electrons would fall and collide with the nucleus since their motion would result in a continuous emission spectrum. The answer to this problem would be found by a young Danish physicist Neils Bohr, who, as it turns out, had worked in Rutherford's laboratory.

According to [CMMC], "Bohr's leading role in the development of atomic theory began in 1913" when Bohr published his work in a paper titled “On the constitution of atoms and molecules”. Bohr observed that atoms exhibited a stability with respect to external perturbation that could not be explained by classical mechanics or electrodynamics. He wanted to preserve Rutherford's atomic model resembling a planetary system, but at the same time explain the stability and discrete spectrum phenomena that classical theory failed to account for. He did this by providing two assumptions under which this bizarre behavior could possibly be explained. First, Bohr introduced the notion of stationary states. He assumed that an atomic system can only exist permanently in a certain series of states corresponding to a series of discrete and discontinuous energy levels of the system. Additionally he added that the transition between these so-called stationary states could only occur upon the emission or absorption of a energetic light quanta described by Einstein's photoelectric effect. His second assumption then was that the difference in energy levels between stationary states was equal to the energy of the absorbed or emitted radiation associated with the transition and that each transition had a frequency

f which defined this energy. That is, each transition has a frequency f such that

$$E_1 - E_2 = hf$$

gives the energy difference hf between two states with energy E_1 and E_2.

More specifically, for the hydrogen atom, Bohr's revolutionary new assumptions meant that the angular momentum of an electron orbiting the atom could only take on certain, discrete, values. In other words, the angular momentum is quantized, given by the relation

$$L = n\hbar, n = 1, 2, 3, ...$$

where $\hbar$ is the reduced Plank constant $\frac{h}{2\pi} = \hbar = 1.054572 \times 10^{-34} J \cdot s$. Additionally, Bohr assumed Newton's second law and $L = mvr$ for an orbiting electron, the usual relation governing the dynamics of circular orbits. This ultimately led to the relation

$$E_n = \frac{-13.6eV}{n^2}, \ n = 1, 2, 3, ...$$

governing the allowed energy of the orbiting electron.

The novelty of this new conception of atomic systems cannot be understated. Bohr's work sparked enormous interest in this area and a resurgence in work on what is now know as the "old" quantum theory. The First World War hindered progress, but did not stop it. During this time Arnold Sommerfield investigated some of the implications of Bohr's assumptions and invented several new quantum numbers as a means of explaining the fine structure of hydrogen.

The end of the war allowed physicists to refocus their efforts on the rapidly expanding field of quantum theory. In 1922 Stern and Gerlach demonstrated the space quantization rule by demonstrating that the magnetic moment of silver atoms could only take on two positions, or a silver atom is a $spin - \frac{1}{2}$ particle.

With the sudden eruption of interest in and discovery of rules describing atomic behavior, many physicist looked forward to obtaining detailed solutions to outstanding problems such as the Zeeman Effect or a coherent description of the helium atom. However, they would soon realize they were drastically underestimating just how exotic the nature of what they were trying to describe truly was. While the successes of Bohr's atomic model had been many, the space quantization rule, as well as Bohr's assumptions and other realizations, were nothing more than convenient descriptions of observations which allowed more for prediction and description rather than any sense of explanation. Additionally, what is know as Bohr's Correspondence Principle promoted a rather restrictive notion of atomic behavior as it held that any quantum theory must agree with classical results in the limit of large energies and large orbits. This inhibited the realization that a radical change in the foundation of physics would be necessary to have any hope of successfully explaining the range of phenomena they hoped to understand.

Physicists were attempting to describe quantum behavior with classical language. In a way, it was like trying to understand a dead language; translation is possible, but the essence and cultural significance of the phrases are often

lost. Similarly, the physicists of the early 1920's were attempting to translate observed quantum behavior in terms of classical ideas and from a classical point of view. But, they did not speak the language of quantum mechanics and so they could not truly understand it. Once they had realized the inadequacy of Bohr's model, and their whole approach really, physicists set their sights on developing some kind of mathematical formalism which would allow more freedom and be unencumbered by classical restrictions.

This originally took two forms and is explored in depth in [CMMC]: one developed by Heisenberg known as matrix mechanics and the other by Schrödinger known as wave mechanics. Most important characteristics of the modern formulation of quantum mechanics can be found fairly explicitly in one of the two theories. The most fundamental assumption of Heisenberg's theory was that atomic physical observables were not dependent on continuous variables, but rather on discrete quantities like the natural numbers as they label orbits in Bohr's model. With this in mind, he replaced the familiar classical function $f(x, p)$ of position x and momentum p with a function of natural numbers $f(m, n)$. To ensure energy conservation he postulated a multiplication rule for observables. One consequence of this rule was that observables in quantum mechanics do not necessarily commute. This is a fundamental part of the modern formulation as well. de Broglie's idea that each particle had an associated wave-like character was at the center of Schrödinger's wave mechanics. Additionally, he supposed that each wave would need to satisfy some equation similar to the fundamental wave equation or Maxwell's equations. He proceeded to study the transition between wave and geometric optics, erroneously assuming that the transition between classical and quantum mechanics should be similar, and ultimately found the equation which is now named after him.

Ultimately it was Von Neumann, while a student of Hilbert, who realized the equivalence of wave mechanics and matrix mechanics and proposed the current formulation. However, the argument at the time was largely heuristic and unconcerned with mathematical detail. In [CMMC] he argues that the bulk of the equivalence proof were and have been historically taken for granted. The reader can consult [CMMC] for more on the technicalities of equivalence proofs.

While the theoretical and conceptual leaps that made the formulation of quantum mechanics possible came largely from Bohr, Heisenberg, and Schrödinger, Von Neumann is responsible for the last unifying mental leap and also for many of the technical mathematical achievements that accompanied it. We discuss Von Neumann and the mathematical perspective of these events next.

1.2 Mathematics

One of the most interesting contextual facts regarding the prolific surge in mathematics that accompanied the development of quantum mechanics is that the concept of a finite dimensional vector space was less than a hundred years old by the time Von Neumann and others realized the necessity of the mathematical formalism we call Hilbert space. A fairly imprecise notion of a finite dimen-

sional vector space was used by Grassman in his work between 1844 and 1862. It would take the work of Riemann, Weierstrass and others before the idea of function spaces would eventually emerge in its modern form. The events in the history of mathematics which led to the emergence of Hilbert space and its applications are arguably less exciting than the story from the physics perspective but there are still several key figures and turning points worth mentioning. Like most areas of mathematics, the time-line of the actual appearance of certain ideas and concepts is not exact. Many mathematicians were working on similar problems at similar times, or the significance of their work was not realized until years after their death, so we will try to credit whoever is historically considered appropriate while still acknowledging the work of others whose importance was not realized at the time.

Hilbert space gives a means by which one can consider functions as points belonging to an infinite dimensional space. The utility of this perspective can be found in our ability to generalize notions of orthogonality and length to collections of objects (i.e. functions) which don't naturally suggest the consideration of these properties. While Hilbert space eventually turned out to be the desired setting for quantum mechanical systems, its applications to physics were not the only motivation for its conception.

The concept of a space was actually suggested by mechanics in the form of the coordinates of a dynamical system depending on arbitrary coordinates (e.g. position $(q_1, ..., q_n)$). This implicit coordinate space appears as early as 1788 in work done by Lagrange. By the 19^{th} century, mathematicians were becoming increasingly interested in studying various concepts of space. Besides Grassman, Reimann also was one of the first to work seriously with the notion of a space, providing an early conception of a manifold. But without a vector space structure or metric, Riemann's notion of space wasn't sufficient to describe all the useful attributes of functional spaces. Crucial steps were made towards achieving this description soon after the work of Riemann, by Weierstrass and Hadamard.

In 1885 Weierstrass considered the distance between two functions, essentially proving that polynomials are dense in the space of continuous functions (with respect to topology of uniform convergence). Then, in 1897 Hadamard presented a paper at the First Congress of Mathematicians in Zurich which called attention to the significance of the possible connections between sets of functions and ideas of set theory developed by Cantor. Hadamard's suggestion marked a major shift in the way mathematicians thought about collections of functions. By considering spaces or sets whose points or elements are functions, known set theoretic results could be combined with analytic ones to invent ways of describing these new structures. Eventually Frechet (as a student of Hadamard) defined what is now called a metric in his Ph.D. thesis of 1906. Hausdorff defined a topological space which eventually led to the concept of a topological vector space.

The most familiar vector space for many is $\mathbb{R}^n$. $\mathbb{R}^n$ is itself a function space as we can view it as the space of all functions $f : \{1, ..., n\} \to \mathbb{R}$ under the identification $x_1 = f(1), ..., x_n = f(n)$. The vector space addition of points in

$\mathbb{R}^n$ obviously corresponds to pointwise addition of functions. We can enrich the structure of $\mathbb{R}^n$ by defining the norm $\|f\|$ or length of a vector. This allows us to further consider the distance between two vectors f and g by the quantity $\|f-g\|$. As mentioned above, this generalized notion of length available in a normed vector space is what motivated the development of these structures and the eventual considerations of vector spaces whose points are functions. The vector space structure also allows for a determination of the angle between two vectors f and g through the definition of an inner product. In $\mathbb{R}^n$ this takes the form

$$\langle f, g\rangle = \sum_{k=1}^{n} f(k)g(k) \tag{1.1}$$

which defines the a norm by

$$\|f\|^2 = \langle f, f\rangle .$$

The inner product also defines the angle θ between two vectors f and g through the relation

$$\langle f, g\rangle = \|f\| \, \|g\| \cos\theta.$$

In particular this relation gives us a notion of orthogonality by defining vectors f and g to be orthogonal whenever $\langle f, g\rangle = 0$. This also allows us to define orthogonal projections onto vector subspaces of $\mathbb{R}^n$.

Now after viewing $\mathbb{R}^n$ as a space of functions, we want to generalize this idea to an infinite dimensional setting. The is done most simply by replacing the set $\{1, ..., n\}$ above with $\mathbb{N}$ and defining the space of all functions $f : \mathbb{N} \to \mathbb{R}(or\ \mathbb{C})$. However, now if we try to define an inner product on this space as before, replacing 1.1 with

$$\langle f, g\rangle = \sum_{k=1}^{\infty} \overline{f(k)}g(k),$$

we are faced with the obvious problem that this sum will usually not converge. This motivated the study of the space $l^2(\mathbb{N})$ which is defined as the space of all functions $f : \mathbb{N} \to \mathbb{C}$ for which the above expression does converge. We will discuss this space a bit more later.

The actual specific mathematical motivation for Hilbert space came from the study of integral equations of the form

$$f(x) + \int_a^b dy K(x, y)f(y) = g(x) \tag{1.2}$$

where f, g, and K are continuous, f unknown. While Volterra and Fredholm were the first to study these equations, it was Hilbert who provided the most substantial breakthrough during his work of 1904-1910. By choosing an orthonormal basis $\{e_k\}$ of continuous functions on $[a, b]$ and defining the generalized Fourier coefficients of f by

$$\widehat{f_k} := \langle f, e_k\rangle$$

with inner product

$$\langle f, g\rangle := \int_a^b dx f(x)\overline{g(x)},$$

Hilbert was able to transform 1.2 into an equation of the form

$$\widehat{f_k} = \sum_l \widehat{K}_{kl} f_l = \widehat{g}_l.$$

Parseval's relation was already know from Fourier analysis so Hilbert was able to conclude that

$$\sum \left|\widehat{f_k}\right|^2 = \int_a^b dx \left|f(x)\right|^2,$$

implying $\widehat{f} \in l^2(\mathbb{Z})$.

This led Hilbert and others to study l^2 abstractly. In 1907 the space $L^2([a,b])$ appeared in the work of Reisz and Fischer. In the theorem bearing their names, Reisz and Fischer established an isomorphism between $L^2([a,b])$ and l^2. However this was only realized years later after the work of Von Neumann. Reisz also identified the abstract properties of the norm, and in 1922 Banach axiomized these properties in his thesis.

These new developments in mathematics and the excitement of emerging quantum mechanics prompted Von Neumann to finally provide a definition for abstract Hilbert space in 1927. Eventually people began studying a new class of functions on Hilbert space, what we now call linear maps. The study of linear functionals had begun long before definitions of Hilbert or Banach space were given and applicable results and techniques where readily applied to linear operators.

Most of the basics and lasting Hilbert space formalism came from the work of Von Neumann. Reisz established the abstract concept of what we now call bounded linear operators in 1913. Hilbert and Schmidt studied compact operators before the concept of compact was even well established and Weyl studied what we now call unbounded operators. With this preliminary work established, Von Neumann went on to generalize most of these results beyond recognition in his book of 1932. Much of Von Neumann's work was aimed specifically at advancing quantum theory using the tools of functional analysis and Hilbert space theory. It is the applications of these tools which we will be interested in discussing.

A note on convention and text

As with most applied mathematics, the notation and convention of physicist and mathematicians does not always coincide and these notes will be slightly biased towards traditional mathematical notation.

Some physics literature refers to *The Hilbert space* in place of the plural phrasing used here. The isomorphism between l^2 and L^2 mentioned above is a specific case of the more general fact that all separable infinite dimensional Hilbert spaces are isomorphic. Since a Hilbert space is separable when it has a

countable orthonormal basis, practically every example of Hilbert space used in the early days of quantum mechanics was isomorphic (because they had countable bases). Therefore *The Hilbert* space was seen as appropriate. However, non-seperable Hilbert spaces do appear and are useful in some areas of quantum field theory and since we will not discuss the concept of isomorphism or separability in depth the singular phrasing *The Hilbert space* will not be used.

These notes are meant to be a summary of some of the most basic elements of Hilbert space theory and their application to modern quantum theory. But, this exposition is also meant to provide a collection of suggested references for the material we will discuss. While appropriate references have been made during the discussion, we will take the time here to mention some of the more useful and unique references for related material we will not have time to explore in depth.

An Introduction to Hilbert Space and Quantum Logic by David W. Cohen contains a unique treatment of the material we will discuss and more using the concepts of *manuals* and *refinements.* It provides an interesting introduction to the logic of nonclassical physics and gives many projects for the reader to investigate independently. The book *Group theory and physics* by *S. Sternberg* provides is a fairly broad and comprehensive presentation of many application of group theory and other algebraic concepts in physics. While we will deal with little to no algebra in our main discussion, parts of [SS] would constitute an appropriate sequel to these notes. It deals with symmetries and representations and provides some algebraic shortcuts for some of the methods we will cover.

The symmetry discussed in the last section of these notes is a good starting point for more advanced quantum field theory. Representation theory as well as the theory of Lie algebras becomes of central importance in many areas. More in depth discussions of the applications of these algebraic concepts can be found in [EJ] and [HS]. More advanced material is covered in [NPL2] and [JM].

There are a plethora of mathematical texts which cover the mathematics we will go over here but there are a couple worth mentioning briefly. [JW] gives a rigorous treatment of the theory of linear operators and [WR] is a classic reference for all basic functional analysis.

For anything you ever wanted to learn about linear operators, see [ND&JS].

There is no shortage of text books out there but hopefully the texts just mentioned can provided anyone interested with useful references for a variety of material.

Chapter 2

Hilbert Space definitions and examples

We will begin with the definition of a vector space

Definition 1 *A **vector space** V is a collection of vectors together with a field $\Bbbk$ and a corresponding addition ($V \times V \longrightarrow V$) and multiplication ($\Bbbk \times V \longrightarrow V$) relation such that V is a commutative group under addition and multiplication is both associative and distributive.*

$\mathbb{C}^n$ and $\mathbb{R}^n$ are both vector spaces with addition and multiplication defined in the obvious way and are probably the most familiar to readers. The set of polynomials with degree less than or equal to n, for some fixed $n \in \mathbb{N}$, with real coefficients is a vector space under usual addition and multiplication by elements from $\mathbb{R}$. The space of polynomials of degree less than or equal to $n \in \mathbb{N}$ with complex coefficients is a vector space under usual addition and multiplication by elements from $\mathbb{R}$ but, $\mathbb{C}$ could also be used as the field for multiplication.

In what follows, we will usually take $\Bbbk = \mathbb{C}$.

Out of the endless numbers of vector spaces available, we will ultimately be most interested in the Hilbert space consisting of functions who's squared norm is Lebesgue integrable on a particular interval. This will be more explicitly defined and discussed later. A rigorous understanding of the Lebesgue integral and theory of measure will not be necessary for what follows but some familiarity is encouraged. Besides the standard treatment available in most undergraduate analysis texts, the reader is referred to [JHW] for more on Lebesgue integration.

2.1 Linear functionals

Most people are familiar with the common definition of an inner product space as a vector space V endowed with a mapping, usually $\langle ., . \rangle : V \times V \to \mathbb{C}$, such

that for all $x, y, z \in V$ and $\alpha, \beta \in \mathbb{C}$ the following hold

$$\begin{aligned} \langle x, y \rangle &= \overline{\langle y, x \rangle}, && (2.1) \\ \langle \alpha x + \beta y, z \rangle &= \alpha \langle x, z \rangle + \beta \langle y, z \rangle, \text{ and} && (2.2) \\ \langle x, x \rangle &\geq 0 && (2.3) \end{aligned}$$

where equality holds in the last expression only if $x = 0$. We will introduce a slight variant of this definition using definitions and theorems related to linear functionals. The majority of these definitions were taken from [PRH].

