2009

An Axiomatic Formulation of Quantum Mechanics

Matthew Mastroeni
Ithaca College

AN AXIOMATIC FORMULATION OF QUANTUM MECHANICS

HONORS THESIS

ITHACA COLLEGE

DEPARTMENT OF MATHEMATICS

BY

MATTHEW MASTROENI

FACULTY MENTOR: JAMES CONKLIN

ITHACA, NY

MAY 2009

ABSTRACT

Von Neumann's axiomatic treatment of non-relativistic quantum mechanics is the archetypal example of the dual interaction between physical theories and the development of mathematical ideas. We examine this interaction by first building up the necessary parts of the theory of unbounded self-adjoint operators on a Hilbert space, emphasizing the physical intuition that motivates the mathematical concepts. We then present a version of the Dirac-von Neumann axioms on a quantum system and deduce some of their elementary consequences, illustrating the converse effect of the mathematical formalism on the physical theory.

Gentlemen: there's lots of room left in Hilbert space.   *–S. Mac Lane*

TABLE OF CONTENTS

CHAPTER 1
HILBERT'S SIXTH PROBLEM

When David Hilbert called for the axiomatization of physical theories as part of his notable twenty-three problems, he singled out in particular the applications of probability theory to statistical mechanics and of Lie theory to classical mechanics. However, underpinning these specific examples was certainly the larger realization that there was something to be gained from the application of mathematics to the study of physical phenomena, both for physical theories and for general mathematical knowledge. Hilbert defended this position by asserting:

> I believe that specialization plays an even more important role than generalization when one deals with mathematical problems. Perhaps in most cases in which we seek in vain the answer to a question, the cause of failure lies in the fact that we have worked out simpler and easier problems either not at all or incompletely. What is important is to locate these easier problems and to work out their solutions with tools that are as complete as possible and with concepts capable of generalization. This procedure is one of the most important levers for overcoming mathematical difficulties... , [11].

As a means of promoting a better understanding of the rich interactions between physical theories and mathematical knowledge, we consider as an archetypal example the work initiated and largely carried out by John von Neumann to create a rigorous axiomatic foundation for the theory of quantum mechanics, arguably "the most important axiomatization of a physical theory up to this time," [11]. Quantum mechanics as a physical theory had largely been discovered by around 1925, but it was not until von Neumann had developed many of the central concepts related to unbounded self-adjoint operators on a Hilbert space that the theory could be given

a suitable mathematical framework. Thus, through attempts to clarify the structure of quantum mechanics, mathematics proper gained several results and new ideas capable of generalization much as Hilbert had envisioned.

In what follows, we will develop the necessary portions of the theory of linear operators on a Hilbert space in order to provide a set of axioms that describe a non-relativistic quantum-mechanical system. We make no assumption of previous knowledge concerning Hilbert space or related concepts; we assume only familiarity with linear and abstract algebra, basic point-set topology, and basic measure theory to remain as self-contained as possible. The interested reader is referred to [3] for the relevant topology and to [6] for the relevant measure theory. The aim is to illustrate both the mathematical theory that arose from the need to axiomatize the physics as well as some of the elementary physical implications that follow as direct results of the mathematical theory.

CHAPTER 2
LINEAR OPERATORS ON A HILBERT SPACE

2.1   Inner Product Spaces

As noted above, the central objects of the mathematical formalism of quantum mechanics are precisely Hilbert spaces and the linear operators that act upon them. Before discussing Hilbert spaces, though, we begin with the slightly more general notion of an inner product space. Additionally, we establish the notational convention that we will let F denote the field $\mathbb{R}$ or $\mathbb{C}$.

Definition   Given a vector space V over the field F, an *inner product* on V is a function $\langle \cdot, \cdot \rangle : V \times V \to F$ satisfying for all $u, v, w \in V$ and $a, b \in F$:

1. $\langle v, v \rangle \geq 0$ with $\langle v, v \rangle = 0$ if and only if $v = 0$

2. $\langle au + bv, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle$

3. $\langle v, w \rangle = \overline{\langle w, v \rangle}$

Definition   A vector space V over F together with a specific inner product $\langle \cdot, \cdot \rangle$ on V is an *inner product space.*

Although at first glance the definition of an inner product seems rather arbitrary and abstract, anyone who has ever taken the dot product of two vector in $\mathbb{R}^n$ or $\mathbb{C}^n$ has worked with an inner product since both these vector spaces together with their respective dot products are inner product spaces. Inner product spaces can be thought of as the natural generalization of these mathematical objects. In particular, just as the dot product induces a notion of distance on $\mathbb{R}^n$, we also have the following important functions induced by the inner product on an inner product space.

Definition For V an inner product space, the function $\| \cdot \| : V \to \mathbb{R}$ defined by $\|v\| = \sqrt{\langle v, v \rangle}$ is the *induced norm* on V.

Definition For V an inner product space, the function $d : V \times V \to \mathbb{R}$ defined by $d(u, v) = \|u - v\|$ is the *induced metric* on V.

Moreover, the generalization to inner products also extends the notion of what it means for two vectors to be perpendicular as captured by the dot product on $\mathbb{R}^2$ and $\mathbb{R}^3$ for example. In terms of inner product spaces, this carries over as the concept of orthogonality.

Definition Given vectors $u, v \in V$ an inner product space, u is *orthogonal* to v, denoted $u \perp v$, if $\langle u, v \rangle = 0$.

Definition An *orthornormal set* of vectors in an inner product space V is a set of vectors $\{v_i\}$ such that $v_i \perp v_j$ and $\|v_i\| = 1$ for all $i \neq j$.

In our discussion that follows, we will make frequent reference to the induced norm and implicit reference to the induced metric, and so, we briefly note a few useful results essentially from [8] relating norms and inner products of vectors, the first of which being an excellent example of how inner product spaces expand on the geometry of normal Euclidean space.

Theorem 1. (The Pythagorean theorem) Let V be an inner product space and $\{v_1, \ldots, v_n\} \subseteq V$ an orthonormal set of vectors, then for any $v \in V$:

$$\|v\|^2 = \sum_{i=1}^{n} |\langle v, v_i \rangle|^2 + \left\| v - \sum_{i=1}^{n} \langle v, v_i \rangle v_i \right\|^2$$

Proof: To start off, it is clear that we may express v as $v = \sum_{i=1}^{n} \langle v, v_i \rangle v_i + (v - \sum_{i=1}^{n} \langle v, v_i \rangle v_i)$.

One may easily check that $v - \sum_{i=1}^{n} \langle v, v_i \rangle v_i \perp \sum_{i=1}^{n} \langle v, v_i \rangle v_i$, and so, it follows that:

$$\|v\|^2 = \langle v, v \rangle = \left\langle \sum_{i=1}^{n} \langle v, v_i \rangle v_i , \sum_{j=1}^{n} \langle v, v_j \rangle v_j \right\rangle + \left\langle v - \sum_{i=1}^{n} \langle v, v_i \rangle v_i , v - \sum_{j=1}^{n} \langle v, v_j \rangle v_j \right\rangle$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle v, v_i \rangle \overline{\langle v, v_j \rangle} \langle v_i, v_j \rangle + \left\| v - \sum_{i=1}^{n} \langle v, v_i \rangle v_i \right\|^2$$

$$= \sum_{i=1}^{n} |\langle v, v_i \rangle|^2 + \left\| v - \sum_{i=1}^{n} \langle v, v_i \rangle v_i \right\|^2$$

In particular, because we know that for any $u \in V$ that $\|u\|^2 = \langle u, u \rangle \geq 0$, we have the immediate consequences:

Corollary 1. (Bessel's inequality) Let V be an inner product space and $\{v_1, \ldots, v_n\} \subseteq V$ an orthonormal set of vectors, then for any $v \in V$:

$$\|v\|^2 \geq \sum_{i=1}^{n} |\langle v, v_i \rangle|^2$$

Corollary 2. (The Cauchy-Schwarz inequality) Let V be an inner product space and $u, v \in V$, then:

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

Proof: If $v = 0$, then clearly we have $\|v\| = \langle u, v \rangle = 0$ so that result is trivial. So we may assume $v \neq 0$, in which case $\|v\| \neq 0$ so that $v/\|v\|$ by itself is an orthonormal set. Bessel's inequality then implies:

$$\|u\|^2 \geq \left| \left\langle u, \frac{v}{\|v\|} \right\rangle \right|^2 = \frac{|\langle u, v \rangle|^2}{\|v\|^2}$$

So we see that $|\langle u, v \rangle| \leq \|u\|\|v\|$ as desired.

With these considerations, it is not overly difficult to verify that the induced norm and induced metric are in fact a norm and a metric respectively in the sense of normed linear spaces and metric spaces; that is, the induced norm has the following properties:

Proposition 1.   Let V be an inner product space with $u, v \in V$ and $a \in F$, then the induced norm on V satisfies:

1. $\|v\| \geq 0$ with $\|v\| = 0$ if and only if $v = 0$

2. $\|av\| = |a| \, \|v\|$

3. $\|u + v\| \leq \|u\| + \|v\|$

4. $|\, \|u\| - \|v\| \,| \leq \|u - v\|$

Proposition 2. (The Parallelogram law)   Let V be an inner product space and $u, v \in V$, then we have:

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2)$$

Hence, as we continue to discuss inner product spaces, especially Hilbert spaces, we can consider natural questions associated with metric spaces such as those related to the convergence of sequences.

2.2   Hilbert Spaces

We now specialize to the case of Hilbert spaces. As noted above, there are topological considerations associated with inner product spaces, and specifically, this is what distinguishes a Hilbert space from a general inner product space.

Definition   An inner product space $\mathscr{H}$ is a *Hilbert space* if it is complete with respect to the metric induced by its inner product.

