L8 – FORMALISM OF QUANTUM MECHANICS

UCSC, PHYS 139A, SPRING 2008
GEY-HONG SAM GWEON

`This note is quite lengthy. I suggest a quick browsing, a topical reading, and then a full reading, if you still have time. For the topical reading, the following information may be helpful. Topics whose titles are underlined and in bold face are essential to understand thoroughly. They are the fruits of this note. Topics whose titles are in bold face but not underlined are those that are less important to study either because they are not really new ("we know them already") or because they are kind of common-sensical. The rest of the topics and text contain basic physics and basic facts of Linear Algebra. They may help you acquire more grounded understanding of what we are doing, but if you have no time to read them, it would not hurt you for your progress in this course.`

1. We are Waves, Maybe

Physics in a "vector space" is a familiar one. We are here or there, and move, and it seems natural to us that we can put an arrow on everything that is somewhere and moves, just like that air flow pattern in the weather map that that weather woman is reciting about in the evening news. This is the view of non-relativistic[1] Classical Mechanics (CM). In the Lagrangian formulation of CM, two vectors – position vector $\vec{x}$ and velocity vector $\vec{v} \equiv \dot{\vec{x}}$ – form complete variables for a motion of a particle. In the Hamiltonian formulation, the complete variables are $\vec{x}$ and the momentum vector $\vec{p}$. In any case, vectors are like arrows. For an arrow in our ordinary world, three numbers suffice to describe it, and so we call it a three dimensional arrow. Since we have two vectors, why don't we just combine the two sets of three numbers and call it a new arrow, characterized by six numbers? We do that and call these new arrows "vectors in the phase space," which is now six dimensional. Indeed, in CM, our world is mapped to the six dimensional phase space in which each point, in other words each arrow connecting from the origin to the point, completely describes the motion of a particle.

Now, when we look at classical objects through powerful microscopes with incredible resolving powers to discern individual atoms, we discover a totally different world. We realize that things that we called "particles" have all these wrinkles, imperfections, jittery motions and interferences. It is a fascinating world, and it took the work of many scientific heroes in the last hundred years to figure out that, fundamentally speaking, *our world is more like the vast ocean surface bustling and vibrating with all kinds of waves rather than the empty space where "point particles" zoom around.* However, at this point, this is just a mathematical picture, rather than a physical picture (so "maybe").

What we call a "point particle" in CM is a simply wave packet in QM, with a special condition[2]: the energy and the momentum scales associated with the wave packet must be comparable to the everyday scale. If this condition is satisfied, QM[3] jitters inside the wave packet, while existent all the time, are of little practical importance, as their magnitudes are negligible compared to everyday scales. In QM terms, only the average position $\langle\hat{\vec{x}}\rangle$ and the average momentum $\langle\hat{\vec{p}}\rangle$ matter, in which case the CM *emerges* from the QM, and *gives* the Newton's law, a la the Ehrenfest theorem.

[1] Since we are not covering the relativistic regime in this course, "non-relativistic" is an assumed adjective and will be omitted whenever we mention Classical Mechanics (CM) or Quantum Mechanics (QM) in this note.

[2] There is also an *extremely* important aspect to be mentioned here. If you have only one particle, you cannot simply crank up its energy/momentum scale to the classical scale. It is just not possible to do so without going to the relativistic regime and beyond (!). Instead, note that a "particle" in CM is never a single particle, but consists of many ($\sim 10^{23}$) particles. Indeed, what is essential for the emergence of CM from QM is actually the existence of *many* particles, so that the total energy/momentum scale can approach the classical scale.

[3] Acronym QM will be used for either "Quantum Mechanics" or "Quantum Mechanical" (as in this example) throughout this note. Similarly for CM.

What is quite amusing, then, is that the "vast ocean surface vibrating with waves" world of QM from which the "empty space filled with point particles" world of CM emerges is *also* a kind of space consisting of "arrows" as the CM world is a phase space. The space consisting of arrows is, in a mathematical term, a "vector space." But, beware, the QM vector space is totally different from, and way more complicated than, the CM vector space. It is not six dimensional[4]. It is not even finite dimensional. Its dimension is *always infinite*[5]. Why is this? It is basically because, in the QM world, what corresponds to the "arrow" is the "wave" itself, and because each wave must be described by infinite numbers.

2. Dirac shows the Hilbert Space

OK, it is not only Dirac but also von Neumann, but the Dirac notation is the *de facto* notation for QM, and Dirac's formulation of QM is much more widely read by scientists than that of von Neumann, although the two are essentially the same.

The kind of mathematical space that fits QM so nicely is the vector space built on complex numbers with properly defined inner product – this type of space is called a Hilbert space, of which the exact definition will be given below. The mathematics of Hilbert space has many applications other than QM, e.g. field theories in CM, diffusion equations and engineering problems. On the other hand, it seems fair to say that no subject imbues the mathematics of Hilbert space with more beauty, depth and mystery than Quantum Mechanics does.

The central object of a Hilbert space is vector.