A linear transformation between two vector spaces V and V' is a map $L : V \to V'$ such that

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y) \tag{2.4}$$

for every $x, y \in V$ and all $\alpha, \beta \in \mathbb{C}$. Linear transformations whose range coincides with $\mathbb{C}$ are of particular interests. These maps are called **linear functionals.**

Definition 2 *A **linear functional** on a vector space V is a map $\xi : V \to \mathbb{C}$ which satisfies the following conditions, formally known as:*

*(i) ξ is **additive** i.e. $\xi(x + y) = \xi(x) + \xi(y)$ $\forall x, y \in V$ and,*

*(ii) ξ is **homogeneous** i.e. $\xi(\alpha x) = \alpha \xi(x)$ $\forall x \in V$ and $\forall \alpha \in \mathbb{C}$*

Also, we have the corresponding notion of a conjugate linear functional where condition (ii) is replaced with $\xi(\alpha x) = \overline{\alpha} \xi(x)$.

We will also need to define what is known as a bilinear functional:

Definition 3 *A **bilinear functional** on a vector space V is a function $\varphi : V \times V \to \mathbb{C}$ such that if*

$$\xi_y(x) = \eta_x(y) = \varphi(x, y), \tag{2.5}$$

then, for every $x, y \in V$ the map ξ_y is a linear functional and the map η_x is a conjugate linear functional.

Note that if φ is a given bilinear functional and $\psi(x, y) := \overline{\varphi(y, x)}$, then ψ itself is obviously a bilinear functional. In the case that $\varphi = \psi$ we say that φ is **symmetric.** Finally, we will define the quadratic form induced by a bilinear functional.

Definition 4 *The **quadratic form** induced by a bilinear functional φ on a complex vector space V is the map φ^* defined for each $x \in V$ as*

$$\varphi^*(x) = \varphi(x, x). \tag{2.6}$$

The preceding definitions now allow us to introduce the inner product and notions of distance in a vector space.

2.2 Metric, Norm and Inner product spaces

A Hilbert space is a special type of vector space. Its definition will require the generalized notion of length and distance in a vector space. For this we will require an inner or scalar product ($V \times V \longrightarrow V$) which will lead us to a norm and metric.

Definition 5 *An **inner product** on a complex vector space V is a strictly positive and symmetric bilinear functional $\langle .,. \rangle : V \times V \longrightarrow \mathbb{C}$. An **inner product space** is then a vector space V equipped with such an inner product.*

In physics, what is known as the "bra-ket" notation is often used to describe elements in an inner product space. This refers to identifying elements of an inner product space V as vectors $|\alpha\rangle$ or $|\beta\rangle$ with inner product $\langle \alpha \mid \beta \rangle = \overline{\langle \beta \mid \alpha \rangle}$. This is mostly for aesthetic and notational convenience and will seldom be used here.

If an inner product $\langle .,. \rangle$ is defined on a vector space V then from the definition, $\langle .,. \rangle$ is a strictly positive, symmetric bilinear functional. Then according to definition 4 we may define a quadratic form $\varphi^* = \|.\|^2$ associated with $\langle .,. \rangle$ as $\langle x, x \rangle := \|x\|^2$. Since by definition $\langle .,. \rangle$ must be symmetric and strictly positive, i.e. $\langle x, x \rangle = \overline{\langle x, x \rangle} \geq 0$, we are guaranteed that $\langle x, x \rangle := \|x\|^2$ will define a positive real number $\|x\| = \sqrt{\langle x, x \rangle}$. This number is called the norm or more specifically the norm induced by the inner product $\langle .,. \rangle$. There is also a more general definition of a norm and metric.

Definition 6 *A **norm** is a real valued function $\|.\| : V \to \mathbb{R}$ which satisfies the following conditions.*

1. $\|\lambda x\| = |\lambda| \cdot \|x\|$
2. *if* $x \neq 0, \|x\| \neq 0$
3. $\|x\| + \|y\| \geq \|x + y\|$

   $\forall x, y \in V$

Definition 7 *A **metric** is a function $d : V \times V \to \mathbb{R}$ which satisfies the following conditions:*

1. $d(x, y) = d(y, x)$
2. $d(x, y) = 0$ *iff* $x = y$
3. $d(x, y) + d(y, z) \geq d(x, z)$ *(triangle equality)*

   $\forall\ x, y, z \in V$

   A metric space (V, d) is then a set V equipped with an associated metric $d : V \times V \to \mathbb{R}$

For simple vector spaces like $\mathbb{R}$ or $\mathbb{C}$ we have standard norms $\|x\|_{\mathbb{R}} = |x|$ and $\|z\|_{\mathbb{C}} = |z| = \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}$. We can easily generalize these examples to a norm in $\mathbb{C}^n$ by defining an inner product $\langle x, y\rangle = \sum_{k=1}^{k=n} x_k \overline{y_k}$ for every $x = (x_1, x_2, ..., x_n), y = (y_1, y_2, ..., y_n) \in \mathbb{C}^n$. This inner product allows us to define, for each $x \in \mathbb{C}^n$, the norm

$$\|x\|_{\mathbb{C}^n} = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{k=1}^{n} |x_k|^2} \tag{2.7}$$

which also defines a metric as

$$d(x, y)_{\mathbb{C}^n} = \|x - y\|_{\mathbb{C}^n} = \sqrt{\langle x - y, x - y\rangle} = \sqrt{\sum_{k=1}^{n} |x_k - y_k|^2}. \tag{2.8}$$

Note that $\mathbb{R}^n$ is a subspace of $\mathbb{C}^n$ and we may define an inner product and norm for $\mathbb{R}^n$ in the same manner as we have just done for $\mathbb{C}^n$.

Notice that all these inner products, norms, and metrics defined in $\mathbb{R}^n$ and $\mathbb{C}^n$ are geometrically and intellectually intuitive and feel quite natural. This is most likely due to the fact that we actually, at least up to $n = 3$ for $\mathbb{R}^n$, have a very good idea of what the spaces look like. So, it seems natural to say the distance between two points x and y in $\mathbb{R}$ is the distance of their separation on the real line $|x - y|$, or in $\mathbb{R}^2$ the metric which gives the distance between two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ as the distance of their separation in the plane,

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{2.9}$$

which follows from a simple application of the Pythagorean theorem. This ease of visualization of distance in the lower dimensional spaces is a rare luxury for mathematicians so the standard norms and metrics just discussed really are quite pleasing and can allow one to develop a good intuition about the geometry of the individual spaces after some acquaintance.

But what about the infinite dimensional case? Often times we are dealing with a vector space which does not have an obvious metric or distance between its elements. What then can we say is the distance between the functions $f(x) = x - 1$ and $g(x) = 2x - 1$ in $L^2([0, 1])$? Definition 6 allows us to abstract notions of length in arbitrary vector spaces which may or may not have had some inherent and intuitively obvious notion of distance already. For instance, the space $L^2(\Omega) = \{f : \Omega \to \mathbb{C} \mid \|f\|_2 := \left(\int_\Omega |f|^2 \, d\mu\right)^{\frac{1}{2}} < \infty\}$ permits a norm

$$\|f\|_2 := \left(\int_\Omega |f|^2 \, d\mu\right)^{\frac{1}{2}} \textit{ for all } f \in L^2(\Omega), \tag{2.10}$$

and therefore a metric

$$d(f, g) = \|f - g\|_2 = \left(\int_\Omega |f - g|^2 \, d\mu\right)^{\frac{1}{2}} \tag{2.11}$$

as well. The definition of $\|f-g\|_2$ gives us a way to interpret the distance between any two functions f and g in $L^2(\Omega)$. $L^2(\Omega)$ is an infinite dimensional vector space and one of our most prominent and useful examples of Hilbert space.

We can define a norm on the space $M_n = \{A = (a_{ij}) : a_{ij} \in \mathbb{C}\ i, j \in \{1, 2, ..., n\}\}$ of n by n matrices with complex entries by

$$\|(a_{ij})\| := \max_{i,j} |a_{ij}| . \tag{2.12}$$

This gives us a way of defining the norm of a linear operator by considering its matrix representation and using 2.12. We will also see it is possible to define the norm of an arbitrary linear operator without considering its matrix form. If, however, we want to consider the norm of a linear operator on an infinite dimensional Hilbert space there is one major consideration that one needs to take into account; the issue of boundedness. This is obviously not a problem in the finite dimensional case and of course means that it will only make sense to look at the norm of bounded linear operators.

Although it will take some work, we will see that Hilbert space is useful precisely because it allows us to generalize our traditional notions of an inner product space and norm space to an infinite dimensional setting. Later we will look at what restrictions we need to make on our infinite dimensional spaces in order for a suitable inner product and norm to be defined.

It is easy to define a norm or metric on a vector space which satisfies definition 7 or 6 but does not necessarily coincide with the traditional idea of distance in the given space. The following examples demonstrate the diversity available for functions that satisfy definition 6. We can equip $\mathbb{R}^2$ with the following distinct norms with associated metric which define different geometries on our vector space.

Example 8 *Consider the norms*

$$\|(x_1, x_2)\|_1 = \sqrt{x_1^2 + x_2^2}, \tag{2.13}$$

$$\|(x_1, x_2)\|_2 = |x_1| + |x_2| , \text{ and} \tag{2.14}$$

$$\|(x_1, x_2)\|_3 = \max\{x_1, x_2\} \tag{2.15}$$

defined on $\mathbb{R}^2$. The corresponding unit ball in $\mathbb{R}^2$ with respect to the preceding norms (i.e. $\|(x_1, x_2)\|_1 \leq 1$, $\|(x_1, x_2)\|_2 \leq 1$, and $\|(x_1, x_2)\|_3 \leq 1$.) is easy to visualize, forming a circle, rotated square ($\frac{\pi}{2}$ radians), and a centered square respectively.

Although the norms in example 8 are all distinct functions on $\mathbb{R}^2 \times \mathbb{R}^2$, it turns out that in any finite dimensional space, as far as convergence is concerned, every function satisfying definition 6 is equivalent i.e. defines the same convergence.

For any inner product space with norm $\|x\| = \sqrt{\langle x, x\rangle}$ we have the following well known result.

Theorem 9 ***(Cauchy-Schwartz Inequality)*** *For any two elements x and y in an inner product space with norm $\|x\| = \sqrt{\langle x, x\rangle}$ we have*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| . \tag{2.16}$$

Proof. *Note: if $y = 0$ we are done. Assume $y \neq 0$. Then from the properties of bilinear functionals we may expand the value $\langle x + \alpha y, x + \alpha y\rangle$ like so:*

$$0 \leq \langle x + \alpha y, x + \alpha y\rangle = \langle x, x\rangle + \overline{\alpha}\,\langle x, y\rangle + \alpha\,\langle y, x\rangle + |\alpha|^2\,\langle y, y\rangle \tag{2.17}$$

Now set $\alpha = -\langle x, y\rangle / \langle y, y\rangle$ and multiply each side of 2.17 by $\langle y, y\rangle$ and we have

$$0 \leq \langle x, x\rangle\,\langle y, y\rangle - |\langle x, y\rangle|^2 \tag{2.18}$$

■

If $x, y \in V$ are such that $x = \lambda y$ for some $\lambda \in \mathbb{C}$ we say x and y are parallel, often written $x \parallel y$. It can be easily verified that when $x \parallel y$ we will have $|\langle x, y\rangle| = \|x\| \cdot \|y\|$. We also have the useful result known as the parallelogram law

Theorem 10 ***(Parallelogram Law)*** *For any two elements x and y in an inner product space we have*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \tag{2.19}$$

Proof. *Expand*

$$\begin{aligned} \langle x + y.x + y\rangle &= \|x + y\|^2 && (2.20)\\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle && (2.21)\\ &= \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2 && (2.22) \end{aligned}$$

Similarly we have $\|x - y\|^2 =$

$$\begin{aligned} &= \langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle && (2.23)\\ &= \|x\|^2 - \langle x, y\rangle - \langle y, x\rangle + \|y\|^2 && (2.24) \end{aligned}$$

Now, by adding 2.22 and 2.24 we obtain the parallelogram law. ■

2.3 Convergence and completeness

We will now briefly introduce notions of convergence and completeness in a metric space before explicitly defining Hilbert space.

Definition 11 *Let (x_n) be a sequence in a metric space (V, d). We say $x_n \to x$ when $\lim_{n\to\infty} d(x_n, x) = 0$.*

Definition 12 *A metric space (V, d) is called* ***complete*** *when every Cauchy sequence converges to a limit in V*

With these definitions in mind we may now formally define a Hilbert space:

Definition 13 *A* ***Hilbert space*** *is an inner product space which is complete with respect to the associated metric or norm induced by the inner product. That is, a Hilbert space is an inner product space in which every Cauchy sequence has a limit.*

It is important to note the difference between a metric space which is complete with respect to the norm induced by the inner product (i.e. Hilbert space) and a metric space which is complete with respect to an arbitrary norm. The latter is know as a Banach space. It is always the case that a Hilbert space is a Banach space but the converse is not necessarily true. Consider the space

$$L^p(\Omega) = \{f : \Omega \to \mathbb{C} \mid \|f\|_p := \left(\int_\Omega |f|^p \, d\mu\right)^{\frac{1}{p}} < \infty\} \tag{2.25}$$

where Ω is a space with associated measure μ. $L^p(\Omega)$ is a Banach space for all $0 < p \leq \infty$ however it is only a Hilbert space for $p = 2$. It was said earlier that Hilbert space allows the generalization of euclidean space to an infinite dimensional setting. To do so we must take care in constructing the elements of the space. The most natural generalization of $\mathbb{R}^n$ to a space of infinite dimension would be the set of all infinite real sequences. It is true that this set forms a vector space according to definition 1 but it is not a very useful one. How would one define an inner product or norm on such a space? Continuing with our generalization of the structure of $\mathbb{R}^n$ one may propose an inner product such as

$$\langle x, y \rangle = \sum_{n=1}^{\infty} x_n y_n \tag{2.26}$$

for two sequences $x = (x_n)$ and $y = (y_n)$. However it is often the case that 2.26 will not converge when x and y are just two arbitrary real sequences. Instead, we can restrict our attention to the set of all infinite sequence whose terms are square summable. This space is know as $\ell^2(\mathbb{N}) = \{(a_1, a_2, a_3, ...), a_i \in \mathbb{C} \mid \sum_{n=1}^{\infty} |a_i|^2 < \infty\}$. There is also the related hardy space H^2 consisting of all holomorphic functions f on the open unit disk satisfying

$$\sup_{0<r<1} \left(\frac{1}{2\pi} \int_0^{2\pi} \left|f(re^{i\theta})\right|^2 d\theta\right)^{1/2} < \infty$$

The added structure of an inner product or normed spaces allows us to define two new types of convergence which will be useful in the future.

Definition 14 ***(Strong Convergence)****A sequence of vectors (x_n) in an inner product space V is said to converge* ***strongly*** *to $x \in V$ if $\|x_n - x\| \to 0$ as $n \to \infty$.*

Definition 15 ***(Weak Convergence)*** *A sequence of vectors (x_n) in an inner product space V is said to converge* ***weakly*** *to $x \in V$ if for each $y \in V$, $\langle x_n, y\rangle \to \langle x, y\rangle$ as $n \to \infty$.*

2.4 Basis system and orthogonality

Before we begin our discussion of linear operators in Hilbert space, we need to define an orthonormal basis and review other notions of orthogonality. Another special aspect of Hilbert space is that it allows our common notion of orthogonality in a finite dimensional vector space to be generalized to an infinite dimensional setting.

Recall, given any Hilbert space H with inner product $\langle ., .\rangle$ we say that two vectors $f, g \in H$ are **orthogonal,** often written $f \perp g$, when $\langle f, g\rangle = 0$.

This idea can be extended to the notion of subspaces as well. That is, $M \subset H$ and $N \subset H$, both subspaces of H, are said to be orthogonal, again, often written $M \perp N$, when $\langle f, g\rangle = 0$ for all $f \in M$ and all $g \in N$.

Here is very elementary example of two subspaces which are orthogonal $\ell^2(\mathbb{N})$.

Example 16 *Consider the Hilbert space $\ell^2(\mathbb{N})$ of square summable sequences of elements of $\mathbb{C}$. Let $M_n = \{(a_1, a_2, a_3, ...) \in \ell^2(\mathbb{N}) \mid a_i = 0 \ \forall i \leq n\}$ and $N_n = \{(b_1, b_1, b_1, ...) \in \ell^2(\mathbb{N}) \mid b_j = 0 \ \forall j > n\}$, for some $n \in \mathbb{N}$. Then using the usual inner product for $\ell^2(\mathbb{N})$, for any $f \in M_n$ and any $g \in N_n$ we have:*

$$\langle f, g\rangle = \sum_{i=1}^{\infty} a_i \bar{b}_i = 0 \cdot \sum_{i=1}^{n} \bar{b}_i + 0 \cdot \sum_{i=n+1}^{\infty} a_i = 0 \tag{2.27}$$

So we can see that $M_n \perp N_n$ in $\ell^2(\mathbb{N})$.

There are numerous other examples of orthogonal subspaces, obviously more complicated than these. For now, we present some theorems showing how some properties of orthogonality can manifest in relations between the norm of different vectors.

While it is very well known, it is useful to remind ourselves that for any $f \perp g$, $f, g \in H$, the vector norms of f and g satisfy the Pythagorean Theorem. That is, we have:

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2 \tag{2.28}$$

whenever $f \perp g$ with $f, g \in H$.