Again, as with inner products, it would be nice to have an intuitive picture of what a Hilbert space should "look like." And just as before, normal Euclidean space with the dot product can serve as a somewhat simplistic model, since the dot product induces the usual notion of distance and Euclidean space is complete with respect to the induced metric. More importantly for later on, in our notation the underlying field F for a Hilbert space $\mathscr{H}$ is always a Hilbert space itself.

Just as in algebra how it is natural to consider the subsets of a given object that retain much of the same structure as the given object, we also consider the such subsets of Hilbert spaces.

Definition   A subset $\mathscr{M}$ of a Hilbert space $\mathscr{H}$ is a *linear manifold* if for all $u, v \in \mathscr{M}$ and $a, b \in F$, $au + bv \in \mathscr{M}$.

While linear manifolds do retain the vector space structure of a Hilbert space, they are not quite Hilbert spaces in their own rights in general. This is due to the fact that Cauchy sequences of elements of any linear manifold $\mathscr{M}$ may not converge to an element of $\mathscr{M}$, although they certainly converge in the Hilbert space $\mathscr{H}$. Thus, for a linear manifold to be a Hilbert space in its own right, we must require that it be closed in the norm topology on $\mathscr{H}$.

Closed linear manifolds have nice properties associated with them, one of which we now demonstrate. But before doing so, we extend the notion of orthogonality slightly.

Definition For $U \subseteq V$ an inner product space and $v \in V$, v is *orthogonal* to U, denoted $v \perp U$, if $v \perp u$ for every $u \in U$.

Theorem 2. (The Projection theorem) Let $\mathcal{M}$ be a closed linear manifold in a Hilbert space $\mathcal{H}$, then for every $v \in \mathcal{H}$ there exists a unique vector u of minimal distance from v such that $v - u \perp \mathcal{M}$.

Proof: We begin by defining $\delta = \inf_{w \in \mathcal{M}} \|v - w\|$, which clearly exists since the distances $\|v - w\|$ are bounded below by 0. By the definition of δ, we know that there must be a sequence in $\{w_n\} \subseteq \mathcal{M}$ such that $\{\|v - w_n\|\} \to \delta$. We show that $\{w_n\}$ is Cauchy.

Let $\epsilon > 0$. Since $\{\|v - w_n\|\} \to \delta$, we know that $\exists N_1 \in \mathbb{N}$ such that $\|v - w_n\| < \frac{\epsilon}{2}$ whenever $n \geq N_1$. Assume $n, m \geq N_1$. Then, it follows that:

$$\|w_n - w_m\| = \|w_n - v + v - w_m\| \leq \|w_n - v\| + \|w_m - v\|$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Thus, $\{w_n\}$ is Cauchy, and so $\{w_n\} \to u$ for some $u \in \mathcal{H}$ since $\mathcal{H}$ is complete. Moreover, $u \in \mathcal{M}$ as $\mathcal{M}$ is closed.

To see that $\|v - u\| = \delta$, and hence that u is of minimal distance from v, note that since $\{w_n\} \to u \ \exists N_2 \in \mathbb{N}$ such that $\|w_n - u\| < \epsilon$ whenever $n \geq N_2$, from which it follows that:

$$\big| \|v - w_n\| - \|v - u\| \big| \leq \|w_n - v + v - u\| = \|w_n - u\| < \epsilon$$

Hence, we see that $\{\|v - w_n\|\} \to \|v - u\|$, but since we also know that $\{\|v - w_n\|\} \to \delta$, we must have that $\|v - u\| = \delta$.

Assume there is another vector $z \in \mathcal{M}$ such that $\|v - z\| = \delta$. Then, $\frac{1}{2}(u + z) \in \mathcal{M}$ since $\mathcal{M}$ is a linear manifold. So $\|v - \frac{1}{2}(u + z)\| \geq \delta$. But we also note that $\|v - \frac{1}{2}(u + z)\| \leq \|\frac{1}{2}(v - u)\| + \|\frac{1}{2}(v - z)\| = \frac{1}{2}(\|v - u\| + \|v - z\|) = \delta$. Hence,

$\|v - \frac{1}{2}(u + z)\| = \delta$. The Parallelogram law then implies that $4\delta^2 + \|u - z\|^2 = 4\|v - \frac{1}{2}(u+z)\|^2 + \|u - z\|^2 = \|2v - (u+z)\|^2 + \|u - z\|^2 = 2(\|v - u\|^2 + \|v - z\|^2) = 4\delta^2$, or equivalently, $\|u - z\| = 0$. Thus, $u = z$, and u is unique.

Finally, to see that $v - u \perp \mathscr{M}$, let $z \in \mathscr{M}$ and $t \in \mathbb{R}$. Then, $u + tz \in \mathscr{M}$ so that:

$$
\begin{aligned}
\delta^2 \leq \|v - (u + tz)\|^2 &= \langle v - (u + tz), v - (u + tz)\rangle \\[2mm]
&= \langle v - u, v - u\rangle - t\langle z, v - u\rangle - t\langle v - u, z\rangle + t^2\langle z, z\rangle \\[2mm]
&= \|v - u\|^2 - 2t\,\mathrm{Re}\,\langle v - u, z\rangle + t^2\|z\|^2 \\[2mm]
&= \delta^2 - 2t\,\mathrm{Re}\,\langle v - u, z\rangle + t^2\|z\|^2
\end{aligned}
$$

This shows that $-2t\,\mathrm{Re}\,\langle v - u, z\rangle + t^2\|z\|^2 \geq 0$ so that we have a real polynomial always at most zero, and so, it must have at most one real root. The quadratic formula then forces $4(\mathrm{Re}\,\langle v - u, z\rangle)^2 \leq 0$, or equivalently, $\mathrm{Re}\,\langle v - u, z\rangle = 0$.

If $\mathscr{H}$ is a Hilbert space over $\mathbb{R}$, then $\langle v - u, z\rangle \in \mathbb{R}$ so we are done. Otherwise, $\mathscr{H}$ is a Hilbert space over $\mathbb{C}$ so that $u + itz \in \mathscr{M}$, and a similar argument to the one above shows that $\mathrm{Im}\,\langle v - u, z\rangle = 0$. Thus, in general, $\langle v - u, z\rangle = 0$, and $v - u \perp z$. Furthermore, because z was arbitrary, we see that $v - u \perp \mathscr{M}$.

The preceding theorem shows us not only why closed linear manifolds are nice objects to work with, but also how the concept of orthogonality plays an important role in determining the structure of a Hilbert space; above, this role allowed us to uniquely decompose any given vector v as $u + w$ for some $u \in \mathscr{M}$ and $w \perp \mathscr{M}$. In a larger sense, orthogonality allows us to arrive at the notion of a basis for a Hilbert space similar to the one that it so useful in dealing with vector spaces.

Definition   A *basis* for a Hilbert space $\mathscr{H}$ is a maximal orthonormal set in $\mathscr{H}$.

That is, a basis is an orthonormal set for which there does not exist any other orthonormal set properly containing it. As with vector spaces, a simple Zorn's lemma argument shows that every Hilbert space has a basis. The following theorem illustrates the similarity of a Hilbert space basis with a vector space basis. A good treatment of the proof can be found in [4].

Theorem 3. Let $\mathscr{H}$ be a Hilbert space with basis $\{v_i\}_{i\in\mathcal{I}}$ (not necessarily countable), then for every $v \in \mathscr{H}$:

$$v = \sum_{i\in\mathcal{I}} \langle v, v_i \rangle \, v_i$$

and

$$\|v\|^2 = \sum_{i\in\mathcal{I}} |\langle v, v_i \rangle|^2$$

where the sum converges to v in the sense that a sum over any finite set of indices can be made arbitrarily close to v provided that it contains a certain subset of indices.

Notice that the second part of the above is in essence a generalization of the Pythagorean theorem for inner product spaces to not necessarily finite sums.

In addition to the preceding property of bases, it would also be nice if the cardinality of bases for a given Hilbert space was an invariant as it is for vector spaces. Again, this turns out to be the case. However, we omit the proofs of the following results once more; while they are important ideas worth noting, they play only a subsidiary role in our subsequent consideration of quantum mechanics. See [2] for a complete proof.

Theorem 4. Let $\mathscr{H}$ be a Hilbert space, then any two bases for $\mathscr{H}$ have the same cardinality.

Definition The *dimension* of a Hilbert space $\mathscr{H}$ is the cardinality of any basis.

In particular, the next theorem tells us when a Hilbert space is somewhat easier to have a handle on.

Theorem 5. Let $\mathscr{H}$ be a Hilbert space, then $\mathscr{H}$ is separable if and only if it has countable basis.

Separability is, therefore, a very nice property for a Hilbert space to have since it assures us that we have to manage at worst countably infinitely many basis vectors. This will be especially useful when we discuss operators on Hilbert spaces. Thus, from this point onward, all Hilbert spaces mentioned are assumed to be separable.

2.3   Linear Operators

Hilbert spaces serve as the stage on which quantum mechanics is carried out. We now shift our attention to the principal actors– linear operators. Furthermore, although we have not yet defined what linear operators are, we establish the convention that we will often abbreviate the action of the operator A on the vector v to Av, as opposed to the usual $A(v)$, as a means of simplifying expressions.

Definition   Given Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$ over F, a *linear operator* is a linear map A, denoted by $A : \mathscr{H} \to \mathscr{K}$, whose domain $\operatorname{dom} A$ is a linear manifold in $\mathscr{H}$.

The key point of the above definition is that linear operators need not be defined on the entire Hilbert space $\mathscr{H}$. It would be nice if we could simply disregard those operators not everywhere defined, but it is those operators that turn out to particularly important in certain cases.