Definition 1. ***Dirac Notation – Vector***

In the "*Dirac notation*," a vector is denoted as $|\alpha\rangle$. Here α stands for an arbitrary label. It is often a quantum number or a group of quantum numbers. Sometimes it is a generic label as in $|\Psi\rangle$.

Note. Usual and unusual Linear Algebra

The mathematics of our concern here is the Linear Algebra. The *usual* Linear Algebra of finite dimensional vectors is easy to handle and clear-cut. In Quantum Mechanics, though, we need more for our analytic work. We need an *unusual* Linear Algebra, extended to handle infinite dimensional vectors whose basis vectors are indexed by either integers or continuous real numbers. The extension to the unusual Linear Algebra is not a trivial process, but, fortunately, the results are very analogous to those results of the usual Linear Algebra, if one keeps in mind some key substitutions to be made ($\sum_n \to \int dr$ and $\delta_{nn'} \to \delta(r - r')$).

Definition 2. ***Vector Space***

We start with numbers, say a set of numbers $\mathcal{C}$. In mathematics, $\mathcal{C}$ can be any "field" of numbers but, in QM, $\mathcal{C}$ should be thought of as the set of complex numbers, which will be our assumption throughout this note. Numbers are called "scalars" with respect to those higher order objects that we build based on them, such as vectors and matrices.

A vector space is defined in the following table by the rules that need to be satisfied by vectors. As long as these rules are satisfied, vectors can be *any* objects – *arrows, functions, waves, wave functions, computer programs, growth patterns of CO on Pt surface in the tail pipe of your car,* etc. etc. In QM, the best view is to regard vector as state. More about this below.

Note. The central notion of a vector space $\mathcal{V}$ is the following: a vector space is closed under the *linear combination*, i.e. $\forall|\alpha\rangle, |\beta\rangle \in \mathcal{V}$ and $\forall c, d \in C$, the linear combination $c|\alpha\rangle + d|\beta\rangle$ should *always* be a valid element of $\mathcal{V}$.

---

[4]In QM the position and the momentum are not independent due to their non-zero commutator. This means that momentum is not an independent variable any more. The independent variable is either position or momentum, but it cannot be both. This is the consequence of "matter as wave" and is aptly summarized by the Heisenberg Uncertainty Principle.

[5]If you consider a particular "sub-space," such as angular momentum space or spin space, the dimension can be finite. However, waves in QM *always* need to exist in the position space, in addition to these *other* spaces, and so the *total* dimension of the QM vector space is always infinite.

#	For *any*	It is required that	Meaning
1	$\|\alpha\rangle, \|\beta\rangle \in \mathcal{V}$	$\|\alpha\rangle + \|\beta\rangle \in \mathcal{V}$	vectors *can* be added
2	$\|\alpha\rangle, \|\beta\rangle \in \mathcal{V}$	$\|\alpha\rangle + \|\beta\rangle = \|\beta\rangle + \|\alpha\rangle$	vector addition is commutative
3	$\|\alpha\rangle, \|\beta\rangle, \|\gamma\rangle \in \mathcal{V}$	$\|\alpha\rangle + (\|\beta\rangle + \|\gamma\rangle) = (\|\alpha\rangle + \|\beta\rangle) + \|\gamma\rangle \equiv \|\alpha\rangle + \|\beta\rangle + \|\gamma\rangle$	vector addition is associative
4	$\|\alpha\rangle \in \mathcal{V}$	$\exists \|null\rangle \, s.t. \, \|\alpha\rangle + \|null\rangle = \|\alpha\rangle$	the null, or zero, vector exists
5	$\|\alpha\rangle \in \mathcal{V}$	$\exists \|-\alpha\rangle \, s.t. \, \|\alpha\rangle + \|-\alpha\rangle = \|null\rangle$	inverse vector exists
6	$\|\alpha\rangle \in \mathcal{V},\, c \in \mathcal{C}$	$c\|\alpha\rangle \in \mathcal{V}$	vectors *can* be scaled
7	$\|\alpha\rangle, \|\beta\rangle \in \mathcal{V},\, c \in \mathcal{C}$	$c(\|\alpha\rangle + \|\beta\rangle) = c\|\alpha\rangle + c\|\beta\rangle$	scaling is a linear operation
8	$\|\alpha\rangle \in \mathcal{V},\, c, d \in \mathcal{C}$	$(c+d)\|\alpha\rangle = c\|\alpha\rangle + d\|\alpha\rangle$	scaling is distributive
9	$\|\alpha\rangle \in \mathcal{V},\, c, d \in \mathcal{C}$	$(cd)\|\alpha\rangle = c(d\|\alpha\rangle) \equiv cd\alpha\rangle$	scaling is associative
10	$\|\alpha\rangle \in \mathcal{V}$	$0\|\alpha\rangle = \|null\rangle$	0 scaling gives the null vector
11	$\|\alpha\rangle \in \mathcal{V}$	$1\|\alpha\rangle = \|\alpha\rangle$	scaling by 1 does nothing
12	$\|\alpha\rangle \in \mathcal{V}$	$-1\|\alpha\rangle = \|-\alpha\rangle \equiv -\|\alpha\rangle$	scaling by -1 gives inverse vector