Recall that for any subset $A \subset V$ where V is a vector space, span(A) (the span of A) is the set of all finite linear combinations of vectors in A. span(A) is the smallest subspace of V containing A. If it happens that span$(A) = V$ and, also the vectors of A are linearly independent, then A forms a basis for V. That is, if A is a basis for V then every vector in V can be written as a finite linear combination of vectors in A. Now we give the definition of orthogonal and orthonormal systems as given in [LD&PM].

Definition 17 *Let V be an inner product space. A family of vectors $S \subset E$ is called an* ***orthogonal system*** *if, for any $x, y \in S$, $x \neq y$ we have $x \perp y$. Additionally, if it happens that $\|x\| = 1$ for all $x \in S$ we say that S is an* ***orthonormal system.***

I am assuming that the reader is familiar with the concept of a basis for a finite dimensional vector space. The infinite dimensional case is similar, however it is not quite as simple as "a set of linearly independent vectors whose span is equal to the vectors space". However, linear independence is still necessary for a basis in our infinite dimensional setting so we will require the following theorem and the definition of an orthonormal sequence.

Theorem 18 *Orthogonal systems are linearly independent.*
Proof. *Let A be an orthogonal system and suppose $\sum_{k=1}^{k=n} a_k e_k = 0$ for some $e_1, e_2, ... e_n \in A$ and $a_1, a_2, ... a_n \in \mathbb{C}$. Then*

$$\begin{aligned} 0 &= \sum_{m=1}^{m=n} \langle 0, a_m e_m \rangle && (2.29) \\ &= \sum_{m=1}^{m=n} \left\langle \sum_{k=1}^{k=n} a_k e_k, a_m e_m \right\rangle && (2.30) \\ &= \sum_{m=1}^{m=n} |a_m|^2 \|e_m\|^2 && (2.31) \end{aligned}$$

This implies that $a_m = 0 \ \forall m \in \mathbb{N}$ so $e_1, e_2, ... e_n$ are linear dependent ■

Definition 19 ***(Orthonormal Sequence)*** *A sequence of vectors which constitute an orthonormal system is called an* ***orthonormal sequence.***

Now the orthogonal condition for a (x_n) sequence can be expressed in terms of the Kronecker delta symbol:

$$\langle x_n, x_m \rangle = \delta_{nm} \tag{2.32}$$

and $\|x_n\|$ for an orthonormal sequence (x_n) where $\delta_{nm} = 1$ for $n = m$ and $\delta_{nm} = 0$ when $n \neq m$.

Again, this is really not very different from the finite dimensional case. As an example, consider $e_n = (0, ..., 0, 1, 0, ...)$ where the 1 is in the nth position. Then the set $S = \{e_1, e_2, e_3, ...\}$ is an orthonormal system in ℓ^2

Definition 20 *(Complete Orthonormal Sequence) An orthonormal sequence (x_n) in an inner product space V is said to be* ***complete*** *if for every $x \in V$ we have:*

$$x = \sum_{n=1}^{\infty} \langle x, x_n \rangle x_n \tag{2.33}$$

Now we may finally define a basis in a infinite dimensional space as follows and give the following theorem regarding complete orthonormal sequences in Hilbert space.

Definition 21 *(Orthonormal Basis) An orthonormal system B in an inner product space V is called an* ***orthonormal basis*** *(o.n.b) if every $x \in V$ has a unique representation*

$$x = \sum_{n=1}^{\infty} \alpha_n x_n \tag{2.34}$$

where each $\alpha_n \in \mathbb{C}$ and each $x_n \in B$.

When a basis B has been specified we often will write $x = (\alpha_1, \alpha_2, ...)$ where $\alpha_n = \langle x, x_n \rangle$ for every $x_n \in B$.

Theorem 22 *An orthonormal sequence (x_n) in a Hilbert space H is complete if and only if $\langle x, x_n \rangle = 0$ $\forall n \in \mathbb{N}$ implies that $x = 0$.*
Proof. *Suppose (x_n) is a complete orthonormal sequence in H. Then we can write every $x \in H$ as*

$$x = \sum_{n=1}^{\infty} \langle x, x_n \rangle x_n \tag{2.35}$$

So if $\langle x, x_n \rangle = 0$ for every $n \in \mathbb{N}$ then $x = 0$.
Now suppose that if $\langle x, x_n \rangle = 0$ for every $n \in \mathbb{N}$ it follows $x = 0$. Let $x \in H$ and define $y = \sum_{n=1}^{\infty} \langle x, x_n \rangle x_n$.NTS $x = y$
We can expand $\langle x - y, x_n \rangle =$

$$= \langle x, x_n \rangle - \left\langle \sum_{n=k}^{\infty} \langle x, x_k \rangle x_k, x_n \right\rangle \tag{2.36}$$

$$= \langle x, x_n \rangle - \sum_{n=1}^{\infty} \langle x, x_k \rangle \langle x_k, x_n \rangle \tag{2.37}$$

to obtain $\langle x, x_n \rangle - \langle x, x_n \rangle = 0$ since $\langle x_k, x_n \rangle = \delta_{k,n}$ and therefore $x - y = 0$ and

$$x = \sum_{n=1}^{\infty} \langle x, x_n \rangle x_n \tag{2.38}$$

■

Chapter 3

Linear Operators

Previously covered was the topic of linear functionals. Recall that a linear functional is nothing more than a linear map whose range is $\mathbb{C}$. Now we will explore the more general case of linear operator between two arbitrary Hilbert spaces. We will however, be mostly concerned with operators $A : H \to H'$ where $H = H'$.

3.1 Basics

Before we get to involved in our discussion, it will help too go over a few notational conventions and elementary facts about Hilbert space.

For a given linear operator A we will refer to the set of vectors which A acts upon as the **domain** of A which we will most often write as $D(A)$. Similarly the **range** of A will often be denoted $R(A)$. We will need the analytic notion of a dense subset:

Definition 23 *A set Q is a dense subset of a set W if the closure of Q is equal to W*

The completeness property of Hilbert space allows us to consider operators defined only on a dense subspace of a Hilbert space. If we are given an operator A defined on $D \subset H$ a dense subset of H, then for $v \notin D$ $(v \in H)$ we can define $Av := \lim_{n\to\infty} Av_n$ where the sequence $(v_n) \subset D$ converges to v. In other words, if we have $(v_n) \subset D(A)$ satisfying $v_n \to 0$ and $Av_n \to w$, we must have $w = 0$ for this process to work. If an operator A admits such an extension it is **closable.** This extension, usually denoted $\overline{A}$, can be shown to be unique. If an operator cannot be extended any further by the method described then it is **closed**. More formally

Definition 24 *An operator A on a Hilbert space H is closed if, for all sequences $(x_n) \subset D(A)$, $x_n \to x$ and $Ax_n \to y$ implies $x \in D(A)$ and $Ax = y$.*

This may also be stated by defining the **graph** $G(A)$ of a linear operator $A : H \to H'$ as $G(A) = \{(x, Ax) : x \in D(A)\}$. A is then closed if $G(A)$ is a closed subspace of $H \times H'$.

It will also help to know that a closed subspace of a Hilbert space is a Hilbert space itself.

Many times we will be interested in the *bound of various operators.*

Definition 25 *For a linear operator $A : H \to H'$ define a positive number $\|A\|$ by*

$$\|A\| := \sup\{\|Ax\|_{H'}, x \in H, \|x\|_H = 1\} \tag{3.1}$$

where $\|x\|_H = \sqrt{\langle x, x\rangle_H}$. Note that from this definition it follows

$$\|Ax\| \leq \|A\| \, \|x\| \tag{3.2}$$

for every $x \in H$.

*We say that A is **bounded** when $\|A\| < \infty$*

It turns out that a linear operator will be bounded if and only if it is continuous. In fact, we have the following theorem demonstrating the equivalence of three important conditions.

Theorem 26 *For a linear map between Hilbert spaces $A : H \to H'$ the following three conditions are equivalent:*

1) A is continuous at zero

2) A is continuous

3) A is bounded

Proof. *2 implies1 does not need proof*

***1 implies 2**: Assume A is continuous at zero. Then $\forall \varepsilon > 0$ we can find $\delta > 0$ such that $\|Ax - 0\|_{H'} < \varepsilon$ whenever $\|x - 0\|_H < \delta$, $x \in H$. Let $\varepsilon > 0$. Now for $x, x' \in H$ with $\|x - x'\|_H < \delta$, from the linearity of A we have*

$$\|Ax - Ax'\|_{H'} = \|A(x - x') - 0\|_{H'} < \varepsilon. \text{So } A \text{ is continuous.}$$

***1 implies 3, 2 implies 3**: Suppose $\delta > 0$ is such that whenever $\|x\|_H \leq \delta$, we have that $\|Ax\|_{H'} < 1$. Then for any nonzero $v \in H$,*

$$\|Av\| = \left\| \frac{\|v\|}{\delta} \cdot A(\delta \frac{\cdot v}{\|v\|}) \right\| = \frac{\|v\|}{\delta} \left\| A(\delta \frac{\cdot v}{\|v\|}) \right\| \leq \frac{\|v\|}{\delta}$$

so A is bounded.

***3 implies 1, 3 implies 2**: Suppose A is bounded. We actually may show that A is uniformly continuous which is obviously a stronger conclusion. If A is bounded, say by $M > 0$, then for $x, y \in H$ we have*

$$\|A(y - x) - Ay\| = \|Ax\| \leq M \, \|x\|$$

Now, letting $x \to 0$ shows that A is continuous at 0. But since M is independent of our choice of y we have shown that A is in fact uniformly continuous.

■

Example 27 *(Shift operator) Consider the map $A : \ell^2 \to \ell^2$ defined by*

$$A(a_1, a_2, ...) = (0, a_1, a_2, ...) \tag{3.3}$$

for every $(a_1, a_2, ...) \in \ell^2$.

Example 28 *(Multiplication operator) Let $z \in C([a,b])$ where $C([a,b])$ is the set of continuous functions on $[a,b]$. Define $A : L^2([a,b]) \to L^2([a,b])$ as $(Ax)(t) = z(t)x(t)$. A is clearly linear and since*

$$\|Ax\|^2 = \int_a^b |x(t)|^2 |z(t)|^2 \, dt \leq \max_{[a,b]} |z(t)|^2 \int_a^b |x(t)|^2 \, dt,$$

we have that $\|Ax\| \leq \max_{[a,b]} |z(t)| \, \|x(t)\|$ so therefore A is also bounded.

Every linear operator on a Hilbert space is completely determined by it's action on the basis. Suppose we have a Hilbert space H with o.n.b. $S = \{e_1, e_2, e_3, ...\}$ and are considering the action of a linear operator A on H. For every $x \in H$ we may write $x = \sum_{n=1}^{\infty} \alpha_n e_n$ where $\langle x, e_n \rangle = \alpha_n \in \mathbb{C}$. Then by the linearity of A we may write

$$\begin{aligned} Ax &= A(\sum_{n=1}^{\infty} \alpha_n e_n) = A(\alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3 + ...) \\ &= A(\alpha_1 e_1) + A(\alpha_2 e_2) + A(\alpha_3 e_3) + ... \\ &= \alpha_1 A(e_1) + \alpha_2 A(e_2) + \alpha_3 A(e_3) + ... \\ &= \sum_{n=1}^{\infty} \alpha_n A e_n \end{aligned}$$

So we can see that the value of Ax is completely described by the value of Ae_i, $i \in \mathbb{N}$. Now, for $A : H \to H'$, if $S' = \{e_1', e_2', e_3', ...\}$ is an o.n.b. for H', then, since $Ae_i \in H'$, we may write

$$Ae_i = \beta_{i_1} e_1' + \beta_{i_2} e_2' + \beta_{i_3} e_3' + ... = \sum_{n=1}^{\infty} \beta_n e_n'$$

where $\beta_{i_j} \in \mathbb{C} \ \forall j \in \mathbb{N}$. Note that $\beta_{i_j} = \left\langle Ae_i, e_j' \right\rangle$. This leads us to the matrix representation of a linear operator A. Given a linear operator $A : H \to H'$, where $S = \{e_1, e_2, e_3, ...\}$ and $S' = \{e_1', e_2', e_3', ...\}$ are o.n.b.'s for H and H' respectively, the matrix representation $A = [a_{i,j}]$, that is, the i, j^{th} entry of the matrix for A is $a_{i,j} = \left\langle Ae_j, e_i^{'} \right\rangle$. From this discussion of the matrix representation of an operator, we can see that the composition of operators is the same as matrix multiplication in the usual sense. In light of this, we will typically drop the $\circ$ when discussing composition or products of operators i.e. $(A \circ B)(x) = A(B(x))$. We will usually just write AB in place of the more bulky $(A \circ B)$.

Next we will prove a lemma which will lead us to the proof of the what is called the Reisz representation theorem. The reader will need to recall that the

null space, $null(L)$, of a linear map L is the subspace of all vectors which L maps to the zero vector.

Lemma 29 *Let f be a bounded linear functional on an inner product space E. Then the dimension of $null(f)^{\perp} \leq 1$.*
Proof. *If $f = 0$ we are done so assume f is non-zero. Since f is continuous $null(f)$ is a closed subspace of E so $null(f)^{\perp}$ is not empty. If we take non-zero $x, y \in null(f)^{\perp}$, since $f(x) \neq 0$ and $f(y) \neq 0$, we can find some scalar $a \neq 0$ such that $f(x)+af(y) = 0$. f is linear so $f(x)+af(y) = f(x+ay)$ which implies $x+ay \in null(f)$. However, we choose $x, y \in null(f)^{\perp}$ which is a vector space so we must have $x + ay \in null(f)^{\perp}$ as well. The only way this may happen is if x and y are linearly dependent since $a \neq 0$. Since x and y where chosen arbitrarily we see that they each span $null(f)^{\perp}$ i.e. it is one dimensional.* ■

Now we can prove

Theorem 30 *(Reisz Representation Theorem) Let f be a bounded linear functional on a Hilbert space H. Then there exist some $x_0 \in H$ such that $f(x) = \langle x, x_0 \rangle$ for all $x \in H$ and $\|f\| = \|x_0\|$.*
Proof. *If $f = 0$ then $x_0 = 0$ works so assume f is a non-zero functional. From lemma 29 we know that $null(f)^{\perp}$ has dimension one. Let $z_0 \in null(f)^{\perp}$ with $\|z_0\| = 1$. Then for every $x \in H$ we may write $x = x - \langle x, z_0 \rangle z_0 + \langle x, z_0 \rangle z_0$. Since $\langle x, z_0 \rangle z \in null(f)^{\perp}$ we must have $x - \langle x, z_0 \rangle z_0 \in null(f)$ so*

$$f(x - \langle x, z_0 \rangle z_0) = 0$$

which means

$$f(x) = \langle x, z_0 \rangle f(z_0) = \left\langle x, \overline{f(z_0)} z_0 \right\rangle$$

so let $x_0 = \overline{f(z_0)} z_0$ and we have $f(x) = \langle x, x_0 \rangle$ for all $x \in H$. We need to make sure x_0 is unique.

Suppose there is some x_1 satisfying $f(x) = \langle x, x_1 \rangle$ for all $x \in H$. Then $\langle x, x_1 \rangle = \langle x, x_0 \rangle$ or $\langle x, x_1 - x_0 \rangle = 0$ for all $x \in H$. Then $\langle x_1 - x_0, x_1 - x_0 \rangle = 0$ which can only happen if $x_0 = x_1$.

To show that $\|f\| = \|x_0\|$ consider $\|f\| = \sup_{\|x\|=1} |f(x)| = \sup_{\|x\|=1} |\langle x, x_0 \rangle| \leq \sup_{\|x\|=1} \|x\| \|x_0\| = \|x_0\|$

On the other hand, $\|x_0\|^2 = |\langle x, x_0 \rangle| = |f(x)| \leq \|f\| \|x_0\|$

Therefore $\|x_0\| = \|f\|$ ■

3.2 The adjoint

For now we are only going to consider linear operators between the same spaces i.e. operators $A : H \to H$. There are plenty of interesting things to study about operators between different spaces but knowledge of them will not be necessary what we're ultimately concerned with. Right now we want to define and discuss what are known as self-adjoint operators which will be of great importance to us later. First we start with the definition of a Hermitian operator

Definition 31 *An operator $A : H \to H$ is said to be Hermitian if*

$$\langle Af, g\rangle = \langle f, Ag\rangle$$

holds for all $f, g \in H$

Using this definition we will present a lemma which will prove to be one of the main reason Hermitian operators are so important in quantum physics.

Lemma 32 *An operator $A : H \to H$ is self-adjoint if and only if $\langle Af, f\rangle \in \mathbb{R}$ for all $f \in H$*

Most people are familiar with this result which comes from a fairly simple application of the properties of inner products so we won't prove it here.