In addition, we would like to give suitable definitions for addition of linear operators and other such manipulations that exist for linear maps between vector spaces. However, the fact that linear operators need not be everywhere defined complicates the issue and forces us to be careful about what the domain of our new operator should be. With these concerns in mind, though, we can define the algebra for linear

operators in a logical way. For $A : \mathscr{H} \to \mathscr{K}$, $B : \mathscr{H} \to \mathscr{K}$, and $c \in F$, we define $A + B$ in the usual way to be $(A+B)v = Av + Bv$ with $\operatorname{dom} A + B = \operatorname{dom} A \cap \operatorname{dom} B$, and we define cA also in the usual way to be $(cA)v = cAv$ with $\operatorname{dom} cA = \operatorname{dom} A$. There is also the usual composite BA of operators $A : \mathscr{H} \to \mathscr{K}$ and $B : \mathscr{K} \to \mathscr{V}$ provided we restrict the domain of the composite to be $\operatorname{dom} BA = A^{-1}(\operatorname{dom} B)$. A special case of these definitions worth noting is the *bracket* $[A, B]$ of two operators $A : \mathscr{H} \to \mathscr{H}$ and $B : \mathscr{H} \to \mathscr{H}$, given by $[A, B] = AB - BA$. It's domain of definition can be deduced easily from the basic combinations of operators already mentioned.

2.3.1 Bounded Operators

It should be clear from the preceding discussion that general linear operators are a general pain to work with; so, now we consider our first type of better-behaved linear operators.

Definition   A linear operator $A : \mathscr{H} \to \mathscr{K}$ is called *bounded* if $\exists\, c > 0$ such that for all $v \in \operatorname{dom} A$:

$$\|Av\| \leq c\,\|v\|$$

One of the reasons that bounded operators is a nice variety of linear operator to deal with is that it is always possible to extend a bounded linear operator $A : \mathscr{H} \to \mathscr{K}$ with domain $\operatorname{dom} A$ to a bounded linear operator with domain $\overline{\operatorname{dom} A}$. For any $v \in \overline{\operatorname{dom} A}$, we know that there must be a sequence $\{v_n\} \subseteq \operatorname{dom} A$ such that $\{v_n\} \to v$. Since $\{v_n\}$ converges, it must be Cauchy, and this forces the sequence $\{Av_n\}$ to be Cauchy because A is a bounded operator. The fact that $\mathscr{K}$ is a Hilbert space then implies that $\{Av_n\}$ converges to some $w \in \mathscr{K}$. It is not overly difficult to show that this w is independent of the sequence converging to v. Hence, we may extend A by defining $Av = w$ for $v \in \overline{\operatorname{dom} A}$. With a little work, one may check that the extension A is bounded and linear as desired.

If we also restrict ourselves to considering only everywhere defined operators, then it turns out that bounded operators are extremely well-behaved. The collection of all such operators from a Hilbert space $\mathscr{H}$ to a Hilbert space $\mathscr{K}$ is given the special designation $\mathscr{L}(\mathscr{H}, \mathscr{K})$, and if $\mathscr{H} = \mathscr{K}$, we simply abbreviate the previous notation to $\mathscr{L}(\mathscr{H})$. We note that, when operators are everywhere defined, all the care we had to take concerning the domains of general linear operators is needless; it follows directly from the definitions given for general linear operators that linear combinations of everywhere defined operators are also everywhere defined. It is also rather clear that linear combinations of operators in $\mathscr{L}(\mathscr{H}, \mathscr{K})$ are bounded, and so, $\mathscr{L}(\mathscr{H}, \mathscr{K})$ has the structure of a vector space over F. Moreover, if $\mathscr{H} = \mathscr{K}$, we see that composites of operators in $\mathscr{L}(\mathscr{H})$ are everywhere defined and bounded as well, and so, $\mathscr{L}(\mathscr{H})$ has the structure of a linear associative algebra over F. Recalling the definition of the bracket of two operators, we see that $[A, B] \in \mathscr{L}(\mathscr{H})$ for $A, B \in \mathscr{L}(\mathscr{H})$, and $\mathscr{L}(\mathscr{H})$ is actually a Lie algebra for which the bracket we defined turns out to be the natural Lie bracket on a linear associative algebra.

Recalling that the underlying field F is a Hilbert space, we make the following definition:

Definition   A linear operator $A : \mathscr{H} \to F$ is called a *linear functional*, and the collection $\mathscr{L}(\mathscr{H}, F)$ is denoted by $\mathscr{H}^*$ and called the *dual space* of $\mathscr{H}$.

Another reason bounded operators behave nicely owes to the fact that they turn out to be continuous in the norm topology on the Hilbert space $\mathscr{H}$, as can easily be shown. Combining this with the above definition, we demonstrate a significant result in which the continuity of bounded linear functionals plays an important role.

Theorem 6.   (The Riesz Representation theorem)   Let $L \in \mathscr{H}^*$, then there exists a unique vector $w \in \mathscr{H}$ such that for every $v \in \mathscr{H}$:

$$Lv = \langle v, w \rangle$$

Proof: We begin with a proof of existence. Let $\mathscr{M} = \ker L$. As in vector space theory, it is clear that $\ker L$ is a linear manifold, but it is also closed since L is continuous and 0 is closed in the standard topology on F. If $\mathscr{M} = \mathscr{H}$, then for every $v \in \mathscr{H}$ we have that $Lv = 0 = \langle v, 0 \rangle$ so we are done.

So we may assume that $\mathscr{M} \neq \mathscr{H}$. The Projection theorem then yields a vector $u \in \mathscr{M}$ such that $v' - u \perp \mathscr{M}$ for some $v' \notin \mathscr{M}$. Note that $v' - u \notin \mathscr{M}$ for if $v' - u \in \mathscr{M}$, then $Lv' = Lu + L(v' - u) = 0 + 0 = 0$ contradicting our assumption that $v' \notin \mathscr{M}$. In particular, $L(v' - u) \neq 0$ and $v' - u \neq 0$ as $0 \in \mathscr{M}$. Hence, there is a vector $u' \neq 0 \perp \mathscr{M}$ such that $Lu' = 1$ which can be found explicitly by taking $u' = \frac{1}{L(v'-u)}(v' - u)$. Also, note that for any $v \in \mathscr{H}$, $v - L(v)u' \in \mathscr{M}$ since $L(v - L(v)u') = Lv - Lv = 0$. And so, if we define $w = \frac{1}{\|u'\|^2} u'$, we see that:

$$
\begin{aligned}
0 = \langle v - L(v)u', u' \rangle &= \langle v, u' \rangle - Lv \langle u', u' \rangle \\[2ex]
&= \langle v, \|u'\|^2 w \rangle - Lv \|u'\|^2 \\[2ex]
&= (\langle v, w \rangle - Lv) \|u'\|^2
\end{aligned}
$$

As $u' \neq 0$, we know that $\|u'\|^2 \neq 0$, forcing $Lv = \langle v, w \rangle$ for any $v \in \mathscr{H}$ as v was arbitrary.

To see that w is unique, assume there is another vector $z \in \mathscr{H}$ such that $Lv = \langle v, w \rangle = \langle v, z \rangle$ for all $v \in \mathscr{H}$. Then, $\langle v, w - z \rangle = \langle v, w \rangle - \langle v, z \rangle = 0$ for every $v \in \mathscr{H}$ so that, specifically, $\langle v - z, v - z \rangle = 0$. Thus, $v = z$.

2.3.2   Self-adjoint Operators

As mentioned when we first began our discussion of the mathematical formalism of quantum mechanics, the theory of not necessarily bounded self-adjoint linear operators was one of the vital components to a rigorous foundation of the theory of

quantum mechanics. This is because self-adjoint operators lie at the heart of quantum mechanics, acting as physical quantities that can be observed and measured. To get to these special operators though, we must first introduce the concept of the adjoint of a linear operator.

To define the adjoint of a linear operator, it is useful to consider operators with domains that are dense subsets of a Hilbert space $\mathscr{H}$, henceforth densely defined operators. The benefit of restricting ourselves to densely defined operators should soon be apparent.

Consider a densely defined operator $A : \mathscr{H} \to \mathscr{K}$ and the set:

$$\operatorname{dom} A^* = \{w \in \mathscr{K} \mid \varphi_w v = \langle Av, w \rangle \text{ is a bounded linear functional on } \operatorname{dom} A\}$$

Because each φ_w is bounded on $\operatorname{dom} A$, we can extend φ_w to a bounded operator on $\overline{\operatorname{dom} A}$ as described in § 2.3.1. And since we assumed that $\operatorname{dom} A$ was dense, we know that $\overline{\operatorname{dom} A} = \mathscr{H}$ so that the extension of φ_w is everywhere defined as well as bounded. The Riesz Representation theorem then applies so that there is a unique $u \in \mathscr{H}$ such that $\langle v, u \rangle = \varphi_w v = \langle Av, w \rangle$ for each $v \in \operatorname{dom} A$. Define an operator $A^* : \mathscr{K} \to \mathscr{H}$ with domain $\operatorname{dom} A^*$ by $A^* w = u$. One may verify that A^* is in fact linear. By construction, A^* is the unique linear operator such that $\langle v, A^* w \rangle = \langle Av, w \rangle$ for every $v \in \operatorname{dom} A$ and $w \in \operatorname{dom} A^*$. This is the quintessential property of adjoints.

Definition The *adjoint* of a linear operator $A : \mathscr{H} \to \mathscr{K}$ is the unique linear operator $A^* : \mathscr{K} \to \mathscr{H}$ such that $\langle v, A^* w \rangle = \langle Av, w \rangle$ for every $v \in \operatorname{dom} A$ and $w \in \operatorname{dom} A^*$.