TABLE 1. Mathematical Definition of a Vector Space $\{|\alpha\rangle\}$

The practical notation for $|null\rangle$ is 0, for which rule 10 above gives the justification. Note that $|null\rangle$ basically means that all "components" of a vector are zero[6], so it is not really equal to 0. They belong in difference spaces. In practice, though, 0 is *all* you will see in the physics literature when the null vector $|null\rangle$ is really what the authors really mean. For instance, in the Harmonic oscillator problem, $|n\rangle$ is the standard notation for the stationary state $\psi_n(x)$. The equation that we derived in chapter 2, $\hat{a}_-\psi_n(x) = \sqrt{n}\psi_{n-1}(x)$, then reads $\hat{a}_-|n\rangle = \sqrt{n}|n-1\rangle$. For $n = 0$, one writes $\hat{a}_-|0\rangle = 0$. Notice that the number 0 on the RHS of this equation should be understood as $|null\rangle$, strictly speaking, but people never bother to write $|null\rangle$ explicitly, and maybe you can understand why.

Here, different from the textbook, I am staying away from using the symbol $|0\rangle$ to mean $|null\rangle = 0$. My practice is consistent with the notations widely accepted by physicists, in which $|0\rangle$ is actually never 0. In the usual physicist's notation, $|0\rangle$ is the ground state or the vacuum, which has rich structures and wave functions that *continue to* vibrate even in vacuum. Put another way, $|null\rangle$ represents the mathematical concept of zero, which is truly nothing, while $|0\rangle$ is represents the physical vacuum (if you like, you can call this the physical zero), which is not nothing at all, and is actually, you might say, a complete opposite of nothing.

The above table is certainly boring and seemingly trivial. As students of math, however, we all know that from these trivial-looking axioms a whole wonderful field of linear algebra quickly springs up and moreover that these axioms are powerful abstractions of diverse real world problems. In short, treat them very preciously!

Definition 3. ***Inner Product Space***

An inner product space is a vector space with an additional feature, which is, well, the inner product of two vectors. Any mapping of two vectors $|\alpha\rangle, |\beta\rangle$ to a number qualifies as an inner product as long as it meets canonical requirements, listed in the table below. The central notion of an inner product is that it provides a way to measure angles between vectors as well as the length of a vector.

Definition 4. ***Dirac Notation – Bra(c)ket***

The cool Dirac notation for an inner product of two vectors $|\alpha\rangle, |\beta\rangle$ is $\langle\alpha|\beta\rangle$. You could write the inner product as $dot(|\alpha\rangle, |\beta\rangle)$ or $|\alpha\rangle \cdot |\beta\rangle$ or in any similar way conceivable but once you get used to the clean and efficient Dirac notation these other notations seem hardly necessary. Note that in

[6] For a vector space of functions, this is not necessarily true. A function that is zero at all points except a few points at which the function has finite values is still to be regarded as $|null\rangle$, not only a function that is zero at all points. When we talk about functions as vectors, we are actually talking about classes of functions. However, we do not need to worry about this, since such discontinuous functions do not occur in physics.

the Dirac notation, the inner product looks like a product of two objects, one $\langle\alpha|$ and the other $|\beta\rangle$. What does $\langle\alpha|$ mean? We will discuss its meaning below. Here, let us just say that it is a mirror image of $|\alpha\rangle$. Dirac called a vector like $|\alpha\rangle$, a "ket" vector, and its mirror image, $\langle\alpha|$, a "bra" vector, so that the inner product is then a "bra(c)ket." Note that the order matters in a bra-ket, namely $\langle\alpha|\beta\rangle$ and $\langle\beta|\alpha\rangle$ are different, according to rule 1 below.

#	Rule	Meaning
1	$\langle\beta\|\alpha\rangle = \langle\alpha\|\beta\rangle^*$	inner product is not commutative if numbers are complex (then complex conjugate)
2	$\langle\alpha\|\alpha\rangle \geq 0$ $\langle\alpha\|\alpha\rangle = 0$ holds only if $\|\alpha\rangle = \|nil\rangle$	positive definite self-inner-product defines a length^2 or norm^2
3	$\langle\alpha\| (c\|\beta\rangle + d\|\gamma\rangle) = c\langle\alpha\|\beta\rangle + d\langle\alpha\|\gamma\rangle$ $[\Rightarrow (c\langle\alpha\| + d\langle\beta\|) \|\gamma\rangle = c\langle\alpha\|\gamma\rangle + d\langle\beta\|\gamma\rangle]$	inner product is a linear operation for *either* argument [not just for the second argument, due to rule 1]

TABLE 2. Mathematical Definition of an Inner Product $\langle\alpha|\beta\rangle$

We will need to consider certain vectors for which $\langle\alpha|\alpha\rangle$ is infinity. Even for those vectors we will still make use of the above rules. Mathematically, one cannot say that such a vector belongs in an inner product space, and so using the rules above may be an absurdity. The fundamental reasons for physicists' "abusing" mathematics in this way are (i) Nature requires us to do so, as far as we can tell, and (ii) physicists really do *not* mean infinity by infinity[7]. Important classes of QM wave functions, such as plane waves, are not normalizable, but it is essential to include them in our vector space.