Definition 33 *Let $A : H \to H$ be a linear operator on a Hilbert space H. Suppose there is $f, \psi \in H$ such that*

$$\langle \psi, A\varphi\rangle = \langle f, \varphi\rangle$$

for some $\varphi \in H$.
*The **adjoint** of A is an operator A^* defined as $A^*\psi = f$. An operator then is said to be **self-adjoint** if it is equal to its adjoint.*

The distinction between Hermitian operators and ones which are self-adjoint is very subtle. The main difference is that an operator that is self-adjoint must have the same domain as its adjoint while an operator is Hermitian simply if it satisfies $\langle Af, g\rangle = \langle f, Ag\rangle$. In the case of bounded operators they are the same. The definitions are famously confused as exemplified by the following anecdote of a meeting between the famous mathematician John von Neumann and physicist Werner Heisenberg. Their encounter is said to have gone something like this:

- During the 1960's John von Neumann was meeting with Werner Heisenberg and felt compelled to thank him for the invention of quantum mechanics, because this had led to the development of so much beautiful mathematics (such as the theory of operators on Hilbert space which we are discussing), adding that mathematics paid back a part of the debt by clarifying, for example, the difference of a self-adjoint operator and one that is only symmetric. Heisenberg replied: "What is the difference?"

In physics literature especially the distinction is usually relaxed and the operators are just referred to as Hermitian.

Sometimes it may be helpful to look at the set $\mathcal{L}(H, H')$, of all linear operators between two Hilbert spaces H and H' and also $\mathcal{B}(H, H')$, the set of all bounded linear operators between H and H' ($\mathcal{L}(H)$ or $\mathcal{B}(H)$ in the event $H = H'$). The adjoint of an operator $A \in \mathcal{L}(H, H')$ will then belong to the set $\mathcal{L}(H, H')$. In fact, if $A \in \mathcal{B}(H, H')$ it turns out $A^* \in \mathcal{B}(H', H)$, that is, if a linear operator is bounded then its adjoint is as well. In fact, we have the following theorem relating the norms of operators and their adjoints.

Theorem 34 *Let $A \in \mathcal{B}(H, H')$, H and H' both Hilbert spaces. Then*

$$\|A\|^2 = \|A^*\|^2 = \|A^*A\|$$

Proof. *Let $h \in H$ with $\|h\|_H \leq 1$.*

Then:

$$\begin{aligned} \|Ah\|_{H'}^2 &= \langle Ah, Ah\rangle_{H'} && (3.4)\\ &= \langle A^*Ah, h\rangle_H \leq \|A^*Ah\|_H \|h\|_H && (3.5)\\ &\leq \|A^*A\| \|h\|_H^2 \leq \|A^*A\| && (3.6)\\ &\leq \|A^*\| \|A\| && (3.7) \end{aligned}$$

So, by definition of $\|A\|$ we have $\|A\|^2 \leq \|A^*\|^2 \leq \|A^*A\|$.

To obtain the reverse inequality, simply substitute A^* for A and A^{**} for A^*. Therefore,

$$\|A\|^2 = \|A^*\|^2 = \|A^*A\| \quad (3.8)$$

■

A very important self-adjoint operator is the projection operator $P_K : H \to K$ which projects every vector onto a subspace $K \subset H$.

Definition 35 *A **projection** on a Hilbert space H is a bounded operator P satisfying $P^2 = P^* = P$*

Consider a closed subspace $K \subset H$. K itself is a Hilbert space and therefore it admits a basis representation, say $B = (e_i)$. Then the projection of H onto K, P_K, is the map

$$f \mapsto \sum_i \langle f, e_i\rangle e_i$$

for all $f \in H$. Recalling our previous discussion of orthogonality, it is clear that $Pf = f$ for all $f \in K$ and $Pf = 0$ for all $f \in K^\perp$.

The well know orthogonal projection theorem will be necessary to complete our proof of the spectral theorem for self-adjoint operators. We will state the theorem here. The proof is fairly straight forward but it is not particularly enlightening so we will state the theorem without proof.

Theorem 36 *If K is a closed subspace of a Hilbert space H then every $x \in H$ has a unique representation as a sum of vectors, $u \in K$ and $v \in K^\perp$. That is, for every $x \in H$ we may write*

$$x = u + v,\ u, v \in H \text{ where } \langle u, v\rangle = 0.$$

Next we state and prove an identity which, again, will aid is in our proof of the spectral theorem.

Theorem 37 *Let A be a self-adjoint operator on a Hilbert space H. Then we have $\|A\| = \sup_{\|x\|=1} |\langle Ax, x\rangle|$.*

Proof. *First set $M = \sup_{\|x\|=1} |\langle Ax, x\rangle|$. Then for $\|x\| = 1$, $|\langle Ax, x\rangle| \leq \|Ax\| \leq \|A\|$ and therefore $M \leq \|A\|$.*

Now let $\alpha := \sqrt{\frac{\|Ax\|}{\|x\|}}$ and $z := \frac{Ax}{\alpha}$. Then

$$\begin{aligned}
\|Ax\|^2 &= \langle A\alpha x, z\rangle \\
&= \frac{1}{4}\left[\langle A\alpha x + z, \alpha x + z\rangle - \langle A(\alpha x - z), \alpha x - z\rangle\right] \\
&\leq \frac{1}{4}M(\|\alpha x + z\|^2 + \|\alpha x - z\|^2) \\
&= \frac{1}{2}M(\|\alpha x\|^2 + \|z\|^2) \\
&= \frac{1}{2}M(\alpha^2\|x\|^2 + \frac{1}{\alpha^2}\|Ax\|^2) \\
&= M\|x\|\|Ax\|.
\end{aligned}$$

This gives $\|Ax\| \leq M\|x\|$ which means $\|A\| \leq M$. ■

Compactness is another quality of linear operator which is relevant to our discussion. We know that a compact set is one in which every sequence has a convergent subsequence. Similarly, a compact operator is simply an operator which preserves compactness. That is, if K is a compact subset of a Hilbert space H, then $A : H \to H$ is compact if the set $\{Ax \mid x \in K\}$ is compact as well. We may also define a **compact operator** as follows:

Definition 38 *A **compact operator** on a Hilbert space H is a bounded operator that maps the closed unit ball $B_1 \subset H$ into a compact set.*

Compact operators are special because they behave like finite rank-operators. In fact, it can be shown that every compact operator on a Hilbert space is the limit of a finite rank operator. Therefore we can regard the set of compact operators as the closure of the set of finite rank operators with respect to the operator norm.

Our first statement about compact operators will be an interesting equivalence statement involving the notion of weak convergence which was defined earlier. For proof see [LB&PM].

Theorem 39 *An operator A on a Hilbert space H is compact if and only if $x_n \to x$ weakly implies $A\ x_n \to Ax$.*

The eigenvalues of linear operator are of particular interest in quantum mechanics. We will require a few theorems regarding the eigenvalues of compact operators which we will develop here.

Lemma 40 *Let $A \neq 0$ be a self-adjoint, compact operator between a Hilbert space H and itself. A has an eigenvalue λ satisfying $|\lambda| = \|A\|$ or $|\lambda| = -\|A\|$*

Proof. Consider a sequence of unit vectors (u_n) in H such that $\|Au_n\| \to \|A\|$ as $n \to \infty$. Consider the quantity $\left\|A^2u_n - \|Au_n\|^2 u_n\right\|^2$ which may be expanded as follows

$$\begin{aligned}
\left\|A^2u_n - \|Au_n\|^2 u_n\right\|^2 &= \left\langle A^2u_n - \|Au_n\|^2 u_n, A^2u_n - \|Au_n\|^2 u_n \right\rangle \\
&= \left\langle A^2u_n, A^2u_n\right\rangle + \left\langle \|Au_n\|^2 u_n, \|Au_n\|^2 u_n\right\rangle - 2\|Au_n\|^2 \left\langle A^2u_n, u_n\right\rangle \\
&= \left\|A^2u_n\right\|^2 + \left\|\|Au_n\|^2 u_n\right\|^2 - 2\|Au_n\|^2 \langle Au_n, Au_n\rangle \\
&= \left\|A^2u_n\right\|^2 - \|Au_n\|^4 \qquad (3.9) \\
&\leq \|A\|^2 \|Au_n\|^2 - \|Au_n\|^4 \\
&= \|Au_n\|^2 (\|Au\|^2 - \|Au_n\|^2). \qquad (3.10)
\end{aligned}$$

Since we chose (u_n) such that $\|Au_n\| \to \|A\|$ as $n \to \infty$, the last expression, equation 3.10, will go to zero as well so $\left\|A^2u_n - \|Au_n\|^2 u_n\right\|^2 \to 0$ as $n \to \infty$.

Since A is compact A^2 will be as well so there exist a subsequence (u_{n_i}) of (u_n) such that $(A^2u_{n_i})$ converges. We may write the limit of $(A^2u_{n_i})$ as $\|A\|^2 v$ for some non-zero $v \in H$. Now, for any $i \in \mathbb{N}$ we have

$$\begin{aligned}
&\left\|\|A\|^2 v - \|A\|^2 u_{n_i}\right\| \qquad (3.11) \\
\leq\ &\left\|\|A\|^2 v - A^2u_{n_i}\right\| + \left\|A^2u_{n_i} - \|Au_{n_i}\|^2 u_{n_i}\right\| + \left\|\|Au_{n_i}\|^2 u_{n_i} - \|A\|^2 u_{n_i}\right\| \qquad (3.12)
\end{aligned}$$

The first quantity in the second half of equation 3.12 goes to zero by definition of $\|A\|^2 v$, the second since $\left\|A^2u_n - \|Au_n\|^2 u_n\right\|^2 \to 0$, and the third quantity of equation 3.11 eventually vanishes because we assume $\|Au_n\| \to \|A\|$. So $\left\|\|A\|^2 v - \|A\|^2 u_{n_i}\right\| = \|A\|^2 \|v - u_{n_i}\|$ as $n \to \infty$, which means that (u_{n_i}) converges to v, since $A \neq 0$, and therefore,

$$\|A\|^2 v = A^2 v$$

or

$$(A - \|A\|)(A + \|A\|)v = 0.$$

If $(A + \|A\|)v \neq 0$ then we may let $w = (A + \|A\|)v$ and then we have $(A - \|A\|)w = 0$. Therefore, w is an eigenvector of A with eigenvalue $\|A\|$. If it happens that $(A + \|A\|)v = 0$ then $-\|A\|$ is an eigenvalue of A. ■

The existence of the eigenvalue in the previous theorem will end up playing an important part in the proof of the spectral theorem.

Corollary 41 *If A is a compact self-adjoint operator on a Hilbert space H, then there is a unit vector $w \in H$ such that*

$$|\langle Aw, w\rangle| = \sup_{\|x\| \leq 1} |\langle Ax, x\rangle| \quad (\|w\| = 1)$$

Proof. *Let $w \in H$ be an eigenvector in lemma 40 corresponding to the eigenvalue λ with $|\lambda| = \|A\|$. Then from theorem 37 we have*

$$|\langle Aw, w\rangle| = |\langle \lambda w, w\rangle| = |\lambda| \|w\| = |\lambda| = \|A\| = \sup_{\|x\|\leq 1} |\langle Ax, x\rangle|$$

■

Finally, we are ready to present the Hilbert-Schmidt and spectral theorem. Most readers familiar with linear algebra know that, for a self-adjoint linear operator of finite rank, we may find an orthonormal basis of eigenvectors. We now seek to generalize this result for infinite dimensional Hilbert space. This idea will be very important in the applications to come.

Theorem 42 *(Hilbert Schmidt) For any compact self-adjoint operator A on an infinite dimensional Hilbert space H we may find an orthonormal system of eigenvectors (u_n) corresponding to nonzero (real) eigenvectors (λ_n) such that every $x \in H$ may be written uniquely as*

$$x = \sum_{n=1}^{\infty} \alpha_n u_n + v$$

where $\alpha_n \in \mathbb{C}$ and $Av = 0$. If A has an infinite number of distinct eigenvalues then we have $\lambda_n \to 0$ as $n \to \infty$.

Proof. *Let A be a compact self-adjoint operator on an infinite dimensional Hilbert space H. First, recall from lemma 40 and corollary 41 that we know A has an eigenvalue, λ_1, such that*

$$|\lambda_1| = \|A\| = \sup_{\|x\|\leq 1} |\langle Ax, x\rangle| .$$

Let u_1 be the normalized eigenvector corresponding to the eigenvalue λ_1 and set

$$Q_1 = \{x \in H : x \perp u_1\}$$

Note that for $x \in H$ we have $\langle Ax, u_1\rangle = \langle x, Au_1\rangle = \lambda_1 \langle x, u_1\rangle = 0$ so $Ax \perp u_1$ and therefore $Ax \in Q_1$ i.e. A maps Q_1 into itself. This means we may consider the map of A restricted to Q_1 and again apply corollary 41 and lemma 40 to find an eigenvalue λ_2 such that $|\lambda_2| = \sup_{\|x\|\leq 1} \{|\langle Ax, x\rangle| : x \in Q_1\}$ corresponding to a normalized eigenvector $u_2 \in Q_1$. Obviously $u_2 \perp u_1$ and we may set

$$Q_2 = \{x \in Q_1 : x \perp u_2\}$$

and continue in the same manner as above. If we have chosen eigenvectors $\lambda_1, ...\lambda_n$ with corresponding normalized eigenvectors $u_1, ..., u_n$ define

$$Q_n = \{x \in Q_{n-1} : x \perp u_n\}$$

and choose

$$|\lambda_{n+1}| = \sup_{\|x\|\leq 1} \{|\langle Ax, x\rangle| : x \in Q_n\}$$

Having chose λ_{n+1} we may take u_{n+1} to be the normalized eigenvector correspond to λ_{n+1}.

It may happen that this procedure terminates after a finite number of steps in which case every $x \in H$ has a unique representation as

$$x = \alpha_1 u_1 + ... + \alpha_k u_k + v$$

with $Av = 0$ and

$$Ax = \alpha_1 \lambda_1 u_1 + ... + \alpha_k \lambda_k u_k.$$

If the procedure terminates the theorem is proved but suppose that we obtain an infinite sequence of eigenvalues (λ_n) and eigenvectors (u_n). By the nature of how (u_n) is chosen it converges weakly to zero. Therefore, by theorem 39 we know $Au_n \to 0$ strongly and therefore

$$|\lambda_n| = \|\lambda_n u_n\| = \|Au_n\| \to 0$$

Now let $S = span(u_1, u_2, ...)$ we the space spanned by the vector or. From theorem 36 we know we know every $x \in H$ may be written uniquely as $x = u + v$, $u \in S$ and $v \in S^\perp$,

$$x = \sum_{n=1}^{\infty} \alpha_n u_n + v$$

Now to show $Av = 0$ let $w = \frac{v}{\|v\|}$. Then $\langle Av, v\rangle = \|v\|^2 \langle Aw, w\rangle$. Note that $w \in S^\perp \subset Q_n$ for all $n \in \mathbb{N}$ and

$$|\langle Av, v\rangle| = \|v\|^2 \langle Aw, w\rangle \leq \|v\|^2 \sup_{\|x\|\leq 1} \{|\langle Ax, x\rangle| : x \in Q_n\} = \|v\|^2 |\lambda_{n+1}| \to 0$$

This implies $\langle Av, v\rangle = 0$. So from theorem 37 , the norm of A restricted to $S^\perp$ is zero giving $Av = 0$ for all $v \in S^\perp$ ■

The work to prove the spectral theorem is now mostly done now. The spectral theorem is presented in various forms depending on the context of discussion but in general it is a way of completely characterizing compact self-adjoint operators.

Theorem 43 *(Spectral Theorem) Let A be a compact self-adjoint operator on a Hilbert space H with a complete orthonormal system of eigenvectors $\{v_1, v_2, ...\}$ with corresponding eigenvalues $\{\lambda_1, \lambda_2, ...\}$. Let $P_i : H \to H$ be the one dimensional projection onto $span(v_1, v_2, ...)$ defined by $x \mapsto \langle x, v_i\rangle$. Then for all $x \in H$ we may write*

$$x = \sum_{i=1}^{\infty} P_i x$$

and

$$A = \sum_{i=1}^{\infty} \lambda_i P_i$$

Proof. *Recall from theorem 42 we have the orthonormal system* $B' = \{u_1, u_2, ...\}$ *which forms a basis for vectors in* S*. Therefore if we add an arbitrary orthonormal basis for* $S^{\perp}$ *to the set* B' *we obtain a complete orthonormal system* $B = \{v_1, v_2, ...\}$*. Given a complete orthonormal basis* B *we may write any* $x \in H$ *as*

$$x = \sum_{i=1}^{\infty} \langle x, v_i \rangle \, v_i.$$

With our definition of P_i *as* $P_i x = \langle x, v_i \rangle$ *for all* $i \in \mathbb{N}$ *we have*

$$x = \sum_{i=1}^{\infty} P_i x,$$

and since our basis consist of eigenvectors we also have

$$Ax = \sum_{i=1}^{\infty} \lambda_i \langle x, v_i \rangle \, v_i = \sum_{i=1}^{\infty} \lambda_i P_i x,$$

which gives

$$A = \sum_{i=1}^{\infty} \lambda_i P_i.$$

■

3.3 The spectrum

The previous discussion had much to do with the eigenvalues of linear operators. In the infinite dimensional setting we have sort of a broader set of values we are interested in known as the spectrum.