For operators in $\mathscr{L}(\mathscr{H})$, we know that adjoints always exist since operators in $\mathscr{L}(\mathscr{H})$ are everywhere defined, hence obviously densely defined. Moreover, there are nice explicit relationships between combinations of operators in $\mathscr{L}(\mathscr{H})$ and their constituents as follows:

Proposition 3. Let $A, B \in \mathscr{L}(\mathscr{H})$ and $c \in F$, then $A^*, B^* \in \mathscr{L}(\mathscr{H})$ and:

1. $(A + B)^* = A^* + B^*$

2. $(cA)^* = \bar{c}A^*$

3. $(A^*)^* = A$

4. $(AB)^* = B^*A^*$

Thus, we see that the adjoints of the well-behaved operators in $\mathscr{L}(\mathscr{H})$ are correspondingly well-behaved in ways that do not necessarily hold for general linear operators.

As an example, consider the operators $0, I \in \mathscr{L}(\mathscr{H})$ where 0 is the operator that sends everything in $\mathscr{H}$ to 0 and I is the identity operator on $\mathscr{H}$. Clearly, both are in $\mathscr{L}(\mathscr{H})$ as claimed since $\|0v\| = \|0\| = 0 \leq \|v\|$ and $\|Iv\| = \|v\|$ for all $v \in \mathscr{H}$. Note also that $\langle 0v, w \rangle = \langle 0, w \rangle = 0 = \langle v, 0 \rangle = \langle v, 0w \rangle$ and $\langle Iv, w \rangle = \langle v, w \rangle = \langle v, Iw \rangle$ for all $v, w \in \mathscr{H}$. But since 0^* and I^* are the unique operators such that $\langle v, 0^*w \rangle = \langle 0v, w \rangle$ and $\langle v, I^*w \rangle = \langle Iv, w \rangle$ for all $v, w \in \mathscr{H}$, we must have that $0^* = 0$ and $I^* = I$. This is not true in general for all operators in $\mathscr{L}(\mathscr{H})$, but it demonstrates that adjoints for such well-behaved operators are not overly difficult to find.

Additionally, we might further ask how the adjoint A^* is related to the original operator A, and the simplest answer to this question is precisely the kind of operator that we care about.

Definition A densely-defined operator $A : \mathscr{H} \to \mathscr{H}$ is *self-adjoint* if $A = A^*$.

It is important to note that cached within the condition for self-adjointness is the implicit assumption that the domains of A and A^* coincide. Proposition 3. shows that this is not an issue for operators in $\mathscr{L}(\mathscr{H})$, as we saw concretely in the case of the self-adjoint operators 0 and I, but it is another complication for general operators. Even in general though, self-adjoint operators have significant properties that are critical for quantum mechanics. These properties will be fully explained in the Spectral theorem at the end of the next section.

2.3.3 Orthogonal Projections

Bridging the gap between bounded and self-adjoint operators is a third type of operator– orthogonal projections.

Definition An operator $P \in \mathscr{L}(\mathscr{H})$ is an *orthogonal projection* if $P^2 = P$ and $P = P^*$.

Orthogonal projections capture much of the structure inherent in a Hilbert space $\mathscr{H}$ in that there is a one-to-one correspondence between orthogonal projections and closed linear manifolds in $\mathscr{H}$.

To see that each orthogonal projection P corresponds to a closed linear manifold, note that $I - P \in \mathscr{L}(\mathscr{H})$ since $I \in \mathscr{L}(\mathscr{H})$ and $\mathscr{L}(\mathscr{H})$ is a vector space as noted in § 2.3.1. From this, it is easy to check that $\ker I - P = \{x \in \mathscr{H} \mid (I - P)x = 0\} = \{x \in \mathscr{H} \mid Px = x\} = \operatorname{im} P$. Since 0 is a closed set in $\mathscr{H}$, we see that $\ker I - P = \operatorname{im} P$ must be closed as $I - P$ is bounded, hence continuous. Thus, we associate the closed linear manifold $\operatorname{im} P$ with the orthogonal projection P. The reverse correspondence comes from the following result which relies heavily on the Projection theorem of § 2.2:

Proposition 4. Let $\mathscr{M}$ be a closed linear manifold in the Hilbert space $\mathscr{H}$ and define the operator $P : \mathscr{H} \to \mathscr{H}$ for every $v \in \mathscr{H}$ by $Pv = u$ for $u \in \mathscr{M}$ the unique vector such that $v - u \perp \mathscr{M}$, then $P \in \mathscr{L}(\mathscr{H})$ and satsifies:

1. $P^2 = P$

2. $\|Pv\| \leq \|v\|$ for all $v \in \mathscr{H}$

3. $P^* = P$

Hence, the operator P defined above is an orthogonal projection. Furthermore, if Q is an orthogonal projection with $\operatorname{im} Q = \mathscr{M}$, then for any $u \in \mathscr{M}$, there exists a $w \in \mathscr{H}$ such that $Qw = u$, which implies that $Qu = Q^2w = Qw = u$. So, for any $v \in \mathscr{H}$, $\langle v - Qv, u \rangle = \langle v, u \rangle - \langle Qv, u \rangle = \langle v, u \rangle - \langle v, Q^*u \rangle = \langle v, u \rangle - \langle v, Qu \rangle = \langle v, u \rangle - \langle v, u \rangle = 0$ and $v - Qv \perp \mathscr{M}$ as u was arbitrary. But since Pv is the unique vector in $\mathscr{M}$ such that $v - Pv \perp \mathscr{M}$, we must have $Pv = Qv$. And as v was arbitrary, $P = Q$. Thus, every closed linear manifold also determines a unique orthogonal projection of which it is the image.

The capacity of orthogonal projections to capture the structure of a Hilbert space is precisely the tool we need to describe the important properties of self-adjoint operators. We do this with the aid of the following:

Definition   Given a measurable space X with σ-algebra $\mathfrak{F}$ and a Hilbert space $\mathscr{H}$, a *projection-valued measure* on X is a function $\mathrm{P} : \mathfrak{F} \to \mathscr{L}(\mathscr{H})$ satisfying:

1. $\mathrm{P}(E)$ is an orthogonal projection for each $E \in \mathfrak{F}$

2. $\mathrm{P}(\emptyset) = 0$ and $\mathrm{P}(X) = I$

3. For $E = \coprod_{n=1}^{\infty} E_n \in \mathfrak{F}$:

$$\mathrm{P}(E) = \lim_{n \to \infty} \sum_{i=1}^{n} \mathrm{P}(E_i)$$

where convergence of the above limit is understood to be pointwise convergence in the norm topology on $\mathscr{H}$.

An important instance of projection-valued measures for our purposes will be when the measurable space is $\mathbb{R}$ with the Borel σ-algebra $\mathscr{B}(\mathbb{R})$. Notice that for a fixed $v \in \mathscr{H}$, the assignment $\mu : \mathscr{B}(\mathbb{R}) \to \mathbb{R}$ defined by $\mu(E) = \langle \mathrm{P}(E)v, v \rangle$ is in fact a finite measure on $\mathbb{R}$. It is straightforward that $\mu(\emptyset) = \langle \mathrm{P}(\emptyset)v, v \rangle = \langle 0v, v \rangle = \langle 0, v \rangle = 0$. We also note that $\mu(E) \geq 0$ for each $E \in \mathscr{B}(\mathbb{R})$ since $\langle Pv, v \rangle \geq 0$ for any

orthogonal projection P; this will be shown explicitly in the next section. Finally, for $E = \coprod_{n=1}^{\infty} E_n \in \mathscr{B}(\mathbb{R})$, we see that:

$$\mu(E) = \langle \mathrm{P}(E)v, v \rangle = \langle \lim_{n \to \infty} \sum_{i=1}^{n} \mathrm{P}(E_i)v, v \rangle = \lim_{n \to \infty} \sum_{i=1}^{n} \langle \mathrm{P}(E_i)v, v \rangle = \sum_{i=1}^{\infty} \mu(E_i)$$

so that μ is a measure as claimed. The finiteness of μ follows similarly since $\mu(\mathbb{R}) = \langle \mathrm{P}(\mathbb{R})v, v \rangle = \langle Iv, v \rangle = \langle v, v \rangle$.

Knowing that $\langle \mathrm{P}(\,\cdot\,)v, v \rangle$ is a measure, we adopt the convention of writing $\mathrm{P}(\lambda)$ in place of $\mathrm{P}((-\infty, \lambda))$ for each $\lambda \in \mathbb{R}$.

Definition A measurable function f on $\mathbb{R}$ is *finite a.e.* with respect to projection-valued measure P if it is finite a.e. with respect to the measure $\langle \mathrm{P}(\,\cdot\,)v, v \rangle$ for all $v \in \mathscr{H}$.

With these considerations, we are ready to state without proof the projection-valued measure version of the Spectral theorem. See [8] for a discussion of the proof.

Theorem 7. (The Spectral theorem) Let A be a self-adjoint linear operator on the Hilbert space $\mathscr{H}$, then:

1. For every $v \in \operatorname{dom} A$, there is a unique projection-valued measure P on $\mathbb{R}$ with respect to the Borel σ-algebra $\mathscr{B}(\mathbb{R})$ such that:

$$\langle Av, v \rangle = \int_{-\infty}^{\infty} \lambda \, d\langle \mathrm{P}(\lambda)v, v \rangle$$

which we denote by:

$$Av = \int_{-\infty}^{\infty} \lambda \, d\mathrm{P}(\lambda)$$

Additionally, every projection-valued measure P on $\mathbb{R}$ with respect to $\mathscr{B}(\mathbb{R})$ determines a linear operator A via the above expression with domain

$$\operatorname{dom} A = \left\{ v \in \mathscr{H} \ \middle|\ \int_{-\infty}^{\infty} \lambda^2 \, d\langle \mathrm{P}(\lambda)v, v \rangle < \infty \right\}$$

on which it is self-adjoint.