Theorem 5. ***Schwartz Inequality***

In the ordinary Euclidean space based on real numbers, we are familiar with the fact that $\vec{a}\cdot\vec{b} = ab\cos\theta$, where θ is the angle between $\vec{a}$ and $\vec{b}$. Thus the following inequality holds: $|\vec{a}\cdot\vec{b}|^2 = (\vec{a}\cdot\vec{a})\left(\vec{b}\cdot\vec{b}\right)\cos^2\theta \leq (\vec{a}\cdot\vec{a})\left(\vec{b}\cdot\vec{b}\right)$. The inner product defined above is a generalization of this dot product, and continues to satisfy the following inequality:

$$|\langle\alpha|\beta\rangle|^2 \leq \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle$$

Proof. See Problem A.5 of Griffiths. Note that the inequality is a direct consequence of the above definitions listed in the table and no other assumptions are necessary (e.g. *specific* properties of column vectors are not necessary). □

When vectors are not normalizable, that is $\langle\alpha|\alpha\rangle = \infty$, the above inequality may not have any meaning. This is not a major concern for us, as all important properties that we will derive using the Schwartz inequality are for normalizable vectors.

Definition 6. *Orthogonal Vectors*

Two vectors $|\alpha\rangle, |\beta\rangle$ are said to be *orthogonal*, if $\langle\alpha|\beta\rangle = 0$.

Definition 7. ***Dual and its Dual***

If a set of vectors $\{|\alpha\rangle\}$ form an inner product space, then the set $\{\langle\alpha|\}$ where $\langle\alpha|$ is a "mapping" $L_\alpha : \mathcal{V} \to \mathcal{C}$ defined by $L_\alpha(|\beta\rangle) = \langle\alpha|\beta\rangle$ also forms an inner product space, which we call the "dual space" of the original inner product space (proof left to readers). But then, you might say, why not view the set of bras as vectors, and then view the set of kets as a set of linear mappings on

[7]At this point, let me note the physical meaning of an expression like $\int_{-\infty}^{\infty} dx \exp(ikx) = 2\pi\delta(k)$. What happens at $k = 0$? It is infinity! What physicists really mean here is that the integral *scales like* the system size L. Ideally, L can go to infinity, but in reality it cannot due to the emergence of other physics. What we are dealing with for any given problem is implicit scales, lower cutoff and upper cutoff. In this view, you might be at peace with the delta functions and infinities and so on, since they never really represent mathematical infinities, but simply a mathematical idealization of infinities.

bras. This is fine, too! Of course, what we are doing here is a trivial (but useful) thing, which is to look at the inner product $\langle\alpha|\beta\rangle$ and interpret it as

$$\langle\alpha|\beta\rangle = L_\alpha(|\beta\rangle) = L_\beta(\langle\alpha|)$$

By virtue of rule #3 of inner product, both L_α and L_β are *linear* mappings. In the language of Linear Algebra, they are *linear transformations.* Having introduced these interpretations, from now on we will simply look at the inner product and be flexible enough to interpret it as any one of these expressions, which are to be understood as product of bra and ket, linear transformation on ket, linear transformation on bra, respectively):

$$\langle\alpha|\beta\rangle = \langle\alpha|(|\beta\rangle) = (\langle\alpha|)|\beta\rangle$$

Notice that the last expression shows the convention of the Dirac notation: *a linear transformation on bras acts from the right side.* This is true in general for any linear transformation on bras.

Definition 8. ***Hilbert Space***

A Hilbert space $\mathcal{W}$[8] is an inner product space *without holes*[9], in the sense that if there is a convergent sequence of vectors $|\alpha_n\rangle$, then $\lim_{n\to\infty}|\alpha_n\rangle$ is also within the inner product space. By definition, we will also require that a Hilbert space is a proper inner product space in the sense that $\langle\alpha|\alpha\rangle < \infty$ for any $|\alpha\rangle$. The central notion of the Hilbert space that it rules out a space like "an inner product space based on purely rational numbers." Such space *has holes.* Such space would be actually OK for numerical work, but certainly is not OK for analytical work for which the standard calculus is one of the most basic mathematical tools.

Note. By virtue of the property of the Hilbert space, we will treat any limiting procedures as well defined, e.g. $\int dx[something] = \lim_{\Delta x\to 0}\sum \Delta x[something]$. Furthermore, we will assume that all mathematical conditions are met for swapping limiting procedures and other procedures, e.g. $\langle\alpha|\int dx|\beta_x\rangle = \int dx\langle\alpha|\beta_x\rangle$.