Definition 44 *Let* A *be a linear operator on a Hilbert space* H*. We define the operator* R_λ *by*

$$R_\lambda := (\lambda I - A)^{-1}.$$

R_λ *is called the* ***resolvent operator*** *of* A*. The values of* λ *for which* R_λ *is defined on all of* H *and is bounded are called* ***regular points.*** *The set of all regular points, denoted* $\rho(A)$*, is called the* ***resolvent set*** *of* A*. The set of all* $\lambda \in \mathbb{C}$ *not in* $\rho(A)$ *is called the* ***spectrum*** *of* A *and is denoted by* $\sigma(A)$*.*

In summary, $\lambda \in \sigma(A)$ if one of the following conditions is true

- $(A - \lambda I)$ is not one-to-one
- The range of $(A - \lambda I)$ is not dense

We can easily see that every eigenvalue of A is in $\sigma(A)$: if v is an eigenvector of A corresponding to eigenvalue λ then v is a solution to the equation $A - \lambda I = 0$. So the operator $(A - \lambda I)$ is not one-to-one and therefore not invertible which means $\lambda \in \sigma(A)$. The spectrum may also contain values which are not eigenvalues of A. In fact, a non-empty spectrum may contain no eigenvalues.

Example 45 *Consider the multiplication operator $A : C([a,b]) \to C([a,b])$ on the space $C([a,b])$ of continuous functions on the interval $[a,b]$ defined by*

$$(Ax)(t) = u(t)x(t)$$

for fixed $u \in C([a,b])$. In this case the resolvent operator $R_\lambda = (A - \lambda I)^{-1}$ is defined by the rule $x(t) \mapsto \frac{x(t)}{u(t)-\lambda}$ so we can see that

$$\sigma(A) = \{\lambda : u(t) = \lambda \text{ for some } t \in [a,b]\},$$

or in other words the spectrum of A consists of the values in the range of u. In the case that u is non-constant and strictly increasing we have $\sigma(A) = [u(a), u(b)]$.

We know that the eigenvalues of a self-adjoint operator are always real and that the spectrum contains all eigenvalues. Now we will prove a stronger statement about the spectrum.

Theorem 46 *If A is a self-adjoint operator then $\rho(A)$ contains all non-real numbers*

Proof. *If $z \in \mathbb{C}$ but $z \notin \mathbb{R}$ then $g(\lambda) = (z - \lambda)^{-1}$ is a bounded continuous function on the real line. Therefore the resolvent operator $R_\lambda = (A - \lambda)^{-1}$ is a bounded operator defined everywhere so $z \in \rho(A)$.* ■

So, the spectrum of a self-adjoint linear operator may only contain real values.

Now that we know that the spectrum of a self-adjoint operator contains only real values we are of course interested in when exactly we can say that a real number $\lambda \in \sigma(A)$. The following theorem gives us an equivalent condition.

Theorem 47 *If $A : H \to H$ is a closed self-adjoint then the real number $\lambda \in \sigma(A)$ if and only if there is a sequence $(u_n) \subset D(A)$ such that*

$$\|u_n\| = 1 \text{ and } \|(\lambda - A)u_n\| \to 0 \text{ as } n \to \infty. \tag{3.13}$$

Proof. *Suppose condition 3.13 holds. If $\lambda \in \rho(A)$ then we would have $\|u_n\| = \|R_\lambda(\lambda - A)u_n\| \to 0$ from the second part of our assumption, contradicting the first part, $\|u_n\| = 1$.*

Conversely, if we suppose that 3.13 does not hold then we may find some constant C such that

$$\|u\| \leq C \|(\lambda - A)u\| \text{ for } u \in D(A). \tag{3.14}$$

If not then we can find $(v_n) \subset D(A)$ *such that* $\|(\lambda - A)v_n\|^{-1} \|v_n\| \to \infty$ *and setting* $u_n = \frac{v_n}{\|v_n\|}$ *gives us a sequence satisfying condition 3.13. Therefore, condition 3.14 holds. This gives us two important bits of information. First, it tells us that the function* $(\lambda - A)$ *is one-to-one on* $D(A)$.

3.14 also implies that $R(\lambda - A)$ *is closed. To see this, consider a sequence* $(f_n) \subset R(\lambda - A)$ *such that* $f_n \to f$ *in* H. *Since* $(\lambda - A)$ *is one-to-one on* $D(A)$ *we may find a unique* $u_n \in D(A)$ *satisfying* $(\lambda - A)u_n = f_n$. *Now condition 3.14 implies that* (u_n) *is Cauchy so* $u_n \to u \in H$. *So, if we take* $v \in D(A)$ *we have*

$$\langle u, (\lambda - A)v\rangle = \lim \langle u_n, (\lambda - A)v\rangle = \lim \langle f_n, v\rangle = \langle f, v\rangle .$$

Therefore $u \in D(A)$ *with* $(\lambda - A)u = f$. *Knowing this allows us to show* $R(\lambda - A) = H$ *which gives* $\lambda \in \rho(A)$ *since we may then define* $R_\lambda f$ *as the solution to* $(\lambda - A)u = f$. *So to show* $R(\lambda - A) = H$ *let* $f \in H$ *and* $w \in R(\lambda - A)$. *Let* $v \in D(A)$ *be the unique solution (in* $D(A)$*) of* $(\lambda - A)v = w$. *We may now define a linear functional on* $R(\lambda - A)$ *by* $Fw := \langle v, f\rangle$. *Since we showed* $R(\lambda - A)$ *is closed and it is automatically a subspace of* H, $R(\lambda - A)$ *is itself a Hilbert space. From condition 3.14 we have*

$$|Fw| \leq \|v\| \|f\| \leq C \|f\| \|w\|$$

so F *is bounded. Therefore, by the Reisz representation theorem, there is* $u \in R(\lambda - A)$ *such that* $Fw = \langle w, u\rangle$ *for all* $w \in R(\lambda - A)$. *This gives*

$$\langle u, (\lambda - A)v\rangle = \langle f, v\rangle \ \text{ for } v \in D(A)$$

Since A *is self-adjoint, this implies* $u \in D(A)$ *and* $(\lambda - A)u = f$. *So* $f \in R(\lambda - A)$. ■

The preceding proof also gives proofs of the following corollaries

Corollary 48 *Self-adjoint operators are closed.*

Corollary 49 *If* A *is closed and*

$$\|v\| \leq C \|Ax\| , \ v \in D(A)$$

then A *is one-to-one and* $R(A)$ *is closed.*

Our next endeavor will be to show that the resolvent set, $\rho(A)$, is always an open set in $\mathbb{C}$. For this we will need a lemma and to employ the notion of convergence of a sequence of operators.

Recall the definitions of strong/weak convergence defined earlier. We have similar notions of convergence which we can define in the space $\mathcal{B}(H)$ of bounded linear operator on a Hilbert space

Definition 50 *Let* $(A_n) \subset \mathcal{B}(H)$ *be a sequence of linear operators. If there exist* $A \in \mathcal{B}(H)$ *such that* $\|A_n - A\| \to 0$ *as* $n \to \infty$ *we say* (A_n) *converges to* A *in the uniform operator topology or* ***converges in norm.*** *We say* A ***converges strongly*** *if* $A_n x \to Ax$ *for all* $x \in H$.

With this definition in mind, it may be shown that the space $\mathcal{B}(H)$ is in fact itself a Banach space.

Lemma 51 *Let $A \in \mathcal{B}(H)$. If $\|A\| < 1$ then $(I - A)$ is invertible in $\mathcal{B}(H)$ and the inverse is given by*

$$(I - A)^{-1} = I + A + A^2 + A^3 + ... = \sum_{k=0}^{\infty} A^k \tag{3.15}$$

Expression 3.15 is called the Neumann series.

Proof. *If $\|A\| < 1$, then $\sum_{k=0}^{\infty} \|A\|^k = \frac{1}{1-\|A\|} < \infty$ so the Neumann series $\sum_{k=0}^{\infty} A^k$ will converges to an operator in $\mathcal{B}(H)$ since it is complete.*

Note that if $S_i, S, T \in \mathcal{B}(H)$ and $S_i \to S$ in $\mathcal{B}(H)$ then $\|S_i T - ST\| \leq \|S_i - S\| \cdot \|T\| \to 0$ and $\|TS_i - TS\| \leq \|T\| \cdot \|S_i - S\| \to 0$. Therefore

$$(I - L)\left(\sum_{k=0}^{\infty} L^k\right) = \lim_{N\to\infty} (I - L)\sum_{k=0}^{N} L^k = \lim_{N\to\infty} (I - L^{N+1}) = I$$

since $\left\|L^{N+1}\right\| \leq \|L\|^{N+1} \to 0$. Similarly we have $\left(\sum_{k=0}^{\infty} L^k\right)(I - L) = I$. So $I - L$ is invertible with $(I - L)^{-1} = \left(\sum_{k=0}^{\infty} L^k\right)$ ■

Now we have

Corollary 52 *The resolvent set $\rho(A)$ of a linear operator A is an open set.*

Proof. *Let $\lambda \in \rho(A)$. So by definition the operator $(\lambda I - A)$ is bounded and invertible. Let $M = \left\|(\lambda I - A)^{-1}\right\|$ and let μ be such that $|\lambda - \mu| < \varepsilon = \frac{1}{M}$. We have*

$$\begin{aligned} \mu I - A &= (\mu - \lambda)I + \lambda I - A && (3.16)\\ &= (\lambda I - A)((\mu - \lambda)(\lambda I - A)^{-1} + I). && (3.17) \end{aligned}$$

Since $|\lambda - \mu| < \varepsilon$ we have

$$\left\|(\mu - \lambda)(\lambda I - A)^{-1}\right\| \leq \|(\mu - \lambda)\| \left\|(\lambda I - A)^{-1}\right\| < 1 \tag{3.18}$$

so $(\lambda I - A)((\mu - \lambda)(\lambda I - A)^{-1} + I) = \mu I - A$ is invertible, implying $\mu \in \rho(A)$.
■

Chapter 4

Physics

The often non-intuitive nature of quantum mechanics is, in part, a consequence of the statistical nature of the current quantum interpretation of atomic theory that has been adopted by modern physics. Before this, the classical description of mechanics was rooted in the notion of a phase space M and and the concept of time evolution. The points in the phase space each correspond to a state. At any point in time, each state is said to completely characterize the system. More specifically, a state determines any observable value of the system such as momentum, energy etc. and also allows one to make predictions about the future of the system so long as the state at $t = 0$ is known. For instance we have the familiar Hamiltonian which gives the energy of the system in a particular state. The relationship between states (the points of phase space M) and observables may be stated as follows

The value of the observable f in the state x is f(x).

This naturally leads to the formulation of certain kinds of yes-no questions: is the system in a state lying in $S \subset M$ i.e. is $x \in S$. The answer is defined by the characteristic function χ_S for S as usual with $\chi_S(x) = 1$ for 'yes' i.e. $x \in S$ and $\chi_S(x) = 0$ for 'no' i.e. $x \notin S$.

We may define a system classically using $M = \mathbb{R}^{2n}$ as our phase space to model a system consisting of point particles moving in $\mathbb{R}^n$. We can define coordinates $(q, p) := (q_i, p_i)$ where $i = 1, ..., n$. These coordinates define the position q of a particle with momentum q. To completely describe the system we will need a function h on $\mathbb{R}^{2n}$ which we will call the Hamiltonian of the system. As stated before, the Hamiltonian can be regarded as an observable that gives the value of the energy of the system, but it also plays a the dual role, determining the time-evolution of the system. Time-evolution is governed by Hamilton's equations:

$$\dot{q}_i := \frac{dq_i}{dt} = \frac{\partial h}{\partial p_i}$$

$$\dot{p}_i := \frac{dp_i}{dt} = \frac{\partial h}{\partial q_i}$$

The most well known example is the Hamiltonian for a particle with mass m moving in a potential V.

$$h(q,p) = \frac{p^2}{2m} + V(q)$$

where $p^2 := \sum_{i=1}^{n} (p_i)^2$

Applying the Hamiltonian equations to h gives us the relations

$$\dot{q}_i = \frac{p_i}{m}$$

$$\dot{p}_i = -\frac{\partial V}{\partial q_i}.$$

By defining the force $F_i := -\frac{\partial V}{\partial q_i}$, then the above equations are precisely Newton's equation $\overrightarrow{F} = m\overrightarrow{a}$.

4.1 Quantum mechanics

The fundamental difference between the classical and quantum interpretation is that in a quantum decision, an infinite number of possible values of an observable are possible. In quantum mechanics the phase space is taken to be a Hilbert space i.e. $M = H$. Because of the probabilistic interpretation of quantum mechanics we demand that only vectors with unit norm correspond to physical states and additionally if $y = \lambda x$, $x, y \in H$ and $|\lambda| = 1$ then x and y correspond to the same state.

As we progress, the postulates of quantum mechanics will be given in order and described in a similar nature as in *Operator Methods in Quantum Mechanics* by [MS]. Different texts and authors vary in minor ways on the precise form and order of the postulates to come, however, the overall picture they form is essentially equivalent.

4.1.1 Linear operators as observables

The motion of a particle in $\mathbb{R}^3$

One of the simplest quantum systems is that of a particle moving in $\mathbb{R}^3$. Here our wave function ψ will be in the Hilbert space $L^2(\mathbb{R}^3)$. Unless otherwise specified, all operations will be over all components of $x = (x_1, x_2, x_3)$ in $\mathbb{R}^3$, however, we will drop the component notation, e.g. dx not d^3x, except for where a distinction is necessary or desired. For a quantum system we cannot ask questions exactly as we did in the classical sense such as when is the particle at position $x = (x_1', x_2', x_3') \in \mathbb{R}^3$. The function ψ now corresponds to a probability density,

specifically $\|\psi\|^2$. This is the essential starting point of quantum mechanics and it is often stated as a more general postulate:

Postulate 1 : There is a function $\psi(x,t)$ of position x and time t such that the probability of the particle being found in the region Δ is given by the expression

$$\Pr(\psi, x \in \Delta) = \langle \psi(x,t), \chi_\Delta \psi(x,t) \rangle = \int_\Delta dx \, \|\psi(x,t)\|^2 .$$

This expression hints at the useful structure of Hilbert space as the last equality comes from the inner product defined on $L^2(\mathbb{R}^3)$ by

$$\langle f, g \rangle = \int_{\mathbb{R}^3} f(x) \cdot \overline{g(x)} dx.$$

The first stipulation above now appears obvious. We are demanding our function ψ have norm 1, i.e. $\int_{\mathbb{R}^3} dx \, \|\psi(x)\|^2 = 1$, so that $\int_{\mathbb{R}^3} d^3x \, \|\psi(x)\|^2 = \langle \psi, \chi_{\mathbb{R}^3} \psi \rangle = \langle \psi, \psi \rangle = \Pr(\psi, x \in \mathbb{R}^3) = 1$, which is interpreted as the likelihood $x \in \mathbb{R}^3$. So $\int_{\mathbb{R}^3} d^3x \, \|\psi(x)\|^2 = \Pr(\psi, x \in \mathbb{R}^3) = 1$ naturally since we are assuming that the particle is somewhere in $\mathbb{R}^3$. We also must demand that the limit of ψ at both negative and positive infinity be zero.

This example is quite specific. More generally we can represent any observable a with a self-adjoint linear operator A which we will give the following interpretation

The value of the observable a in the state ψ is $\langle \psi, A\psi \rangle$.

At this point one may call into question our ability to represent any observable with a self-adjoint operator. It turns out we will take this as one of our postulates of quantum mechanics. We will provide a more acceptable justification of this later, but for now we will discuss the some applications of this operator-observable association while taking for granted our ability to do so.

The drastic difference in the classical and quantum descriptions of a system can be somewhat intellectually uncomfortable at first. The statement "the probability of a particle in state ψ being found in region Δ is $\Pr(\psi, x \in \mathbb{R}^3)$" would seem rather useless since when we actually perform an experiment we will find that the particle either is indeed in Δ or it is not. Unfortunately this is the nature of quantum mechanics. As we will see, there is a limit on the degree of accuracy one can expect from the calculated value of an observable in a particular state. That is, there is a degree of uncertainty associated with any measurement of a quantum system. We cannot say "a particle is at position x_0 at time t". We instead may interpret $\Pr(\psi, x \in \Delta)$ in the following way. Suppose we perform N experiments trying to detect the particle in state ψ in some region Δ. Say the particle is detected i times in Δ. As $N \to \infty$, the value

of $\frac{i}{N}$ will approach $\Pr(\psi, x \in \Delta)$. In this sense, quantum mechanics is usually not particularly useful in calculating specific results for a single experiment, but rather what value we can expect the results to center around or converge to given a large number of trials. The expectation value represents the statistical mean of the value of an observable. It is the nature of quantum behavior that, in general, identical experiments repeated several times will not yield the same results. But, the statistical mean of the measured values from experiment will tend to the expectation value as more experiments are performed.

Projections

Projections can be thought of as the quantum analogues to characteristic functions in classical mechanics. For a given projection $P_K : H \to K$, we may ask, is the system in K. However, the fundamental difference between classical mechanics and the quantum mechanical interpretation of expectation value is that we have more possible answers to this question than just yes or no. It becomes a *yes-no* and varying degrees of *maybe* situation. If the answer to the question is yes, i.e. $\psi \in K$, then as with the characteristic function, we have $1 = \langle P_K\psi, \psi\rangle$ since $\|\psi\| = 1$. Similarly, if the answer is no, that is the system is not in K ($\psi \notin K$), then $\langle P_K\psi, \psi\rangle = \langle 0, \psi\rangle = 0$. There is also the possibility that ψ is of the form $\psi = \psi^{\perp} + \psi^{\parallel}$ where $\psi^{\perp} \in K^{\perp}$ and $\psi^{\parallel} \in K^{\parallel}$. In this case, we have

$$\langle P_K\psi, \psi\rangle = \left\langle \psi^{\parallel}, \psi \right\rangle = \left\langle \psi^{\parallel}, \psi^{\parallel} \right\rangle = \left\| \psi^{\parallel} \right\|^2 \in [0, 1].$$

Note: $0 \leq \left\| \psi^{\parallel} \right\|^2 \leq \|\psi\|^2 = 1$, (Recall $\|P_K\| = 1$ for projections). This assures that any predictions we make based on expectation values make sense; at least statistically. Born and Pauli gave the following final generalized interpretation to wavefunctions:

The number $\langle P_K\psi, \psi\rangle$ is the probabilty that the state ψ lies in K.