2. For every measurable function g on $\mathbb{R}$ finite a.e. with respect to the projection-valued measure P from 1., there is a linear operator $g(A)$ defined by the expression:

$$\langle g(A)v, v \rangle = \int_{-\infty}^{\infty} g(\lambda) \, d\langle \mathrm{P}(\lambda)v, v \rangle$$

which we denote by:

$$g(A)v = \int_{-\infty}^{\infty} g(\lambda) \, d\mathrm{P}(\lambda)$$

that has the domain

$$\operatorname{dom} g(A) = \left\{ v \in \mathscr{H} \ \middle|\ \int_{-\infty}^{\infty} |g(\lambda)|^2 \, d\langle \mathrm{P}(\lambda)v, v \rangle < \infty \right\}$$

on which it is self-adjoint.

The great utility of self-adjoint operators, as the Spectral theorem shows us, is then that we may express inner products with self-adjoint operators as integrals on the real line, and specifically, those integrals have forms reminiscent of expected values of continuous random variables and functions of continuous random variables from probability theory. We noted previously that self-adjoint operators play the role of physical quantities that can be observed and recorded in quantum mechanics, and so, loosely speaking, taking inner products with self-adjoint operators yields

the expected value that a certain physical quantity will give a specific result when measured.

These are all rather imprecise ideas from a mathematical standpoint and will be clarified later, but they convey the physical intuition fairly well and should justify our interest in such operators.

2.3.4   Trace Class Operators

Trace class operators, when subject to a few extra conditions, are the other fundamental type of operator for quantum mechanics, essentially encapsulating all physical conditions that make up the state of a quantum system. However, as with self-adjoint operators, we discover the utility of trace-class operators by first introducing an auxiliary concept.

Furthermore, in developing the theory of quantum mechanics, we will be concerned only with infinite-dimensional complex Hilbert spaces, and so, from this point on, all Hilbert spaces are assumed to be infinite-dimensional and complex as well as separable. That is, by Theorem 4. of § 2.2, the Hilbert space $\mathscr{H}$ has a countably infinite basis.

Definition   A linear operator $A \in \mathscr{L}(\mathscr{H})$ is *positive*, denoted by $A \geq 0$, if $\langle Av, v \rangle \geq 0$ for all $v \in \mathscr{H}$.

Positive operators fit in well amidst the various kinds of operators we have already mentioned. By definition, they are bounded. If $\mathscr{H}$ is a complex Hilbert space, it is also possible to show that they are self-adjoint. Hence, for our purposes, positive operators are also self-adjoint. Moreover, orthogonal projections are always positive since for P an orthogonal projection and $v \in \mathscr{H}$ we have $\langle Pv, v \rangle = \langle P^2 v, v \rangle = \langle Pv, Pv \rangle \geq 0$. It can also be shown that a positive operator A satisfies the Cauchy-Schwarz type inequality $|\langle Av, w \rangle|^2 \leq \langle Av, v \rangle \langle Aw, w \rangle$ for all $v, w \in \mathscr{H}$.

The name positive operator is somewhat suggestive since, as the next result shows, they behave to a degree like nonnegative real numbers.

Theorem 8. (The Square Root lemma) Let $A \in \mathscr{L}(\mathscr{H})$ and $A \geq 0$, then there exists a unique $B \in \mathscr{L}(\mathscr{H})$ such that $B \geq 0$ and $B^2 = A$.

Definition For $A \in \mathscr{L}(\mathscr{H})$ and $A \geq 0$, the operator $\sqrt{A}$ is the unique operator in $\mathscr{L}(\mathscr{H})$ such that $A \geq 0$ and satisfies $(\sqrt{A})^2 = A$.

By way of motivation for the definition of trace class operators, we digress momentarily to reflect on some basic measure theory. Recall that the class of Lebesgue integrable functions on $\mathbb{R}$ is defined to be the set of all functions defined on $\mathbb{R}$ such that the Lebesgue integral $\int_{-\infty}^{\infty} |f(x)| \, dx$ is finite, and prior to defining the Lebesgue integral for all measurable functions, the integral is first defined for characteristic and then nonnegative simple functions. It then follows from properties of the Lebesgue integral that the assignment $f \mapsto \int_{-\infty}^{\infty} f(x) \, dx$ is a linear map, although $L^1(\mathbb{R})$ is not a Hilbert space.

We make these observations because our approach to trace class operators follows essentially the same route with positive operators standing in for nonnegative simple functions. And in the process, we obtain an analogous linear map, the trace, on the collection of all trace class operators. But true to the aforementioned pattern, we first we define the trace for positive operators.

Definition For $\mathscr{H}$ a Hilbert space with basis $\{v_i\}$ and $A \in \mathscr{L}(\mathscr{H})$ with $A \geq 0$, the *trace* of A is:

$$\operatorname{tr} A = \sum_{i=1}^{\infty} \langle Av_i, v_i \rangle$$

Proposition 5. Let $A \in \mathscr{L}(\mathscr{H})$ with $A \geq 0$, then $\operatorname{tr} A$ is well-defined.

Proof: Let $\{v_i\}$ and $\{w_i\}$ be two bases for $\mathscr{H}$. We begin by noting that, since A is assumed to be the positive, $\operatorname{tr} A$ takes values in $[0, \infty]$ for any basis. Then, as $A \geq 0$, $\sqrt{A}$ exists and is positive, hence self-adjoint. And so, appealing to Theorem 3. of § 2.2 shows that:

$$\sum_{i=1}^{\infty} \langle Av_i, v_i \rangle = \sum_{i=1}^{\infty} \langle (\sqrt{A})^2 v_i, v_i \rangle \;=\; \sum_{i=1}^{\infty} \langle \sqrt{A}v_i, \sqrt{A}v_i \rangle$$

$$= \sum_{i=1}^{\infty} \|\sqrt{A}v_i\|^2$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |\langle \sqrt{A}v_i, w_j \rangle|^2$$

Because the sums involve only nonnegative quantities, we may reverse the order of the sums since if they converge, they do so absolutely so that:

$$\sum_{i=1}^{\infty} \langle Av_i, v_i \rangle = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} |\langle \sqrt{A}v_i, w_j \rangle|^2 \;=\; \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} |\langle \sqrt{A}w_j, v_i \rangle|^2$$

$$= \sum_{j=1}^{\infty} \|\sqrt{A}w_j\|^2$$

$$= \sum_{j=1}^{\infty} \langle Aw_j, w_j \rangle$$

Thus, we see that $\operatorname{tr} A$ is independent of the basis chosen.

Having a well-defined trace for positive operators, we can define trace class operators. Notice that for any $v \in \mathscr{H}$ and $A \in \mathscr{L}(\mathscr{H})$ $\langle A^*Av, v \rangle = \langle Av, Av \rangle \geq 0$ so that $A^*A \geq 0$, which motivates the following definition:

Definition For $A \in \mathscr{L}(\mathscr{H})$, the operator $|A|$ is defined by $|A| = \sqrt{A^*A}$.

Definition A linear operator $A \in \mathscr{L}(\mathscr{H})$ is *trace class* if and only if $\operatorname{tr}|A| < \infty$. The subset of $\mathscr{L}(\mathscr{H})$ of trace class operators is denoted by $\mathscr{I}_1$.

Definition The *trace* is the map $\operatorname{tr} : \mathscr{I}_1 \to \mathbb{C}$ defined for all $A \in \mathscr{I}_1$ by:

$$\operatorname{tr} A = \sum_{i=1}^{\infty} \langle Av_i, v_i \rangle$$

where $\{v_i\}$ is a basis for the Hilbert space $\mathscr{H}$.

As with the trace defined on positive operators, it can be proved that the trace on trace class operators converges absolutely and is independent of the basis chosen for $\mathscr{H}$. Moreover, the trace on trace class operators is an extension of the trace on positive operators with finite trace since for every positive operator A we know that $A^* = A$ so that $|A| = \sqrt{A^2} = A$, and so, $A \in \mathscr{I}_1$ provided $\operatorname{tr} A < \infty$.

Additionally, the set of trace class operators $\mathscr{I}_1$ and the trace itself have nice properties summarized as follows. The proofs of these facts are somewhat too involved to admit a nice presentation here, but we collect the results for later reference and refer the interested reader to [8].

Theorem 9. Let $A, B \in \mathscr{I}_1$, $C \in \mathscr{L}(\mathscr{H})$, and $c, d \in \mathbb{C}$, then the following are in $\mathscr{I}_1$:

1. $cA + dB$

2. AC and CA

3. A^*

Theorem 10. Let $A, B \in \mathscr{I}_1$, $C \in \mathscr{L}(\mathscr{H})$, and $c, d \in \mathbb{C}$, then:

1. $\operatorname{tr} cA + dB = c \operatorname{tr} A + d \operatorname{tr} B$

2. $\operatorname{tr} AC = \operatorname{tr} CA$

3. $\operatorname{tr} A^* = \overline{\operatorname{tr} A}$

We remarked at the beginning of this section that trace class operators are critical to the mathematical formalism of quantum mechanics and that they capture the physical conditions of a quantum system in some way. Intuitively speaking, it makes sense that the state of a quantum system should have some bearing on the outcome of attempting to measure a quantity associated with the system. As previously noted, quantum mechanics is a probabilistic theory, and so this dependence on the state of a system manifests itself as trace class operators that alter the probability distribution that a measurement of a given physical quantity will yield a specific result. The trace itself serves as one of the primary components for constructing the probability measure. Again, this is merely a rough sketch outlining the importance of trace class operators, but we will make the details exact shortly.

2.4   One-Parameter Groups

The final type of operator on a Hilbert space that we discuss is unitary operators, which determine the time evolution of quantum systems.

Definition   An operator $U \in \mathscr{L}(\mathscr{H})$ is *unitary* if $UU^* = U^*U = I$.

Equivalently, the definition of a unitary operator implies that such an operator U is invertible as a linear map and that $U^* = U^{-1}$. It is not hard to check that if U is unitary then so is U^{-1}. In relation to previously defined concepts, unitary operators have the nice properties:

Proposition 6.   Let $U \in \mathscr{L}(\mathscr{H})$ be unitary and P be a projection-valued measure on the measurable space X with σ-algebra $\mathfrak{F}$, then $U\mathrm{P}(\,\cdot\,)U^{-1}$ is also a projection-valued measure.