Definition 9. ***Rigged Hilbert (or "e-Hilbert") space (Advanced Topic)***

An "e-Hilbert" space (equipped Hilbert space or rigged Hilbert space – the latter seems to be the more "standard" term for this rather modern concept, but I don't like it), as an Hilbert space with some "suburban" area in which functions like $\exp(ikx)$ or $\delta(x - x_0)$ reside. Why do we need it? Well, since we have to deal with vectors with diverging norms! The mathematical definition of a rigged Hilbert space involves considering a subset Φ of the Hilbert space $\mathcal{W}$ in which all vectors are not only normalizable but "nice" (e.g. analytic), and then consider a superset Φ^* of $\mathcal{W}$, Φ^* is a set of functionals on Φ. An element of Φ^* is called a generalized function, a distribution, or a test function. For instance δ function belongs there. In this mathematical sense, δ function only exists when it is used with a nice function $f(x)$ inside the integral as in $\int dx f(x)\delta(x)$. These three *vector spaces* satisfy the relationship $\Phi \subset \mathcal{W} \subset \Phi^*$. In this modern view, Dirac bras belong to Φ^* (and kets belong to another such superset Φ') and all physical states belong to Φ.

This definition of e-Hilbert space does help to clear things up a little bit, as compared to formulations of Quantum Mechanics made in Dirac's time and Neumann's time. Nevertheless, one is left with the usual feeling that one gets almost whenever one tries to find a watertight mathematical framework to contain physics – it is *not* perfect. You see, this *vector space* Φ does not allow you to take inner products, when they diverge, but we need to handle inner products, even diverging ones[10], on a constant basis!

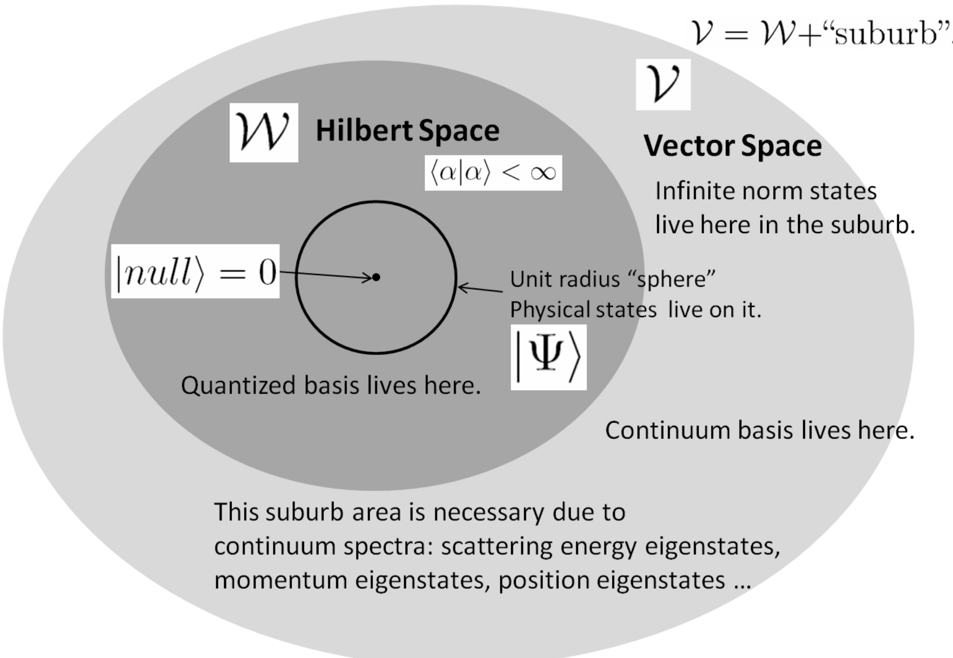
Note. ***Terminology***

[8]The symbol $\mathcal{W}$ will be used to denote a Hilbert space. Usually the notation is $\mathcal{H}$, but I do not like it since it may be confused for the Hamiltonian. I am using $\mathcal{W}$ to mean a space by which our QM "world" is described.

[9]The correct *mathematical expression* here is a "complete" inner product space. But I am staying away from using the word "complete" here, since it has a much much more important other usage in Quantum Mechanics.

[10]However, see footnote 7.

Apart from the subtle mathematical issue of "Hilbert space" and "e-Hilbert or rigged Hilbert space," perhaps what is the most important is the terminology. "Hilbert space" always means $\langle\alpha|\alpha\rangle < \infty$. The definition of the "e-Hilbert space" is given above, and you are welcome to do research on it further if you feel curious about it. We will not mention "e-Hilbert space" or "rigged Hilbert space" again in this document. For our purpose here, let us just note that there is this vector space $\mathcal{V} = \Phi^*$, which always contains $\mathcal{W}$. Keep in mind that in the end the Hilbert space is the only thing we should care about. For instance, if some property holds for any vectors in the Hilbert space but breaks down outside it, this is generally not a problem, since our world is only within the Hilbert space. Also, note that the need to invoke $\mathcal{V}$ which is larger than $\mathcal{W}$ arises only when a continuous spectrum is involved. $\mathcal{V} = \mathcal{W}+$"suburb". A *schematic* representation is given in the diagram below.