If we are considering an observable a and a range or set I of possible values of a then a natural question to ask would be how likely is it that $a \in I$. Projections allow us to answer this question which we will explore in more detail soon.

Mixed States

The above discussion of expectation values is in fact very limited. It is often the case that the precise state of a system cannot be determined. In this case we will have to settle for a list of candidate pure states (ψ_i) and assign each state a probability $\Pr_i$ where naturally we demand that $\Pr_i \in [0, 1]$ for all $i \in \mathbb{N}$ and $\sum_{i=1}^{\infty} \Pr_i = 1$. We will use ρ to denote the so-called **density matrix** corresponding to this weighted series of state vectors.

Now for an observable A, the value of A in the given state ρ is

$$\langle A\rangle = \sum_{i=1}^{\infty} \Pr_i \langle \psi_i, A\psi_i\rangle \tag{4.1}$$

Often ρ is referred to as a density matrix $\rho = \sum_{i=1}^{\infty} \Pr_i[\psi_i]$, where $[\psi_i]$ denotes the one-dimensional projection onto the subspace spanned by ψ_i. The term 'density matrix' suggests the following reformulation of 4.1 as

$$\langle A \rangle = Tr(\rho A)$$

where the trace of a bounded operator B acting on a Hilbert space with respect to o.n.b. $\{e_i\}$ is

$$Tr(B) := \sum_i \langle Be_i, e_i \rangle . \tag{4.2}$$

In quantum mechanics, density matrices are generally referred to as **mixed states**.

In some physics literature the convergence of the sum in equation 4.2 is taken for granted as well as its invariance with respect to the basis used. In reality however, the situation is not so simple. It is easily verified that for a finite-dimensional vector space $\sum_i \langle Be_i, e_i \rangle = \sum_i \langle B\mu_i, \mu_i \rangle$ for any choice of orthonormal bases $\{e_i\}$ and $\{\mu_i\}$. The situation changes for an infinite dimensional space. In fact, it is possible to find a bounded operator B and two distinct orthonormal bases $\{u_i\}$ and $\{v_i\}$ such that $Tr(B) := \sum_i \langle Bu_i, u_i \rangle = \infty$ and $Tr(B) := \sum_i \langle Bv_i, v_i \rangle = \infty$.

To solve this problem we need to construct a new class of operators for which 4.2 converges to a finite value independent of the basis chosen. While there is more than one approach, we will follow the method suggested in [NPL]. For a well defined trace we must demand A be bounded and compact. This implies the Hermitian product of operators A^*A is bounded and compact as well. The spectral theorem tells us that A^*A may be written

$$A^*A = \sum \lambda_i P_i,$$

where P_i is the projection onto the eigenspace corresponding to eigenvalue λ_i. Now we know if ψ belongs to the image of P_i then we have $A^*A\psi = \lambda_i \psi$ and we see $\|A\psi\|^2 = \langle A^*A\psi, \psi \rangle = \lambda_i \|\psi\|^2 \geq 0$ i.e. $\lambda_i \geq 0$. Now we may define the class of operators we are interested in constructing,

Definition 53 *For a bounded and compact linear operator $A : H \to H$, we say that A is **trace class** if the sum $\sum \lambda_i < \infty$ where λ_i are the eigenvalues of A^*A.*

This class is admittedly limited, but useful. The most important attributes of trace class operators may be summarized in the following theorem.

Theorem 54 *If $A : H \to H$ is trace class then the following hold*

- *The expression 4.2 is absolutely convergent and independent of basis. That is, $\sum_i \langle Be_i, e_i \rangle = \sum_i \langle B\mu_i, \mu_i \rangle$ for any choice of orthonormal bases $\{e_i\}$ and $\{\mu_i\}$.*

- *For any bounded linear operator $B : H \to H$ we have*

$$Tr(AB) = Tr(BA),$$

and the products AB and BA are both trace class.

- *If $U : H \to H$ is a unitary operator (meaning $\langle x, y\rangle = \langle Ux, Uy\rangle$ for all $x, y \in H$) then*

$$Tr(UAU^{-1}) = Tr(A),$$

Finally, the above discussion has motivates the following definition which generalizes the idea of mixed states and density matrices.

Definition 55 *A bounded operator $\rho : H \to H$ is called a **density operator** (or density matrix) if:*

- *ρ is positive*
- *ρ is trace class*
- *$Tr(\rho) = 1$*

We say that ρ is positive if $\langle \rho f, f\rangle \geq 0$ for all $f \in H$.

The approach outlined in this section, this observable-operator correspondence, is the essence of the mathematical model of quantum mechanics. However, we have limited our discussion to mostly bounded operators. As stated in [MR&BS], "It is a fact of life that many of the most important operators which occur in mathematical physics are not bounded". While this may seem rather discouraging, at this point most readers familiar with the study of unbounded operators should realize that all is not lost. Much of the tools and machinery developed earlier related to bounded or compact operators can be brought to bear in the study of unbounded operators. Because of this, we will not devote much time to discussion or proving any new results for operators specifically in the unbounded case. Rather, we will continue discussing observables, specifically those of position, momentum, and energy, and their corresponding operators, shedding more light on their application without concerning ourselves with certain seemingly important operator properties (such as boundedness...). In short, we are taking the physicist's approach. The justification for this is that most of the expressions and relations which define the interesting unbounded operators we will encounter do not necessarily make sense on all of H. These operators are usually defined naturally on some dense subspace, say $K \subset H$, making the specification of this domain crucial. As was discussed early in the section on linear operators, the definition of an operator on a dense subspace is sufficient. There is a great deal of literature devoted to rigorously developing the formalities of unbounded operators and the reader is referred to [TFJ] and [KS] for a better discussion.

4.1.2 Position, Momentum, and Energy

In classical mechanics the relation $p = m\frac{dx}{dt}$ where p is the momentum, defined by mass, m, multiplied by the time derivative of position x. Our second postulate of quantum mechanics is

Postulate 2 : The probability that the momentum p of a particle is contained in some interval I is given by

$$\frac{1}{\hbar}\int_I \left\|\widehat{\psi}(\frac{p}{\hbar},t)\right\|^2 dp$$

where $\widehat{\psi}$ denotes the Fourier transform of ψ

Here we will define the **Fourier transform** $\widehat{\psi}(k,t)$ of $\psi(x,t)$ with respect to x by

$$\mathcal{F}(\psi(x,t)) = \widehat{\psi}(k,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi(x,t)e^{-ikx}dx$$

and

$$\mathcal{F}^{-1}(\widehat{\psi}(k,t)) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\widehat{\psi}(k,t)e^{ikx}dk = \psi(x,t)$$

defines $\mathcal{F}^{-1}$ the **inverse Fourier transform**. In quantum mechanics the traditional **wave number** k for the wave function ψ defines the momentum of the particle by

$$p = \hbar k. \tag{4.3}$$

That is, the Fourier transform and its inverse are maps between functions on position and momentum space. The relation in 4.3 highlights the significance of the universal constant $\hbar$. The convention of distributing the factor of 2π evenly between the transform and its inverse is necessary to ensure the Fourier transform remains unitary.

The momentum operator

We wish to define the momentum operator $\mathcal{P}$ by

$$\mathcal{P}_{x_i}\psi = -i\hbar\frac{\partial\psi}{\partial x_i}, \tag{4.4}$$

but it will take a little work to get there.

Here we will borrow a few basic ideas from statistics. Our interpretation of the function ψ defines a probability density for the quantity x. Similarly, $\widehat{\psi}$ defines a probability density for the quantity p. Therefore, we may use the well known equation for the average value of a variable with a corresponding non-discrete probability density to define the average value of the momentum as

$$\overline{p} = \frac{1}{\hbar}\int_{-\infty}^{\infty} p\left\|\widehat{\psi}(\frac{p}{\hbar},t)\right\|^2 dp = \hbar\int_{-\infty}^{\infty} k\left\|\widehat{\psi}(k,t)\right\|^2 dk.$$

In order to justify our desired definition of L_{x_i} in 4.4 we will need to derive the following basic relations between Fourier transforms and their inverses,

Claim 56 *a. (Parseval's identity)*

$$\int_{-\infty}^{\infty} \psi(x,t)\overline{\varphi(x,t)}dx = \int_{-\infty}^{\infty} \widehat{\psi}(k,t)\overline{\widehat{\varphi}(k,t)}dk \tag{4.5}$$

b.

$$\frac{\partial\widehat{\psi}(k,t)}{\partial k} = -i\mathcal{F}(x\psi(x,t)) \tag{4.6}$$

c.

$$\mathcal{F}(\frac{\partial\psi(x,t)}{\partial x}) = ik\widehat{\psi}(k,t) \tag{4.7}$$

Proof. *a. Recall the inner product defined on $L^2(\mathbb{R}^3)$ in section 4.1.1. For any two elements $f, g \in L^2(\mathbb{R}^3)$, writing $f(x) = \frac{1}{\sqrt{2\pi}}\int \widehat{f}(k)e^{ikx}dk$, we can see that*

$$\begin{aligned}
\langle f, g\rangle &= \frac{1}{\sqrt{2\pi}}\int\left[\int \widehat{f}(k)e^{ikx}dk\right]\overline{g(x)}dx \\
&= \int \widehat{f}(k)\overline{\int g(x)e^{-ikx}dxdk} \\
&= \int \widehat{f}(k)\overline{\widehat{\varphi}(k)}dk.
\end{aligned}$$

b. Writing $\widehat{\psi}(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi(x)e^{-ikx}dx$ we may differentiate with respect to k so

$$\begin{aligned}
\frac{\partial\widehat{\psi}(k,t)}{\partial k} &= \frac{-i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x\psi(x)e^{-ikx}dx \\
&= -i\mathcal{F}(x\psi(x,t)).
\end{aligned}$$

c. Consider the integral $\int_{-r}^{r}\frac{\partial\psi(x)}{\partial x}\cdot e^{-ikx}dx$. Using integration by parts we find $\int_{-r}^{r}\frac{\partial\psi(x)}{\partial x}\cdot e^{-ikx}dx =$

$$\begin{aligned}
&= ik\int_{-r}^{r}\psi(x)\cdot e^{-ikx}dx + e^{-ikx}\psi(x)\mid_{x=-r}^{x=r} \\
&= ik\int_{-r}^{r}\psi(x)\cdot e^{-ikx}dx + e^{-ikr}\psi(r) - e^{ikr}\psi(-r)
\end{aligned}$$

Since our wave function goes to zero at negative infinity the last line goes to $\sqrt{2\pi}ik\widehat{\psi}(k)$ as $r\to\infty$, proving (c). ■

If we take another look at the average momentum we can use the above properties to see

$$\begin{aligned}
\overline{p} &= \hbar \int_{-\infty}^{\infty} k\widehat{\psi}(k,t)\overline{\widehat{\psi}(k,t)}dk \\
&= -i\hbar \int_{-\infty}^{\infty} \mathcal{F}(\frac{\partial \psi(x,t)}{\partial x})\overline{\widehat{\psi}(k,t)}dk \\
&= -i\hbar \int_{-\infty}^{\infty} \frac{\partial \psi(x,t)}{\partial x}\overline{\psi(x,t)}dx \\
&= \left\langle -i\hbar \frac{\partial \psi(x,t)}{\partial x}, \psi(x,t) \right\rangle \\
&= \langle \mathcal{P}_x \psi(x,t), \psi(x,t) \rangle .
\end{aligned}$$

Now our definition of $\mathcal{P}_{x_i} = -i\hbar\frac{\partial}{\partial x_i}$ seems completely natural since we can see that the expectation value of $\mathcal{P}_{x_i}$ in state ψ is simply equal to the average value of the momentum.

In classical physics the kinetic energy T of an object moving with momentum p and mass m is given by

$$T = \frac{p^2}{2m}.$$

Since we have an expression for the quantum mechanical momentum operator $\mathcal{P}_{x_i} = -i\hbar\frac{\partial}{\partial x_i}$ we may apply the identity 4.6 to get the general expression $\mathcal{F}(\frac{\partial^n \psi(x,t)}{\partial x^n}) = (ik)^n \widehat{\psi}(k,t)$ and write

$$\overline{p^2} = \left\langle \mathcal{P}^2 \psi, \psi \right\rangle$$

so we will take

$$\langle T \rangle = \overline{T} = \frac{\left\langle \mathcal{P}^2 \psi, \psi \right\rangle}{2m}. \tag{4.8}$$

as the expectation value of kinetic energy.

If we are given a real valued function V, and if additionally V is continuous with $\int_{-\infty}^{\infty} V(x) \left|\psi(x,t)\right|^2 dx < \infty$, then we may also consider the expectation value of the potential energy

$$\langle V \rangle = \overline{V} = \int_{-\infty}^{\infty} V(x) \left|\psi(x,t)\right|^2 dx. \tag{4.9}$$

Since we have an expression for the expectation value of kinetic and potential energy, equations 4.8 and 4.9 respectively, we can use the fact the mathematical expectation of a sum is equal to the sum of mathematical expectation to express the expectation value for the total energy of the system as

$$\langle E \rangle = \langle H\psi, \psi \rangle ,$$

where H is known as the **Hamiltonian** of the system or the **energy operator**

$$H = \frac{\mathcal{P}^2}{2m} + V.$$

Observables and the spectrum

Now that we have our basic operators of position, momentum, and energy, we can have a bit more involved discussion about observables and develop techniques to apply some of the mathematics discussed earlier. Here we will state our next postulate

Postulate **3** : For every real valued observable a there corresponds a Hermitian operator A with a dense domain $D(A)$ such that

$$\overline{a} = \langle A\psi, \psi \rangle \text{ for all } \psi \in D(A)$$

Since our expectation value for an observable comes from the quantity $\langle A\psi, \psi \rangle$, where A is Hermitian, our expectation value $\langle A\psi, \psi \rangle$ is real. So given an observable a one question we want to ask is how likely is it that a belongs to some some interval $I \subset \mathbb{R}$. This suggests we are looking to employ the characteristic function χ_I for I. The value of $\chi_I(a)$ is then an observable itself taking on the value 1 when $a \in I$ and 0 if not. But we want the expectation value of $\chi_I(a)$ so we will define

$$\begin{aligned} \overline{\chi_I(a)} &= P(a \in I) \cdot 1 + P(a \notin I) \cdot 0 \\ &= P(a \in I) \end{aligned}$$

as our expectation value of $\chi_I(a)$. Now according to our postulate there is some Hermitian operator, we'll call it E_I, corresponding to $\chi_I(a)$ such that

$$\overline{\chi_I(a)} = P(a \in I) = \langle E_I\psi, \psi \rangle .$$

This is a good start to answering our question, but we still don't know exactly how E_I is constructed or related to A. We spoke earlier about how projection operators in Hilbert space can be thought of as the quantum analogue of the characteristic function. Now we will shed some more light on this idea.

It turns out that, given an observable a represented by a Hermitian operator A, we want to construct E_I as follows. Suppose $I \subset \mathbb{R}$ is given and λ_1, λ_2,... are the eigenvalues of A contained in I. Let $\{u_1, u_2, ...\}$ be the eigenvectors associated with the eigenvalues contained in I. We define E_I to be the projection operator which projects vectors in the domain of A onto $span\{u_1, u_2, ...\}$. From this definition follows a few of the statistical properties we would expect for the value of $\overline{\chi_I(a)}$. In particular

1. $E_{I \cup J} = E_I + E_J - E_{I \cap J}$
2. $E_{I \cap J} = E_I E_J$
3. $\overline{\chi_I(a)} \to 1$ as $I \to \mathbb{R}$.

From the definition of E_I we can now see why the the definition of the spectrum is so important.

Corollary 57 *An observable may only assume values which belong to the spectrum of its corresponding operator.*

Proof. *Let $\lambda_0 \in \rho(A)$ where A is a self-adjoint linear operator representing an observable a. We know $\rho(A)$ is open so there is an open interval $I \subset \rho(A)$ containing λ_0. All the eigenvalues of A are contained in the spectrum of A so, from the definition E_I we have*

$$P(a \in I) = \langle E_I \psi, \psi \rangle = 0.$$

■

This result suggests that the next question we want to ask is what values are in the spectrum of the operators we are interested in. First we will look at the position operator. Since the state function $\psi(x)$ defines a probability density function of the position it is clear the operator corresponding to the position will be multiplication by x i.e.

$$\begin{aligned} \overline{x} &= \langle x\psi, \psi \rangle \\ &= \int x \left\| \psi(x) \right\|^2 dx. \end{aligned}$$

Since an operator is not technically defined unless its domain is specified we will take the time to do that here. First we need to mention a few facts about operators. This will be brief and mostly without proof.

For an operator A corresponding to a real valued observable it may be shown that for all $f \in L^2(\mathbb{R}^3)$ we may find a vector $u \in D(A)$ such that $f = u + A^2 u$, i.e. the operator $1 + A^2$ is onto.