Proof: Since P is a projection-valued measure, it is immediate that $U\mathrm{P}(\emptyset)U^{-1} = U0U^{-1} = 0$ and $U\mathrm{P}(X)U^{-1} = UIU^{-1} = I$. Let $E \in \mathfrak{F}$. Then since $\mathrm{P}(E), U, U^{-1} \in \mathscr{L}(\mathscr{H})$, $U\mathrm{P}(E)U^{-1} \in \mathscr{L}(\mathscr{H})$, and so, Proposition 3. yields $(U\mathrm{P}(E)U^{-1})^* =$

$(U\mathrm{P}(E)U^*)^* = (U^*)^*(U\mathrm{P}(E))^* = U\mathrm{P}(E)^*U^* = U\mathrm{P}(E)U^{-1}$. It is also easy to see that $(U\mathrm{P}(E)U^{-1})^2 = U\mathrm{P}(E)U^{-1}U\mathrm{P}(E)U^{-1} = U\mathrm{P}(E)^2U^{-1} = U\mathrm{P}(E)U^{-1}$ so that $U\mathrm{P}(E)U^{-1}$ is an orthogonal projection.

Assume $E = \coprod_{n=1}^{\infty} E_n \in \mathfrak{F}$. Let $\epsilon > 0$. Since $\mathrm{P}(E) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{P}(E_i)$, for every $v \in \mathscr{H}$ $\exists N \in \mathbb{Z}^+$ such that $\|(\mathrm{P}(E) - \sum_{i=1}^{n} \mathrm{P}(E_i))v\| < \epsilon$ whenever $n \geq N$. In particular, there is an N for $U^{-1}v$ for arbitrary v. Assume that $n \geq N$. Then:

$$\left\| \left(U\mathrm{P}(E)U^{-1} - \sum_{i=1}^{n} U\mathrm{P}(E_i)U^{-1} \right) v \right\|$$

$$= \left\| U \left(\mathrm{P}(E) - \sum_{i=1}^{n} \mathrm{P}(E_i) \right) U^{-1}v \right\|$$

$$= \sqrt{ \left\langle \left(\mathrm{P}(E) - \sum_{i=1}^{n} \mathrm{P}(E_i) \right) U^{-1}v, U^*U \left(\mathrm{P}(E) - \sum_{i=1}^{n} \mathrm{P}(E_i) \right) U^{-1}v \right\rangle }$$

$$= \left\| \left(\mathrm{P}(E) - \sum_{i=1}^{n} \mathrm{P}(E_i) \right) U^{-1}v \right\| < \epsilon$$

Thus, $U\mathrm{P}(E)U^{-1} = \lim_{n\to\infty} \sum_{i=1}^{n} U\mathrm{P}(E_i)U^{-1}$, and $U\mathrm{P}(\,\cdot\,)U^{-1}$ is a projection-valued measure.

For each $t \in \mathbb{R}$ and self-adjoint A on $\mathscr{H}$, the Spectral theorem gives that there exists a self-adjoint operator e^{itA}. This operator turns out to be unitary, moreover, and it has the attributes:

Proposition 7. Let A be a self-adjoint operator on the Hilbert space $\mathscr{H}$ and $\{U(t)\}$ the collection of unitary operators such that $U(t) = e^{itA}$ for each $t \in \mathbb{R}$, then for all $s, t \in \mathbb{R}$:

1. $U(s)U(t) = U(s+t)$

2. $\lim_{s\to t} U(s) = U(t)$

3. If $\lim\limits_{t \to 0} \dfrac{U(t) - I}{t} v$ exists, then $v \in \text{dom}\, A$

4. For each $v \in \text{dom}\, A$, $\lim\limits_{t \to 0} \dfrac{U(t) - I}{t} v = iAv$

The previous result motivates making the following definition.

Definition A *strongly continuous one-paramter group* $\{A(t)\}$ is the image of a group homomorphism $\phi : \mathbb{R} \to \mathscr{L}(\mathscr{H})$ continuous in the norm topology on $\mathscr{H}$ such that $\phi(t) = A(t)$ and satisfies for all $s, t \in \mathbb{R}$:

1. $A(s)A(t) = A(s + t)$

2. $A(0) = I$

3. $\lim\limits_{s \to t} A(s) = A(t)$

where convergence of the above limit is understood to be pointwise.

The analogy with the time evolution of quantum systems should be somewhat clear since we often think of time as a parameter. The converse to Proposition 7. is given by:

Theorem 11. (Stone's theorem)   Let $\{U(t)\}$ be a strongly continuous one-parameter unitary group on the Hilbert space $\mathscr{H}$, then there exists a self-adjoint operator A on $\mathscr{H}$ such that $U(t) = e^{itA}$ for each $t \in \mathbb{R}$. It has domain:

$$\text{dom}\, A = \left\{ v \in \mathscr{H} \;\middle|\; \lim_{t \to 0} \frac{U(t) - I}{t} v \quad \text{exists} \right\}$$

and for all $v \in \text{dom}\, A$ is given by:

$$\lim_{t \to 0} \frac{U(t) - I}{t} v = iAv$$

Definition   Given a strongly continuous one-parameter unitary group, the self-adjoint operator A such that $U(t) = e^{itA}$ for all t is the *infinitesimal generator* of the one-parameter group.

As concerns quantum mechanics, the self-adjoint operator A guaranteed to exist by Stone's theorem is often rescaled by a real constant to obtain the operator $H = -\hbar A$, where $\hbar$ is Planck's constant. This operator is easily seen to be self-adjoint as well. It is called a *Hamiltonian operator*. Substituting for A, we see that $U(t) = e^{-\frac{i}{\hbar}tH}$ for each t so that H is more frequently considered the infinitesimal generator for $\{U(t)\}$ within the context of quantum mechanics.

CHAPTER 3
THE DIRAC-VON NEUMANN AXIOMS

At this point, we have developed the mathematics that grew out of the need to axiomatize quantum mechanics, and we are now ready to see how the various pieces fit together to provide a rigorous mathematical framework for the theory. In the presentation of axioms that follows, we essentially follow [9], giving a variant of the Dirac-von Neumann axioms on a quantum system. We make no claim as to the ability of these axioms to model physical reality. Instead, we simply note, as frequently claimed, that experimental results regarding non-relativistic phenomena continue to agree perfectly with the predictions of the mathematical framework.

3.1   Axioms of Quantum Mechanics

Formally, we can understand quantum mechanics as a theory with the concepts of a quantum system, observable, state, and measurement as its undefined terms subject to the following axioms:

1. To every *quantum system* there is an associated separable infinite-dimensional complex Hilbert space $\mathscr{H}$.

2. There is a one-to-one correspondence between *observables* and self-adjoint operators on the Hilbert space $\mathscr{H}$. The set of all self-adjoint operators on $\mathscr{H}$ is denoted by $\mathscr{A}$.

3. There is a one-to-one correspondence between *states* and operators $M \in \mathscr{I}_1$ on the Hilbert space $\mathscr{H}$ such that $M \geq 0$ and $\operatorname{tr} M = 1$. The set of all such trace class operators is denoted by $\mathscr{S}$.

4. There is a one-to-one correspondence between *measurements* and assignments to each ordered pair $(A, M) \in \mathscr{A} \times \mathscr{S}$ of a Borel probability measure $\mu_{A,M}$ on $\mathbb{R}$.

   For each Borel set $E \in \mathscr{B}(\mathbb{R})$, $\mu_{A,M}(E)$ is the probability that a measurement of the quantum system in the state represented by M of the observable represented by A gives a value in E.

5. For each $(A, M) \in \mathscr{A} \times \mathscr{S}$, the measure $\mu_{A,M}$ is given by:

$$\mu_{A,M}(E) = \operatorname{tr} \mathrm{P}_A(E)M$$

   for each $E \in \mathscr{B}(\mathbb{R})$, where P_A is the unique projection-valued measure associated with the self-adjoint operator A on $\mathbb{R}$ with the σ-algebra $\mathscr{B}(\mathbb{R})$.

Observe how the axioms match our intuition about physical systems that guided our development of the mathematics. Specifically, the likelihood of measuring a specific value for an observable of a quantum system depends directly on the observable in question and the state of the system via the probability measure $\mu_{A,M}$, which we will prove is a probability measure in the next section. Intuitively, we understand the state of a quantum system to be all the relevant physical conditions that might affect the outcome of an experiment. Observables can thought of as the important physical quantities associated with a quantum system such as the momentum or energy of a particle for example. But beyond this natural model, we could technically consider a quantum system with observables and states to be anything that satisfies the above axioms.

As we now turn to deducing some elementary consequences of these axioms, we will be slightly more casual with our terminology. While it makes sense as a formal theory to speak of the undefined terms such as an observable, because there is a one-to-one correspondence between observables and self-adjoint operators, we will frequently identify an observable with its corresponding self-adjoint operator A and speak of the observable A, understanding implicitly that we mean the observable

represented by A. Similarly, we identify a state with it associated trace class operator M and so on for measurements and the quantum system.

3.2   Some Consequences

Perhaps a somewhat dissatisfying first observation is that, unlike certain sets of operators that we worked with in the pure mathematical theory, the sets $\mathscr{A}$ and $\mathscr{S}$ do not have the nice structure of vector spaces since the sum of two unbounded self-adjoint operators need not be self-adjoint, and if $M, N \in \mathscr{S}$, then clearly $\operatorname{tr} M + N = \operatorname{tr} M + \operatorname{tr} N = 2$ by Theorem 10. The one interesting property of the set $\mathscr{S}$ is that it is *convex* in the sense that, for any $M, N \in \mathscr{S}$ and $0 \leq t \leq 1$, $tM + (1-t)N \in \mathscr{S}$. With Theorem 10., this is also not hard to verify. A simple induction argument can then show that any *convex combination* $\sum_{i=1}^{n} a_i M_i$ such that $\sum_{i=1}^{n} a_i = 1$ with $M_i \in \mathscr{S}$ and $0 \leq a_i \leq 1$ for each i is also in $\mathscr{S}$.