> **Axiom 10.** ***I. Physical States in Quantum Mechanics***
> *A physical state corresponds to a vector of length 1 in the Hilbert space $\mathcal{W}$. By necessity, we may consider a larger vector space $\mathcal{V}$, which encloses $\mathcal{W}$ and may now contain non-normalizable vectors. All vectors in $\mathcal{V}$ are referred to simply as states.*

In keeping with customs, we will tend to use a symbol like $|\Psi\rangle$ for a generic *physical* state. Geometrically, physical states can be considered as forming the surface of a unit radius "sphere" in $\mathcal{V}$. Up to the normalization factor, any state in $\mathcal{W}$, the Hilbert space, can be sort of *thought of* as a physical state, with the single exception of $|null\rangle = 0$, which is not physical by any stretch of imagination.

Note. ***State and Wave Function – Are They the Same?***

For the most part, we will regard them as the same. But beware, the right answer is actually No. They are subtly different, just as vector as arrow is different from vector as column vector, which we may use to *represent* the arrow. We will spend some time discussing the difference between the state and its representation (wave function) in Section 9. Until we clarify their differences, though, it makes a much more pedagogical sense to consider them as one and the same, and so we will do just that.

Note. ***No x in $|\Psi\rangle$***

In the notation for a state, such as $|\Psi\rangle$, we never write down variable of the function like x. It make no more sense to write it than to write the integer index i or n for ordinary vectors. These are dummy indices that we need when we need to express the vector component, and is completely replaceable by any other symbol and so cannot be included as a symbol of the vector. Of course, a more fundamental reason is that state and wave function are different. The time t is a different matter. Often an expression such as $|\Psi(t)\rangle$ will be used to explicitly note that time t is an external parameter, external here meaning outside the vector space.

Example 11. ***Hilbert Space of Functions***

Let $\mathcal{V}$ be the set of *any* [ordinary or generalized] functions of x. The following standard definitions apply:

$$\begin{aligned} |f\rangle &\equiv f(x) \\ \langle f| &\equiv f(x)^* \\ \langle f|g\rangle &\equiv \int dx f(x)^* g(x) \end{aligned}$$

These definitions are completely analogous to those of the usual Linear Algebra, in which $|\alpha\rangle$ means a column vector, $\langle\alpha|$ means a row vector, and the summation is implied only when the inner product is invoked. Namely, a function can be thought of as a list of numbers, albeit indexed by real numbers, and $\int dx$ plays the role of $\sum_n$, as the integration is the only way, up to a multiplicative constant, to define a well defined "sum" of a function. This approach is definitely simpler than an alternative approach to think of $\langle f|$ as meaning the operation $\int dx f(x)^*[...]$ acting on whatever function [...] that appears on the right hand side of it (like in Griffiths), which is OK but not particularly desirable since it tends to be cumbersome.

Let $\psi_n(x)$ be the stationary state solutions of the Harmonic oscillator problem. Due to the Fourier-Hermite theorem, it can be shown that any piecewise continuous function can be written as $\sum_{n=0}^{\infty} c_n \psi_n(x)$. This is not enough to define the Hilbert space. For instance, the exponential function $|\alpha\rangle = \exp(ikx)$ is certainly piecewise continuous, but its norm $\sqrt{\langle\alpha|\alpha\rangle}$ is ∞. Only if one imposes the condition that $\sum_n |c_n|^2 < \infty$, then the Hilbert space $\mathcal{W}$ is obtained. The physical states are the subset of $\mathcal{W}$, for any member of which $\sum_n |c_n|^2 = 1$ is satisfied.

3. Hats to Operators

Before we define operators, we need to discuss linear transformations, which are more general than operators. In the usual Linear Algebra lingo, linear transformations are matrices, and operators are square matrices. As such, we do not give a hat to a linear transformation just yet.

Definition 12. ***Linear Transformation***

A linear transformation L is a mapping from a vector space $\mathcal{V}$, *domain vector space*, to another vector space $\mathcal{V}'$, *image vector space*, with the following requirement: $L\left(c|\alpha\rangle + d|\beta\rangle\right) = cL\left(|\alpha\rangle\right) + dL\left(|\beta\rangle\right)$. As discussed in Definition 7, a bra vector is a linear transformation with $\mathcal{V}' = \mathcal{C}$, and a ket vector is a linear transformation acting on the bra space to $\mathcal{V}' = \mathcal{C}$. More generally, the following notations are used to denote a linear transformation acting on the ket space or a linear transformation from the bra space.

$$\begin{aligned} L|\alpha\rangle &\equiv L\left(|\alpha\rangle\right) \\ \langle\alpha|L &\equiv L\left(\langle\alpha|\right) \end{aligned}$$

Again, it is worth noting that *a linear transformation on bras acts from the right side* in the Dirac notation.

Fact 13. *Linear Transformations Form a Vector Space*

It is straightforward to show that a set of linear transformations is a vector space, assuming that the addition and the scaling of a linear transformation is defined as one may naturally expect: $\left(L_1 + L_2\right)|\alpha\rangle \equiv L_1|\alpha\rangle + L_2|\alpha\rangle$ and $\left(cL\right)|\alpha\rangle \equiv c\left(L|\alpha\rangle\right)$.

Definition 14. *Identical Linear Transformations*

- If $L_1|\alpha\rangle = L_2|\alpha\rangle$ for any $|\alpha\rangle$, then $L_1 = L_2$.
- If $\langle\alpha|L_1 = \langle\alpha|L_2$ for any $\langle\alpha|$, then $L_1 = L_2$.