Corollary 58 *For every $f \in L^2(\mathbb{R}^3)$ there is $u, w \in D(A)$, where A corresponds to a real valued observable, such that*

$$(A - i)u = (A + i)u = f.$$

Proof. *We know for any $f \in L^2(\mathbb{R}^3)$ there is $u \in D(A)$ such that $(A^2 + I)u = (A - i)(A + i)u = (A + i)(A - i)u = f$. Therefore both $(A + i)u$ and $(A - i)u$ are in $D(A)$.* ■

The following three properties are enough to show that a densely defined Hermitian operator $A : H \to H$ is self-adjoint:

- $R(1 + A^2) = H$
- $R(A - i) = H$
- $R(A + i) = H$

Now we can specify the domain of the position operator which we will from here on denote as X. Since we want X to map between $L^2(\mathbb{R}^3)$ the simplest domain we can choose for X is the set of all $\psi \in L^2(\mathbb{R}^3)$ such that $x\psi \in L^2(\mathbb{R}^3)$

as well. We can show X is in fact self-adjoint on this domain as follows. If $f \in L^2(\mathbb{R}^3)$ then $\psi = \frac{f}{x+i} \in D(X)$ and $(X+i)\psi = f$. Therefore $R(X+i) = L^2(\mathbb{R}^3)$. By a similar argument $R(X-i) = L^2(\mathbb{R}^3)$. Moreover, we may show that $D(X)$ is a dense subset of $L^2(\mathbb{R}^3)$ since for $f \in L^2(\mathbb{R}^3)$ if we define $\psi_\varepsilon := \frac{f}{\varepsilon x^2+1}$ we get $\psi_\varepsilon \in D(X)$ for every $\varepsilon > 0$. Therefore $\|\psi_\varepsilon - f\| \leq \frac{\varepsilon x^2 \|f\|}{(1+\varepsilon x^2)}$ so $\psi_\varepsilon \to f$ as $\varepsilon \to 0$. This along with the condition stated in 4.1.2 is enough to conclude that X is self-adjoint.

Now we turn our attention back to determining the spectrum of X. We know the eigenvalues of X will be in the spectrum so we will look for those points first.

The eigenvalues of X will be the $\lambda \in \mathbb{R}$ for which there are solutions to the equation

$$(\lambda - x)\psi = 0 \tag{4.10}$$

A solution of 4.10 must satisfy $\psi(x) = 0$ for all $x \neq \lambda$. So $\psi = 0$ almost everywhere so X evidently has no eigenvalues. Now we must employ the method proposed in 47. This is easily done by constructing a sequence of functions with unit norm which converge to zero when $(\lambda - x)$ is applied and also for $x \neq \lambda$. Let

$$\psi_n = c_n e^{-\frac{n^2(x-\lambda)^2}{2}}$$

with

$$c_n = \sqrt{n}\pi^{-\frac{1}{4}}$$

This gives

$$\|x - \lambda\psi_n\|^2 = c_n n^{-3} \int_{-\infty}^{\infty} y^2 e^{-y^2} dy \to 0$$

So $\lambda \in \sigma(X)$. Since λ is an arbitrary real number this shows that $\sigma(X) = \mathbb{R}$.

We can follow a process similar to the one just described to determine the domain and spectrum of the momentum operator L. From Parseval's identity and the other relations in claim 56 we can see that we can define the action of L by

$$L\psi = \hbar\mathcal{F}^{-1}(k\widehat{\psi}),$$

and this will be equivalent to our previous definition. This essentially converts L into an operator similar to X and we can follow an almost identical procedure as with X to show that L is self-adjoint.

The spectrum can also be determined in a similar fashion. For any $\lambda \in \mathbb{R}$ we can define φ_n as

$$\varphi_n = c_n e^{-\frac{n^2(k-\lambda)^2}{2}},$$

where c_n is defined as above. Again we have $\|\varphi_n\| = 1$ and $\|(k-\lambda)\varphi_n\| \to 0$. If we let ψ_n be the inverse transform of φ_n this gives

$$(L - \lambda)\psi_n = \hbar\mathcal{F}^{-1}[(k-\lambda)\varphi_n],$$

and therefore

$$\|(L-\lambda)\psi_n\| = \hbar\,\|(k-\lambda)\varphi_n\| \to 0.$$

Parseval's identity again gives $\|\psi_n\| = 1$ so we can finally conclude that $\sigma(L)$ is also the entire real line.

Now we want to consider the spectral properties of the Hamiltonian operator H. This will be a little trickier than our previous treatment of the operators X and L since the Hamiltonian operator is dependent on a potential which varies depending on the situation. Recall the Hamiltonian is defined by

$$H = \frac{L^2}{2m} + V.$$

Define the operator $H_0 := \frac{L^2}{2m}$. It is relatively simple to show that H_0 is self-adjoint on $L^2(\mathbb{R}^3)$. The exercise of finding the spectrum of H_0 would seem rather redundant at this point so we will just state

$$\text{The spectrum of } H_0 \text{ is } [0, \infty).$$

That is, the spectrum of H_0 consists of all non-negative real numbers.

So H_0 is fairly easy to deal with, but what happens if we are considering a non-zero potential? We need to ensure that H is self-adjoint on its domain and we would also like this domain to be as large as possible. It is actually possible to construct a potential for which H is only self-adjoint on $D(H) = 0$. It turns out there is a simple theorem which we will state here without proof that allows us to easily determine if H is self-adjoint.

Theorem 59 *If the exist $a, b \in \mathbb{R}$ $(a < 1)$ such that*

$$\|V\psi\| \leq a \|H_0\psi\| + b \|\psi\|, \ \psi \in D(H_0),$$

then H is self-adjoint.

Fortunately there is also a list of equivalent conditions which we may use to verify whether H is self-adjoint as well. The conditions we want to check are as follows:

Theorem 60 *The following statements are equivalent and imply H is self-adjoint:*

1. $D(H_0) \subset D(V)$
2. $\|V\psi\|^2 \leq C(\|H_0\psi\|^2 + \|\psi\|^2)$, $\psi \in D(H_0)$
3. $\sup_x \int_x^{x+1} |V(y)|^2 \, dy < \infty$
4. *For every $\varepsilon > 0$ there is a constant $K \in \mathbb{R}$ such that $\|V\psi\|^2 \leq \varepsilon \|H_0\psi\|^2 + K \|\psi\|^2$, $\psi \in D(H_0)$.*

The last inequality can be replaced by the simpler $\|V\psi\| \leq \varepsilon \|H_0\psi\| + K \|\psi\|$, $\psi \in D(H_0)$.

The possibility of a non-zero potential adds a whole new degree of complexity to the analysis of H. We wont go any further into this here. We have demonstrated the statistical motivation for our definitions of L and X and gone through how to determine their spectrum. The ultimate goal of this process was to be define H which complete defines the time evolution of the state of a system. For a more involved and in depth treatment of the topics covered in the most recent sections the reader is advised to consult [MS].

4.2 Deviation and Uncertainty

One of the well know results of quantum mechanics is Heisenberg's uncertainty principle. This is, generally speaking, a product of the statistical nature of quantum mechanics and the probabilistic interpretation of the state function. More specifically, the uncertainty principle may be viewed as a product of the **canonical commutator relation** $[,]$ (most often just commutator relation) which is often taken as a postulate of quantum mechanics. It may also be derived from our previous work through the definition of the operator's $\mathcal{P}$ and X. This relationship is of particular importance. In the classical theory, observables commute: If we have functions of momentum and position which we represent by observables $\mathcal{P}$ and X then the classical theory tells us that the commutator relation is $[\mathcal{P}, X] = 0$. However, for the postulates of quantum mechanics Born demanded the commutator relation be

$$[\mathcal{P}, X] = \frac{\hbar}{i}$$

where $\hbar = \frac{h}{2\pi}$ is the reduced Plank constant.

With the expectation value of an observable represented by a self-adjoint operator A in state ψ given by

$$\langle A \rangle = \langle A\psi, \psi \rangle$$

we may define the root-mean square deviation from this expectation value as follows

Definition 61 *The **root-mean square deviation** (or **standard deviation**) $(\Delta A)_\psi$ of the expectation value $\langle A \rangle$ of an observable A in state ψ is defined by the expression*

$$(\Delta A)_\psi^2 = \left\langle (A - \langle A\psi, \psi \rangle)^2 \psi, \psi \right\rangle. \tag{4.11}$$

We can easily expand the expression given above in the definition for the standard deviation to obtain the somewhat more intuitive expression

$$\begin{aligned} (\Delta A)_\psi^2 &= \left\langle (A - \langle A\psi, \psi \rangle)^2 \psi, \psi \right\rangle \\ &= \left\langle A^2\psi, \psi \right\rangle - 2 \left\langle \langle A\psi, \psi \rangle \cdot A\psi, \psi \right\rangle + \langle A\psi, \psi \rangle^2 \\ &= \left\langle A^2\psi, \psi \right\rangle - \langle A\psi, \psi \rangle^2 \end{aligned}$$

This naturally leads us to an essential observation of quantum mechanics:

Conclusion 62 *The standard deviation of an observable represented by a self-adjoint operator A is exactly zero if ψ is an eigenvector of A. In this case $\langle A\rangle$ is equal to the corresponding eigenvalue.*

The **bound states** of a quantum mechanical system are defined to be the ψ_n such that $(\Delta A)_{\psi_n} = 0$. Therefore the bounds states can be obtained by solving the eigenvalue equation for a particular operator.

If we consider two Hermitian operators A and B we may use the above definition of standard deviation along with the Cauchy-Schwartz inequality to obtain the following expression

$$\begin{aligned}
\left|\left\langle (A-\overline{A})(B-\overline{B})\psi,\psi\right\rangle\right| &= \left|\left\langle (A-\overline{A})\psi,(B-\overline{B}\psi)\right\rangle\right| \leq \left\|(A-\overline{A})\psi\right\|\left\|(B-\overline{B}\psi)\right\| \\
&= \sqrt{\left\langle (A-\overline{A})\psi,(A-\overline{A})\psi\right\rangle\left\langle (B-\overline{B})\psi,(B-\overline{B}\psi\right\rangle} \\
&= \sqrt{\left\langle (A-\overline{A})^2\psi,\psi\right\rangle\left\langle (B-\overline{B})^2\psi,\psi\right\rangle} \\
&= (\Delta A)_\psi(\Delta B)_\psi
\end{aligned}$$

where $\overline{A} = \langle A\psi,\psi\rangle$ and $\overline{B} = \langle B\psi,\psi\rangle$. This expression will soon prove to be quite useful. For any complex number $z \in \mathbb{C}$ we can write the imaginary part of z as $im(z) = \frac{z-\overline{z}}{2i}$ and it will always be the case that $|im(z)| \leq |z|$. With this in mind, we may write $\left|im(\left\langle (A-\overline{A})(B-\overline{B})\psi,\psi\right\rangle)\right| \leq \left|\left\langle (A-\overline{A})(B-\overline{B})\psi,\psi\right\rangle\right|$. Using the previously mentioned identity we can expand

$$\left|im(\left\langle (A-\overline{A})(B-\overline{B)}\psi,\psi\right\rangle)\right| = \tag{4.12}$$

$$\begin{aligned}
&= \left|\frac{\left\langle (A-\overline{A})(B-\overline{B})\psi,\psi\right\rangle - \overline{\left\langle (A-\overline{A})(B-\overline{B})\psi,\psi\right\rangle}}{2i}\right| \\
&= \frac{\left|\left\langle (A-\overline{A})(B-\overline{B})\psi,\psi\right\rangle - \left\langle \psi,(A-\overline{A})(B-\overline{B})\psi\right\rangle\right|}{2} \\
&= \frac{1}{2}\left|\left\langle (A-\overline{A})(B-\overline{B})\psi,\psi\right\rangle - \left\langle (B-\overline{B)}(A-\overline{A})\psi,\psi\right\rangle\right| \\
&= \frac{1}{2}\left|\left\langle (A-\overline{A})(B-\overline{B})\psi - (B-\overline{B})(A-\overline{A})\psi,\psi\right\rangle\right| \\
&= \frac{1}{2}\left|\left\langle (AB-BA)\psi,\psi\right\rangle\right| = \frac{1}{2}\left|\left\langle [A,B]\psi,\psi\right\rangle\right|
\end{aligned}$$

Now using this result with 4.2 we have

$$\frac{1}{2}\left|\left\langle [A,B]\psi,\psi\right\rangle\right| \leq (\Delta A)_\psi(\Delta B)_\psi \tag{4.13}$$

Note that equation 4.13
is only meaningful for ψ which are not eigenvectors of A or B.

Recall the brief mention of the commutator of two linear operators earlier. It was stated that, for the position and momentum operators X and $\mathcal{P}$ respectively, the commutator relationship was given by $[\mathcal{P}, X] = \frac{\hbar}{i}$.Therefore we can insert this into the above relation to obtain the well known uncertainty principle or **Heisenberg uncertainty principle**

$$\frac{\hbar}{2} = \frac{1}{2}\left|\left\langle \frac{\hbar}{i}1\psi, \psi \right\rangle\right| = \frac{1}{2}\left|\langle[\mathcal{P}, X]\psi, \psi\rangle\right| \leq (\Delta\mathcal{P})_\psi(\Delta X)_\psi.$$

Note that this inequality is independent of ψ i.e. $(\Delta\mathcal{P})_\psi(\Delta X)_\psi \geq \frac{\hbar}{2}$ for any state function ψ.This is incredibly important because it implies that neither $(\Delta\mathcal{P})_\psi$ nor $(\Delta X)_\psi$ ever vanish. This is one of the more mysterious results of the quantum theory. An observation dependent on two variables may never be made with perfect accuracy.

This all seems quite non-intuitive and indeed it was very alarming to many physicists of the early 1900's who where studying these phenomena. If electrons and other particles do exhibit this quantized behavior, then why is it that larger bodies of mass do not appear to? When I look outside my window in the morning there doesn't appear to be any uncertainty as to the position of the trees or birds in the sky, or anything for that matter. But it is important to keep in mind the value of $\hbar = 1.054572 \times 10^{-34} J \cdot s$; it is quite small. $\hbar$ basically acts as a scale of quantum effects, and since it is so small, we do not observe extreme quantum effects on a macroscopic level in our day to day life. The quantization of certain properties e.g. energy, momentum is only noticeable on a very small scale. If we imagine a universe with a different value of $\hbar$ we will be imagining a universe very very different from our own. Since the ionization energy of hydrogen is inversely proportional to $\hbar^2$, if say $\hbar$ where larger than its value in our universe, it is likely atoms would not form. An increase in $\hbar$ would result in a decrease in ionization energy which would make atoms less and less stable as $\hbar$ increased. On the other hand, if $\hbar$ where perhaps smaller, the effects of quantization would be present on a much smaller scale. Perhaps if $\hbar$ had been smaller the instrumentation needed to observe these effects would have taken longer to become available and the transition from classical to quantum theory may have occurred much later in history.

4.3 The Qubit

Our discussion of the basic mathematics of quantum mechanics would not be complete without a brief mention of the qubit. We will describe this notion here and briefly cover an application of group representations that enable us to better understand the qubit. We will avoid going into very much detail here and the reader is referred to [SFS] for a more thorough investigation of these topics. The latter part of this section and the next was added to give the reader a basic idea of how representation theory and other algebraic concepts start to play a very important role in quantum theory and the study of quantum symmetry.

The spin of a quantum particle is an intrinsic form of angular momentum. We may model the space of all spin states for a spin 1/2 particle. This space is called the qubit ("cue-bit"). A spin 1/2 particle begins with two physically distinguishable states which are usually labeled by kets as $|z_+\rangle$ and $|z_-\rangle$. One then assumes that every quantum state may be written as the superposition of kets as

$$c_+ |z_+\rangle + c_- |z_-\rangle , \tag{4.14}$$

where $c_+, c_- \in \mathbb{C}$ and $|c_+|^2 + |c_-|^2 = 1$.

For a spin-1/2 particle we may build a Stern-Gerlach machine which will allows us to determine the orientation of the spin of a particular particle or beam of particles. Those unfamiliar with a Stern Gerlach device may refer to [SFS]. The interpretation we give to the values of c_+ and c_- is as follows: if we put a beam of particles in the state described by expression 4.14 through the Stern-Gerlach machine, the fraction of particles found to be in the state $|z_+\rangle$ will be $|c_+|^2$ and similarly the fraction in state $|z_-\rangle$ is $|c_-|^2$.