Note also that for $v \in \mathscr{H}$ with $\|v\| = 1$, the orthogonal projection P onto the closed linear manifold generated by v is in $\mathscr{S}$ since v is a basis for the linear manifold it generates so that $\operatorname{tr} P = \langle Pv, v \rangle = \langle v, v \rangle = 1$. This suggests the following definition.

Definition   A *pure state* is an orthogonal projection onto a one-dimensional linear manifold in $\mathscr{H}$. For $v \in \mathscr{H}$ with $\|v\| = 1$, we denote orthogonal projection onto the linear manifold generated by v as P_v. A *mixed state* is a state that is not pure.

The reason for the terminology pure state and mixed state is explained by the result proved in [9]:

Proposition 8.   Let $M \in \mathscr{S}$, then M is a pure state if and only if it is not a nontrivial convex combination of states.

As a physical theory, the crux of quantum mechanics lies in the ability to perform probabilistic computations, such as the expected position of an electron at a given moment in time for example. From a mathematical standpoint, however, we should first verify that the function $\mu_{A,M}$ given to us by the axioms for each observable A and state M is in fact a probability measure as claimed.

Proposition 9. Let $A \in \mathscr{A}$ and $M \in \mathscr{S}$, then $\mu_{A,M}$ is a probability measure on $\mathbb{R}$.

Proof: Let $\{v_i\}$ be a basis for $\mathscr{H}$. Clearly, we have that $\mu_{A,M}(\emptyset) = \operatorname{tr} \mathrm{P}_A(\emptyset)M = \operatorname{tr} 0 = \sum_{i=1}^{\infty} \langle 0v_i, v_i \rangle = 0$ since P_A is a projection-valued measure. Also, $\mu_{A,M}(\mathbb{R}) = \operatorname{tr} \mathrm{P}_A(\mathbb{R})M = \operatorname{tr} M = 1$.

For arbitrary $E \in \mathscr{B}(\mathbb{R})$, first note that $\mu_{A,M}(E)$ is finite since $\mathrm{P}_A(E) \in \mathscr{L}(\mathscr{H})$ as it is an orthogonal projection, and so, $\mathrm{P}_A(E)M \in \mathscr{I}_1$ by Theorem 9. So $\operatorname{tr} \mathrm{P}_A(E)M$ converges absolutely for any basis.

Let $\mathscr{M} = \operatorname{im} \mathrm{P}_A(E)$. Then $\mathscr{M}$ is a closed linear manifold, and hence, a Hilbert space in its own right. Thus, $\mathscr{M}$ has a basis $\mathcal{B}$. A Zorn's lemma argument similar to the one that shows that every Hilbert space has a basis shows that there is a basis $\mathcal{A}$ for $\mathscr{H}$ such that $\mathcal{B} \subseteq \mathcal{A}$. Assume $\mathcal{B}$ is a proper subset of $\mathcal{A}$ so that we may write $\mathcal{A} = \mathcal{B} \coprod \mathcal{C}$ for some set $\mathcal{C} \neq \emptyset$. Since $\mathcal{B}$ is a basis for $\mathscr{M}$, each $w \in \mathscr{M}$ may be expressed as $w = \sum_{u \in \mathcal{B}} \langle w, u \rangle u$. And so, for each $v \in \mathcal{C}$, one may check that $\langle v, w \rangle = \sum_{u \in \mathcal{B}} \langle w, u \rangle \langle v, u \rangle = 0$. This shows that $\mathrm{P}_A(E)v = 0$ for each $v \in \mathcal{C}$ since 0 must be the unique vector such that $v - 0 = v \perp \mathscr{M}$. Additionally, the discussion following Proposition 4. shows that $\mathrm{P}_A(E)v = v$ for each $v \in \mathcal{B}$. These facts together with the absolute convergence of $\operatorname{tr} \mathrm{P}_A(E)M$ and the positivity of M then imply that:

$$\mu_{A,M}(E) = \operatorname{tr} \mathrm{P}_A(E)M = \sum_{v \in \mathcal{B}} \langle \mathrm{P}_A(E)Mv, v \rangle + \sum_{v \in \mathcal{C}} \langle \mathrm{P}_A(E)Mv, v \rangle$$

$$= \sum_{v \in \mathcal{B}} \langle Mv, \mathrm{P}_A(E)v \rangle + \sum_{v \in \mathcal{C}} \langle Mv, \mathrm{P}_A(E)v \rangle$$

$$= \sum_{v \in \mathcal{B}} \langle Mv, v \rangle \geq 0$$

If $\mathcal{A} = \mathcal{B}$, a comparable argument can be made to show that, in general, $\mu_{A,M}(E) \geq 0$.

Assume $E = \coprod_{n=1}^{\infty} E_n \in \mathscr{B}(\mathbb{R})$. We claim that for each i:

$$\langle \mathrm{P}_A(E)Mv_i, v_i \rangle = \lim_{n \to \infty} \sum_{j=1}^{n} \langle \mathrm{P}_A(E_j)Mv_i, v_i \rangle$$

Let $\epsilon > 0$. Since P_A is a projection-valued measure, $\mathrm{P}_A(E) = \lim_{n \to \infty} \sum_{j=1}^{n} \mathrm{P}_A(E_j)$ where convergence is pointwise. That is, for every $v \in \mathscr{H}$, $\exists N \in \mathbb{Z}^+$ such that $\|(\mathrm{P}_A(E) - \sum_{j=1}^{n} \mathrm{P}_A(E_j))v\| < \epsilon$ whenever $n \geq N$. In particular, for each i, $\exists N_i$ for Mv_i. Assume $n \geq N_i$. Then, the Cauchy-Schwarz inequality shows that:

$$\left| \langle \mathrm{P}_A(E)Mv_i, v_i \rangle - \left\langle \sum_{j=1}^{n} \mathrm{P}_A(E_j)Mv_i, v_i \right\rangle \right| = \left| \left\langle \left(\mathrm{P}_A(E) - \sum_{j=1}^{n} \mathrm{P}_A(E_j) \right) Mv_i, v_i \right\rangle \right|$$

$$\leq \left\| \left(\mathrm{P}_A(E) - \sum_{j=1}^{n} \mathrm{P}_A(E_j) \right) Mv_i \right\| \|v_i\|$$

$$= \left\| \left(\mathrm{P}_A(E) - \sum_{j=1}^{n} \mathrm{P}_A(E_j) \right) Mv_i \right\|$$

$$< \epsilon$$

Hence, we see that $\langle \mathrm{P}_A(E)Mv_i, v_i \rangle = \lim_{n \to \infty} \sum_{j=1}^{n} \langle \mathrm{P}_A(E_j)Mv_i, v_i \rangle$ for each i. From this, it follows that:

$$\mu_{A,M}(E) = \mathrm{tr}\, \mathrm{P}_A(E)M = \sum_{i=1}^{\infty} \langle \mathrm{P}_A(E)Mv_i, v_i \rangle$$

$$= \sum_{i=1}^{\infty} \left(\lim_{n \to \infty} \sum_{j=1}^{n} \langle \mathrm{P}_A(E_j)Mv_i, v_i \rangle \right)$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \langle P_A(E_j)Mv_i, v_i \rangle$$

$$= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \langle P_A(E_j)Mv_i, v_i \rangle$$

$$= \sum_{j=1}^{\infty} \operatorname{tr} P_A(E_j)M$$

$$= \sum_{j=1}^{\infty} \mu(E_j)$$

where the rearrangement of sums is justified since we know that $\operatorname{tr} P_A(E)M$ converges absolutely.

Therefore, $\mu_{A,M}$ is in fact a probability measure on $\mathbb{R}$.

Having established that $\mu_{A,M}$ is in fact a probability measure for any $A \in \mathscr{A}$ and $M \in \mathscr{S}$, we can consider all the usual concepts of probability such as expected values and variance. We noted earlier that one of the primary uses of the measure $\mu_{A,M}$ is for computing expected values of observables. This suggests that for a fixed state M, we define the expected value of the observable A to be:

$$E_M[A] = \int_{-\infty}^{\infty} \lambda \, d\mu_{A,M}(\lambda)$$

where we adapt our notation from § 2.3.3 to the measure $\mu_{A,M}$ by denoting $\mu_{A,M}((-\infty, \lambda))$ by $\mu_{A,M}(\lambda)$ for each $\lambda \in \mathbb{R}$. From this, one can then compute the associated variance and standard deviation. This should seem reasonable since, intuitively, observables are the measurable physical quantities associated with a quantum system.

Theorem 12. Let $A \in \mathscr{A}$ and $M \in \mathscr{S}$ such that $E_M[A] < \infty$ and $\operatorname{im} M \subseteq \operatorname{dom} A$, then $AM \in \mathscr{I}_1$ and:

$$\mathrm{E}_M[A] = \operatorname{tr} AM$$

Moreover, if $M = P_v$ and $v \in \operatorname{dom} A$, then we have that $\mathrm{E}_M[A] = \langle Av, v \rangle$ and $\mathrm{E}_M[A^2] = \|Av\|^2$.

Using the above result, we prove a fundamental consequence of the mathematical structure of quantum mechanics.