Definition 15. *Product of Linear Transformations*
The product of two linear transformations is defined as follows:

$$(L_1 L_2)\,|\alpha\rangle \equiv L_1\,(L_2|\alpha\rangle)$$

This product is inherently associative, and leads to the following fact.

Fact 16. *Product of Linear Transformations is Associative*
The product $L_1L_2...L_n$ is completely associative, that is it can be grouped in neighbors in any way conceivable. For example, $(L_1L_2)L_3 = L_1(L_2L_3)$ and $(L_1L_2)\,(L_3L_4) = (L_1\,(L_2L_3))\,L_4 = L_1\,(L_2\,(L_3L_4)) = ... = L_1L_2L_3L_4$.

Proof. Directly follows from Definition 15. □

Note. Valid Product of Linear Transformations
You might have noticed by now that any expression that we wrote so far using combinations of bras, kets, and linear transformations is actually a product of linear transformations or a sum of products of linear transformations. In general, not all products of the form L_1L_2 are valid. For instance, if L_1 is a 3×3 matrix and L_2 is a 2×2 matrix, then L_1L_2 is not allowed. For this reason, let me make it clear that all products appearing in this note are *required* to be valid. What I mean here is that when product expressions such as L_1L_2 are written down, we are automatically *requiring* that the domain space of L_1 is the same as the image space of L_2, even if we do not state that requirement explicitly.

Definition 17. ***Adjoint, or Hermitian Conjugate, of a Linear Transformation***
The adjoint or Hermitian conjugate $L^\dagger$ of a linear transformation L is defined as follows.

(1) If L is a linear transformation acting on kets, then $L^\dagger$ is defined as $(L|\alpha\rangle)^\dagger = \langle\alpha|L^\dagger$.
(2) If L is a linear transformation acting on bras, then $L^\dagger$ is defined as $(\langle\alpha|L)^\dagger = L^\dagger|\alpha\rangle$.

This definition summarizes the following meaning, which is familiar from the usual Linear Algebra but now extended to the unusual Linear Algebra. (i) The adjoint for linear transformations is what the complex conjugate is for numbers. (ii) The bra $\langle\alpha|$ is the adjoint of the ket $|\alpha\rangle$. (iii) For more general linear transformations ("matrices"), the adjoint means transpose followed by complex conjugation.

Fact 18. ***Bra and Ket as Hermitian Conjugate Linear Transformations***

$$|\alpha\rangle^\dagger = \langle\alpha|,\ \langle\alpha|^\dagger = |\alpha\rangle$$

Proof. Put $L = 1$ in the above definition. See Fact 7 to see why bras and kets are linear transformations. □

Fact 19. *Hermitian Conjugate for Number is Complex Conjugate*
If $L = c$, i.e. a scaling linear transformation, then $L^\dagger = c^*$.

Proof. Put $|\beta\rangle = L|\alpha\rangle = c|\alpha\rangle$. Then, by the above definition, $\langle\beta| = \langle\alpha|c^\dagger$. Consider $\langle\beta|\gamma\rangle = c^\dagger\langle\alpha|\gamma\rangle$. On the other hand, this is equal to $\langle\gamma|\beta\rangle^* = c^*\langle\gamma|\alpha\rangle^* = c^*\langle\alpha|\gamma\rangle$. Thus, $c^\dagger = c^*$. □

Fact 20. *Hermitian Conjugate of Hermitian Conjugate is the Original*

$$\left(L^\dagger\right)^\dagger = L$$

Proof. If L acts on kets, then proceed as follows. Taking $\dagger$ on both sides of $(L|\alpha\rangle)^\dagger = \langle\alpha|L^\dagger$, we have $\left((L|\alpha\rangle)^\dagger\right)^\dagger = \left(L^\dagger\right)^\dagger|\alpha\rangle$. Note that by definition $L|\alpha\rangle$ is a ket. Since a bra and a ket are Hermitian conjugates of each other, we see that $\left((L|\alpha\rangle)^\dagger\right)^\dagger = L|\alpha\rangle$. Thus, $L|\alpha\rangle = \left(L^\dagger\right)^\dagger|\alpha\rangle$ for any $|\alpha\rangle$, which proves our claim. If L acts on bras, the proof is entirely symmetrical to the one given here. □

Fact 21. ***"Reverse Chain Rule" for Hermitian Conjugation***
For linear transformations $L_1, L_2, ..., L_n$

$$(L_1 L_2)^\dagger = L_2^\dagger L_1^\dagger$$

and in general

$$(L_1 L_2 ... L_n)^\dagger = L_n^\dagger ... L_2^\dagger L_1^\dagger$$

Proof. Consider $(L_1 L_2|\alpha\rangle)^\dagger$ where $|\alpha\rangle$ is a ket that L_2 takes. Let $|\beta\rangle \equiv L_2|\alpha\rangle$ be the ket that L_1 takes. $(L_1|\beta\rangle)^\dagger = \langle\beta|L_1^\dagger$ from Definition 17. The same definition means that $\langle\beta| = (L_2|\alpha\rangle)^\dagger = \langle\alpha|L_2^\dagger$. And so we have $\langle\alpha|\,(L_1L_2)^\dagger \equiv (L_1L_2|\alpha\rangle)^\dagger = (L_1|\beta\rangle)^\dagger = \langle\beta|L_1^\dagger = \langle\alpha|L_2^\dagger L_1^\dagger$ for any $\langle\alpha|$. By Definition 14, we then have $(L_1L_2)^\dagger = L_2^\dagger L_1^\dagger$. By applying this rule repeatedly, we obtain $(L_1L_2...L_n)^\dagger = L_n^\dagger...L_2^\dagger L_1^\dagger$. □