We know however, that we may measure the outcome of our Stern Gerlach experiment with the machine oriented in any particular direction we choose and that the spin state of a spin-1/2 particle is completely determined by the probabilities associated with emerging spin up or down along all three coordinate axes $x-$, $y-$, and $z-$. Since the state described in 4.14 is determined by $|c_+|$ and $|c_-|$, the pair $(c_+, c_-) \in \mathbb{C}^2$ contains more information than necessary suggesting there is a more appropriate space to model the qubit. We will not go into details but it turns out that the natural model for a spin-1/2 particle is the projective space $P(\mathbb{C}^2)$ which can be viewed as the $2-sphere$ S^2 in $\mathbb{R}^3$. $P(\mathbb{C}^2)$ is the set of equivalence classes $[c_+ : c_-]$, $c_+, c_- \in \mathbb{C}$ and $|c_+|^2 + |c_-|^2 = 1$, defined by the relation

$$[c_+ : c_-] = [\widetilde{c}_+ : \widetilde{c}_-] \tag{4.15}$$

if and only if

$$(c_+, c_-) = (\lambda\widetilde{c}_+, \lambda\widetilde{c}_-) \tag{4.16}$$

for $\lambda \in \mathbb{C}$ with $|\lambda| = 1$. Viewing $P(\mathbb{C}^2)$ as S^2 we have $[1 : 0]$ as the north pole of S^2 and $[0 : 1]$ the south (along the $z-axis$). Also, $[1 : 1]$ corresponds to the pole of S^2 along the positive $x-axis$, $[-1 : 1]$ is the pole of S^2 along the negative $x-axis$ and $[1 : i]$ and $[1 : -i]$ are the poles of S^2 along the positive and negative $y-axis$, respectively. We then assign the north and south poles of the $2-sphere$ to the spin up and spin down states for a Stern Gerlach machine oriented along the $z-axis$, respectively. That is, $|z_+\rangle := [1 : 0]$ and $|z_-\rangle := [0, 1]$. Similarly we set $|x_+\rangle := [1 : 1]$ and $|x_-\rangle := [-1 : 1]$ where $|x_+\rangle$ is the x-spin up state for a Stern Gerlach machine oriented along the $x-axis$,. $|x_-\rangle$ is the x-spin down state. Finally, $|y_+\rangle := [1 : i]$ and $|y_-\rangle := [1 : -i]$ where $|y_+\rangle$ and $|y_-\rangle$ are the y-spin up and down states respectively. Now we have a more natural space to model the qubit where spin states simply corresponds to points on S^2. This model is much more visually pleasing than the redundant modeling of spin states as points in $\mathbb{C}^2$.

For this model we give the following experimental/predictive interpretation. Suppose we would like to use a Stern Gerlach Machine oriented along the $z-axis$ to create a beam of z-spin up particles and then send them through a Stern Gerlach Machine oriented along the $y-axis$. Using the qubit model we just constructed we can calculate the fraction of particles which emerge y-spin up from the second machine in the following way. Take any point $(y_0, y_1) \in S^3 \subset \mathbb{C}^2$ correspond to the y-spin up state, i.e. choose $(y_0, y_1) \in S^3 \subset \mathbb{C}^2$ such that $[y_0 : y_1] = |y_+\rangle$, say $(y_0, y_1) = \frac{1}{\sqrt{2}}(1, i)$. Similarly take a point $(z_0, z_1) \in S^3 \subset \mathbb{C}^2$, say $(z_0, z_1) = (1, 0)$, such that $[z_0 : z_1] = |z_+\rangle$. The fraction of particles exiting the second machine y-spin up will then be

$$|\langle (y_0, y_1), (z_0, z_1) \rangle|^2, \tag{4.17}$$

or,

$$\left| \left\langle \frac{1}{\sqrt{2}}(1, i), (1, 0) \right\rangle \right|^2 = \frac{1}{2}, \tag{4.18}$$

with $\frac{1}{2}$ of the particles exiting y-spin up, as it should.

Next we want to define a representation which arises from the effect of rotation in three dimensional space on the qubit. We'll need to briefly define a few ideas and objects.

The reader is hopefully familiar with the general linear group and some of its useful subgroups. The **general linear group,** $GL(V)$, of a vector space V is the set of all invertible linear transformations from V to itself. We are interested in the useful subgroups of $GL(\mathbb{C}^2)$ and $GL(\mathbb{R}^3)$,the **special unitary group,** $SU(\mathbb{C}^2) = \{M \in GL(\mathbb{C}^2) : M^*M = I \text{ and } \det(M) = 1\}$ and the **special orthogonal group,** $SO(\mathbb{R}^3) = \{M \in GL(\mathbb{R}^3) : M^T M = I \text{ and } \det(M) = 1\}$. Typically these groups are thought of as matrix groups with group multiplication corresponding to matrix multiplication. $SO(\mathbb{R}^3)$ then consists of all length preserving three dimensional rotations and similarly $SU(\mathbb{C}^2)$ is the set of unitary operators from $\mathbb{C}^2$ to itself.

Since we have defined the projective space $P(\mathbb{C}^2)$ to model the qubit, we need to define the projective unitary group of $\mathbb{C}^2$ which will define a corresponding equivalence class of operators.

Definition 63 *Let V be a complex inner product space. The **projective unitary group of** V is*

$$P\mathcal{U}(V) := U(V)/\backsim \tag{4.19}$$

where $\mathcal{U}(V)$ is the group of unitary operators from V to itself and $T_1 \backsim T_2$ if and only if $T_1 T_2^{-1}$ is a scalar multiple of the identity.

We will not show details here but we may construct a homomorphism $\Phi : SU(\mathbb{C}^2) \to SO(\mathbb{R}^3)$ such that Φ is two-to-one, that is, if $\Phi(U) = \Phi(\widetilde{U}) \in SO(\mathbb{R}^3)$ then $U = \pm \widetilde{U}$ so we can be sure that $[U] = [\widetilde{U}] \in PU(\mathbb{C}^2)$. We can also construct Φ such that

$$\mathbf{X}_\theta = \Phi \begin{pmatrix} e^{-\frac{\theta i}{2}} & 0 \\ 0 & e^{\frac{\theta i}{2}} \end{pmatrix} \tag{4.20}$$

where $\mathbf{X}_\theta \in SO(\mathbb{R}^3)$ is a rotation of angle θ around the $x-axis$.

Definition 64 *For a given vector space V, group G, and homomorphism $\rho : G \to GL(V)$, the triplet (G, V, ρ) is called a* ***representation of*** *G* ***on*** *V.*

Therefore, we may define the representation $\rho_{\frac{1}{2}} : SO(\mathbb{R}^3) \to P\mathcal{U}(\mathbb{C}^2)$ by

$$\rho_{\frac{1}{2}}(g) := [U] \in P\mathcal{U}(\mathbb{C}^2) \tag{4.21}$$

where $U \in SU(\mathbb{C}^2)$, $g \in SO(\mathbb{R}^3)$ satisfies $\Phi(U) = g$. It can be easily verified that $\rho_{\frac{1}{2}}$ is indeed a homomorphism.

It has been shown experimentally that if two observers position differs by a rotation $g \in SO(\mathbb{R}^3)$, then their observations of states of the qubit differ by a projective unitary transformation $[U]$ such that $\Phi(g) = U$. From equation 4.20 we have

$$\rho_{\frac{1}{2}}(\mathbf{X}_\theta) = \left[\begin{pmatrix} e^{-\frac{\theta i}{2}} & 0 \\ 0 & e^{\frac{\theta i}{2}} \end{pmatrix}\right] \tag{4.22}$$

so as an observer physically rotates around the $x-axis$ the corresponding equivalence class rotates at half the speed of the observer. This would seem to be a problem, however we are not dealing with vectors we are dealing with equivalence classes. We need $\rho_{\frac{1}{2}}(\mathbf{X}_0) = \rho_{\frac{1}{2}}(\mathbf{X}_{2\pi})$, (Note $[c_1 : c_2] = [-c_1 : -c_2]$), and indeed

$$\rho_{\frac{1}{2}}(\mathbf{X}_0) = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] = \left[\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right] = \rho_{\frac{1}{2}}(\mathbf{X}_{2\pi}) \tag{4.23}$$

This idea can be generalized to define $\rho_{\frac{n}{2}}$ for a particle with spin $\frac{n}{2}$.

Since it is not our intent to focus on algebraic application in this paper we will stop our discussion of the qubit here but the reader is strongly encouraged to consult [SFS] for a more detailed treatment and continuation of these themes and ideas.

4.4 Symmetry and Time Evolution

Our last major discussion will concern the role played by unitary operators in describing the time evolution of a quantum system and the so called symmetries of a quantum system.

It was said earlier, the Hamiltonian describes the time evolution of a system. Now we will try to be a bit more explicit.

Here we will state the well known Schrödinger equation as a postulate, similar to the treatment in [MS].

Postulate 4	:	If H is the Hamiltonian operator for a particle in state $\psi(t)$ at time t then $\psi(t)$ is a solution of
$i\hbar\psi'(t)$	=	$H\psi(t)$

This postulate describes what is known as the time-dependent Schrödinger picture. Up until this point we have not been very concerned with time dependence. Now we want to distinguish two descriptions of time evolution.

Schrödinger and Heisenberg picture

Given the Hamiltonian H of the a system, the unitary operator

$$U_t = e^{-\frac{iHt}{\hbar}}$$

is called the **time evolution** operator. Given an initial state ψ_0 at $t = 0$, $U_t\psi_0$ gives the state of the system at time t. In the Schrödinger picture, operators are taken to be time-independent, $A(t) = A$, while the states are time dependent, $\psi = \psi(t) = U_t\psi_0$. In the Heisenberg picture, time dependence is transferred to the operators and the states are considered to be time independent i.e. $\psi(t) = \psi$. In the Heisenberg picture, the time dependence of an operator comes again from the time evolution operator in the expression

$$A(t) = U_t^{-1} A_0 U_t.$$

We can now state the equations governing the time dependence of operators or state vectors for either "picture". In the Schrödinger picture, the state ψ of the system with Hamiltonian H must satisfy Schrödinger's equation

$$i\hbar\frac{\partial}{\partial t}\psi = H\psi$$

and we demand that for an operator A,

$$i\hbar\frac{\partial}{\partial t}A \equiv 0$$

For the Heisenberg picture the equations

$$i\hbar\frac{\partial}{\partial t}\psi \equiv 0$$

and

$$i\hbar\frac{\partial}{\partial t}A = [A, H]$$

govern the time dependence of the system with Hamiltonian H in state ψ. Unless otherwise stated, we will continue to work in the Schrödinger picture, assuming the time dependence comes from the state of the system.

Time dependence of expectation values

We naturally want to know how the expectation value of an observable will evolve in time. We have the expression

$$\frac{\partial}{\partial t}\langle A\psi, \psi\rangle = \langle \partial_t(A)\psi, \psi\rangle + \langle A\partial_t(\psi), \psi\rangle + \langle A\psi, \partial_t(\psi)\rangle$$

Applying the Schrödinger equation $i\hbar\partial_t(\psi) = H\psi$ and noting $\partial_t(A) = 0$ we can obtain what is know as Ehrenfest's theorem:

$$\frac{\partial}{\partial t}\langle A\psi, \psi\rangle = \frac{i}{\hbar}\langle [H, A]\psi, \psi\rangle .$$

This tells us that $\frac{\partial}{\partial t}\langle A\psi, \psi\rangle = 0$ precisely when A and H commute, in other words, the expectation value of operators that commute with H are the constants of motion.

Symmetry

We will end with a brief discussion suggesting some of the symmetric qualities of quantum systems. The observables we have discussed act as generators of transformation groups.

Suppose we are given the Hamiltonian H of a system and consider the group G_H of unitary transformations which act invariantly on H, that is $[H, U] = 0$ for all $U \in G_H$. G_H is said to be the set of symmetries of the quantum system.

Just as we defined the time evolution or translation operator U_t we may define translations in position and momentum space as well.

Suppose we wish to define the spatial translation $\psi(r) \to \psi(r + a)$ for some constant vector a in the coordinate space. It can be shown that the unitary operator

$$U_r = e^{-\frac{i}{\hbar}a\cdot\mathcal{P}}$$

gives the desired translation $U_r^{-1}\psi(r)U_r = \psi(r + a)$ and generates the group of spatial translations.

Similarly, we may define translations in momentum space with the unitary operator

$$U_{\mathcal{P}} = e^{-\frac{i}{\hbar}a\cdot X}$$

where X is the position operator. $U_{\mathcal{P}}$ also acts as a generator of the group of translations in momentum space.

To generate the group of spatial rotations one needs the operator L_θ which corresponds to the observable of angular momentum. We will not discuss L_θ in detail however it may be used to define another unitary operator

$$U_\theta = e^{-\frac{i}{\hbar}\varphi e_n\cdot L_\theta}$$

which generates the group of spatial rotations, φ the angle of rotation around fixed axis e_n.

Representation Symmetry Recall definition 64. A **subrepresentation** of a representation $\rho : G \to GL(V)$ is a subspace $W \subset V$ such that for every $g \in G$, $w \in W$, $\rho(g)w \in W$. For a given representation $(G, \mathbb{R}^3, \rho)$ and Hilbert space H we may define an associated representation $\widetilde{\rho}$ of G on H by

$$\widetilde{\rho}\psi(r, t) = \psi(\rho(g^{-1})r, t) \tag{4.24}$$

We say the representation ρ is a **symmetry** of the Hamiltonian H if for every $g \in G$, $\psi \in H$ we have

$$H(\widetilde{\rho}(g)\psi) = \widetilde{\rho}(g)(H\psi). \tag{4.25}$$

Let $\mathcal{A}_H \subset H$ denote the subspace of H consisting of solutions to Schrödinger's equation. Then for $\psi \in \mathcal{A}_H$ we have

$$\begin{aligned} -iH(\widetilde{\rho}(g)\psi) &= -i(\widetilde{\rho}(g)(H\psi)) && (4.26)\\ &= \widetilde{\rho}(g)(\frac{\partial}{\partial t}\psi) && (4.27)\\ &= \frac{\partial}{\partial t}(\widetilde{\rho}(g)\psi). && (4.28) \end{aligned}$$

This shows that if ρ is a symmetry of H then its associated representation $\widetilde{\rho}$ has the subspace $\mathcal{A}_H$ as a subrepresentation. We can also see from equation 4.25 that if $\psi \in H$ is an eigenvector of the Hamiltonian with eigenvalue λ then $\widetilde{\rho}(g)\psi$ is one as well so the subspace $\mathcal{A}_{H,\lambda}$ of eigenvectors corresponding to eigenvalue λ gives another subrepresentation.

One can now use Lie representations to prove conservation of energy and a special case of **Noether's Theorem.** The reader should consult [EJ] for more details and further investigation into representation theory and symmetry.

Bibliography

[AP] A. Pankov, Vinnitsa State Pedagogical University, Vinnitsa, Ukraine, *Introduction to Spectral Theory of Schrodinger Operators,* 2000

[CMMC] Carlos M. Madrid-Casado, *Mathematical History of Wave and Matrix Quantum Mechanics,* Department of Mathematics, Lázaro Cárdenas Institute, and Department of Statistics and Operations Research II, Complutense University, Madrid, Spain , 2007

[EJ] Evan Jenkins, *Representation Theory, Symmetry, and Quantum Mechanics,* notes from lecture given at University of Chicago, August 15, 2008

[EK&GB] E. Kreyszig and Garret Birkhoff, *The Establishment of Functional Analysis,* Historia Mathematica 11 (1984) 258-321

[GB] Garret Birkhoff, *The Establishment of Functional Analysis,* Historia Mathematica 11, 1984, pg 258-321

[HS] Heinrich Saller, *The Basic Physical Lie Operations*, Max-Planck-Institute for Physics, 2004

[JD] John DePree, *Introduction to Real Analysis, John Wiley & Sons, New York, 1988*

[JHW] J.H. Williamson, *Lebesgue Integration,* Holt, Rinhart, and Winston, New York, 1962

[JM] Joseph Maciejko, *Representations of Lorentz and Poincaré groups,* MITP and Department of Physics, Stanford University, Stanford, California, 2012

[JW] Joachim Weidmann, *Linear Operators in Hilbert Space,* Springer Karstedt, 1980

[KS] Konrad Schmüdgen, *Unbounded Self-adjoint Operators on Hilbert Space,* Springer, Dordrecht, 2012

[LB&PM] L. Debnath, P. Mikusinski, *Hilbert Space with Applications*, Elsevier Inc., New York, 2005

[MR&BS] Michael Reed and Barry Simon, *Methods of Modern Mathematical Physics: Functional analysis. 1,* Acadedmic Press, San Diego, California, 1980

[MS] Martin Schecter, *Operator Methods in Quantum Mechanics*, Elsevier North Holland, Inc., New York 1981

[MS2] Martin Schecter, *Spectra of Partial Differential Operators,* American Elsevier Publishing Company, Inc., New York 1971

[ND&JS] Nelson Dunford and Jacob Schwartz, *Linear Operators Parts I-III,* Wiley, New York, 1963

[NPL] N.P. Landsman, *2006 Lecture Notes on Hilbert Spaces and Quantum Mechanics,* Institute for Mathematics, Astrophysics, and Particle Physics Radboud University Nijmegen, The Netherlands, 2006

[NPL2] N.P. Landsman, *Quantum Mechanics and Representation Theory: The New Synthesis,* Kluwer Academic Publishers, 2004

[PRH] Paul R. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity*, AMS Chelsea Publishing, Providence, RI, 1998. Reprint of the second (1957) edition.

[SFS] Stephanie Frank Singer, *Linearity, Symmetry, and Prediction in the Hydrogen Atom,* Springer Science+Business Media, Inc., New York, 2005

[SS] S. Sternberg, *Group theory and physics*, Cambridge University Press, 1994

[TCM] Pdf from Simons Group of University of Cambridge Website, http://www.tcm.phy.cam.ac.uk/~bds10/aqp/handout_operator.pdf, *Chapter 3: Operator Methods in Quantum Mechanics*

[TFJ] Thomas F. Jordan, *Linear Operators for Quantum Mechanics,* John Wiley & Sons, Inc., New York, 1969

[WR] Walter Rudin, *Functional Analysis*, McGraw-Hill, Inc. Singapore, 1991