Theorem 13. (The Heisenberg Uncertainty relation) Let $A, B \in \mathscr{A}$ and $M = P_v$ such that $v, Av, Bv \in \operatorname{dom} A \cap \operatorname{dom} B$, then:

$$\mathrm{SD}_M[A] \, \mathrm{SD}_M[B] \; \geq \; \frac{1}{2} \, |\langle i[A, B]v, v \rangle|$$

Proof: Let $M = P_v$. We begin by noting that for any $\lambda, \mu \in \mathbb{C}$ $[A - \lambda I, B - \mu I] = [A, B]$. That is, the domains of both operators coincide and the operators are equal everywhere on their common domain. Also, it is easy to check that, for example, $(A - \lambda I)^* = A^* - \overline{\lambda} I$. This suggests defining the operators $A' = A - \mathrm{E}_M[A]I$ and $B' = \mathrm{E}_M[B]I$. The previously mentioned facts then imply that $[A', B'] = [A, B]$ and $A', B' \in \mathscr{A}$. Furthermore, by Theorem 11., $\mathrm{E}_M[A'] = \langle (A - \mathrm{E}_M[A]I)v, v \rangle = \langle Av, v \rangle - \mathrm{E}_M[A] \langle v, v \rangle = \mathrm{E}_M[A] - \mathrm{E}_M[A] = 0$ since $\|v\| = 1$ by assumption so that $\mathrm{Var}_M[A'] = \mathrm{E}_M[(A' - \mathrm{E}_M[A']I)^2] = \mathrm{E}_M[A'^2] = \mathrm{E}_M[(A - \mathrm{E}_M[A]I)^2] = \mathrm{Var}_M[A]$. Similar arguments also hold for B'. Hence, it is sufficient to show that:

$$\mathrm{SD}_M[A'] \, \mathrm{SD}_M[B'] \; \geq \; \frac{1}{2} \, |\langle i[A', B']v, v \rangle|$$

Another application of Theorem 10. and the Cauchy-Schwarz inequality yield the desired relation since:

$$|\langle i[A', B']v, v \rangle| \;\; = \;\; |\langle B'A'v, v \rangle - \langle A'B'v, v \rangle|$$

$$= \ |\langle A'v, B'v \rangle - \langle B'v, A'v \rangle|$$

$$= \ 2 \, |\operatorname{Im} \langle A'v, B'v \rangle|$$

$$\leq \ 2 \, |\langle A'v, B'v \rangle|$$

$$\leq \ 2 \, \|A'v\| \, \|B'v\|$$

$$= \ 2 \operatorname{SD}_M[A'] \operatorname{SD}_M[B']$$

The Heisenberg Uncertainty relation is often hailed as one of the most basic, yet paradoxical results of the quantum mechanics. As an inequality, it seems harmless enough, but it gives an important lower bound on the standard deviations of observables when two observables do not commute. The paragon example of this is related to another axiom of quantum mechanics that we do not treat here. Namely, it postulates the existence of self-adjoint operators P and Q corresponding to the observables of momentum and position respectively related by $[P, Q] = -i\hbar I$. For these operators and v and M satisfying the hypotheses of the theorem, applying the Heisenberg inequality yields $\operatorname{SD}_M[P] \operatorname{SD}_M[Q] \geq \frac{1}{2}|\hbar\langle Iv, v\rangle| = \frac{h}{2} > 0$. This is unsettling from a physical point of view because it asserts that we can never know both the position and momentum of such a quantum system exactly, which may seem very counterintuitive to our experiences of reality on the macroscopic scale. Nonetheless, it is a biproduct of the mathematical formalism of quantum mechanics.

3.3   Dynamics

Although there is much to be said of the view of quantum mechanics we have presented already, we do not usually think of physical systems as being static. In

order to handle such changes in a quantum system, we must introduce an additional axiom that determines its dynamical behavior. However, in doing so we are confronted with a choice. Of the two principal actors related to a quantum system, observables and states, there is no absolute reason for choosing one over the other as the agent of change in a quantum system. We conclude our discussion of quantum mechanics by proving that this seeming ambivalence is actually another consequence of the mathematical structure that we have already postulated.

On the one hand, if we assume that the states of a quantum system change over time and not the observables, we are adopting the Schrödinger picture of quantum mechanics encapsulated in the axiom:

Axiom 1. There exists a strongly continuous one-parameter unitary group $\{U(t)\}$ on $\mathscr{H}$ and a pair of bijections for each $t \in \mathbb{R}$ both denoted by U_t such that:

$$U_t : \mathscr{A} \to \mathscr{A} \quad \text{is given by} \quad U_t(A) = A(t) = A$$

$$U_t : \mathscr{S} \to \mathscr{S} \quad \text{is given by} \quad U_t(M) = M(t) = U(t)MU(t)^{-1}$$

For H the infinitesimal generator of $U(t)$:

$$\frac{dM}{dt} = i\hbar \, [H, M] \qquad \text{for all} \qquad M \in \mathscr{S}$$

Alternatively, we could adopt the converse view that the observables change over time and the states are fixed. This is the Heisenberg picture of dynamics and is axiomatized as follows.

Axiom 2. There exists a strongly continuous one-parameter unitary group $\{U(t)\}$ on $\mathscr{H}$ and a pair of bijections for each $t \in \mathbb{R}$ both denoted by U_t such that:

$$U_t : \mathscr{A} \to \mathscr{A} \quad \text{is given by} \quad U_t(A) = A(t) = U(t)^{-1}AU(t)$$

$$U_t : \mathscr{S} \to \mathscr{S} \quad \text{is given by} \quad U_t(M) = M(t) = M$$

For H the infinitesimal generator of $U(t)$:

$$\frac{dA}{dt} = -i\hbar\,[H, A] \qquad \text{for all} \qquad A \in \mathscr{A} \text{ and } A \in \mathscr{L}(\mathscr{H})$$

These axioms are actually equivalent in the following sense:

Theorem 14. Let $t \in \mathbb{R}$, then for any $A \in \mathscr{A}$ and $M \in \mathscr{S}$ we have that $\mu_{A,M(t)} = \mu_{A(t),M}$ for $M(t) = U(t)MU(t)^{-1}$ and $A(t) = U(t)^{-1}AU(t)$.

Proof: Let $E \in \mathscr{B}(\mathbb{R})$. We claim that $\mathrm{P}_{A(t)} = U(t)^{-1}\mathrm{P}_A U(t)$. We know that for U a unitary operator and P a projection-valued measure $U\mathrm{P}(\cdot)U^{-1}$ is also a projection-valued measure, so the claim should seem plausible at least. Assume $v \in \mathrm{dom}\,A(t) = \mathrm{dom}\,U(t)^{-1}AU(t) = U(t)^{-1}\left(A^{-1}(\mathrm{dom}\,U(t)^{-1})\right) = U(t)^{-1}\,(\mathrm{dom}\,A)$. Then $U(t)v \in \mathrm{dom}\,A$ so that the Spectral theorem applied to A shows that:

$$\langle A(t)v, v \rangle = \langle U(t)^{-1}AU(t)v, v \rangle \;=\; \langle AU(t)v, U(t)v \rangle$$

$$= \int_{-\infty}^{\infty} \lambda\, d\langle \mathrm{P}_A(\lambda)U(t)v, U(t)v \rangle$$

$$= \int_{-\infty}^{\infty} \lambda\, d\langle U(t)^{-1}\mathrm{P}_A(\lambda)U(t)v, v \rangle$$

Because $A(t)$ is also easily seen to be self-adjoint and $\mathrm{P}_{A(t)}$ is the unique projection-valued measure such that $\langle A(t)v, v \rangle = \int_{-\infty}^{\infty} \lambda\, d\langle \mathrm{P}_{A(t)}(\lambda)v, v \rangle$, we must have that $\mathrm{P}_{A(t)} = U(t)^{-1}\mathrm{P}_A U(t)$ as claimed. From this observation, it easily follows that:

$$\mu_{A(t),M}(E) = \mathrm{tr}\,\mathrm{P}_{A(t)}(E)M \;=\; \mathrm{tr}\,U(t)^{-1}\mathrm{P}_A(E)U(t)M$$

$$= \mathrm{tr}\,\mathrm{P}_A(E)U(t)MU(t)^{-1}$$

$$= \operatorname{tr} \mathrm{P}_A(E) M(t)$$

$$= \mu_{A,M(t)}(E)$$

Thus, we have $\mu_{A,M(t)} = \mu_{A(t),M}$ since both measures agree on every Borel set.

Because the two measures are equal, it makes no difference from the practical perspective of computing probabilities and expected values whether we adopt Schrödinger's or Heisenberg's picture of quantum mechanics. We need only append one of the axioms to our former list to capture the dynamics of quantum theory.

REFERENCES

[1] Baez, John C., "Quantum Theory and Analysis."
From: http://math.ucr.edu/home/baez/, April 15, 2009.

[2] Conway, John B., *A Course in Functional Analysis*, 2nd ed., Spring-Verlag New York, Inc., 1990.

[3] Dugundji, James, *Topology*, Allyn and Bacon, Inc., 1966.

[4] Halmos, Paul R., *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, 2nd ed., McGraw-Hill, Inc., 1973.

[5] Kreyszig, Erwin, *Introductory Functional Analysis with Applications*, John Wiley & Sons, Inc., 1978.

[6] Royden, H.L., *Real Analysis*, 3rd ed., Macmillan Company, 1963.

[7] Rudin, Walter, *Functional Analysis*, McGraw-Hill, Inc., 1973.

[8] Simon, Barry and Reed, Michael, *Methods of Modern Mathematical Physics: Functional Analysis*, Academic Press, Inc., 1972.

[9] Takhtajan, Leon A., *Quantum Mechanics for Mathematicians*, American Mathematical Society, 2008.

[10] von Neumann, John, *Mathematical Foundations of Quantum Mechanics*, transl. by Beyer, Robert T., 2nd ed., Princeton University Press, 1955.

[11] Wightman, A.S., "Hilbert's Sixth Problem: Mathematical Treatment of the Axioms of Physics", *Proceedings of Symposia in Pure Mathematics*, vol. 28, 1976.