Fact 22. ***Hermitian Conjugate $\dagger$ as $t*$***
Suppose that a linear operator L, assume that $\langle\alpha|L|\beta\rangle$ is a number. Then,

$$\langle\alpha|L|\beta\rangle^* = \langle\beta|L^\dagger|\alpha\rangle$$

This fact is consistent with the *definition* of the Hermitian conjugate of a matrix M, $M^\dagger \equiv M^{t*}$ in the usual Linear Algebra[11] .

Proof. This is the direct consequence of the reverse chain rule (Fact 21), with $L_1 = \langle\alpha|$, $L_2 = L$, $L_3 = |\beta\rangle$, and the fact that the h.c. of a number is its complex conjugate (Fact 19). □

Example 23. Here are couple of examples of the Hermitian conjugate operation in the usual Linear Algebra.
$\begin{bmatrix} 1 \\ 2+i \end{bmatrix}^\dagger = \begin{bmatrix} 1 & 2-i \end{bmatrix}$ and $\begin{bmatrix} 1 & 4i \\ -3i & i \end{bmatrix}^\dagger = \begin{bmatrix} 1 & 3i \\ -4i & -i \end{bmatrix}$.

Definition 24. ***Quantum Mechanical Operator***

For a given Hilbert space $\mathcal{W}$, a linear transformation from $\mathcal{W}$ to $\mathcal{W}$ is defined[12] as a *Quantum Mechanical operator* $\hat{O}$, or simply an *operator*. In this sense, an operator may be viewed as a *"square matrix."* As such, *both* of the expressions $\hat{O}|\alpha\rangle$ and $\langle\alpha|\hat{O}$ are valid for any $|\alpha\rangle$ of $\mathcal{W}$. If necessary and appropriate, the space within which the operator is defined can be extended from $\mathcal{W}$ to a vector space $\mathcal{V} \supset \mathcal{W}$.

Definition 25. ***Operator in a Hat***

In this note, I use the term "operator" to mean "Quantum Mechanical operator" and absolutely nothing else. With this strict definition, it is redundant to say "linear operator." Notation-wise, operators will be distinguished by the hat $\hat{}$, like in $\hat{x}$ and $\hat{p}$, consistent with other lecture notes of this course. For the first formal course in Quantum Mechanics, I think this practice is strongly recommended. When you become an expert, that is, when you feel sufficiently comfortable with the formalism of Quantum Mechanics as it is laid out in full detail below, you might be tempted, understandably, to finally take hats away from operators, and indeed this is what every physicist

[11] In that type of definition, $|\alpha\rangle, |\beta\rangle$ must be considered as basis vectors. Our definition here is more general, so has no problem dealing with that case.

[12] Strictly speaking, this definition is not enough. The time reversal operator in QM is an *anti-linear* operator, $\hat{\mathbb{T}}\left(c|\alpha\rangle + d|\beta\rangle\right) = c^*\hat{\mathbb{T}}|\alpha\rangle + d^*\hat{\mathbb{T}}|\beta\rangle$, just like the complex conjugation operator. So, let us beware that it will become necessary, at some point of doing QM, to include anti-linear transformations as operators as well, but in this note we do not use them ever, so let us not worry about this fine point.

does more or less as long as there is no ambiguity for readers and writers alike. It would be fine if *you* practice like such an expert during this course, as long as you feel comfortable doing that and you show expertise in operator algebra [commutators, functions, different representations]. *I* will stick with those hats in my notes and lectures, so that there is no ambiguity between numbers and operators throughout.

Definition 26. ***Commutator***

Consider two operators, $\hat{O}_1$ and $\hat{O}_2$, and two products $\hat{O}_1\hat{O}_2$ and $\hat{O}_2\hat{O}_1$. These two are not the same. For instance, stretching along the $x$axis could be one operation and rotation of the xy plane by some angles could be the other operation. Clearly, the order in which these operations are performed matters in figuring out what the outcome is. So, the commutator

$$\left[\hat{O}_1, \hat{O}_2\right] = \hat{O}_1\hat{O}_2 - \hat{O}_2\hat{O}_1$$

is non-zero in general. Familiar examples in QM are $[\hat{x}, \hat{p}] = i\hbar$ and $[\hat{a}_-, \hat{a}_+] = 1$.

Theorem 27. ***Equality of Operators***

Two operators $\hat{O}_1$ and $\hat{O}_2$ are the same if $\langle\alpha|\hat{O}_1|\beta\rangle = \langle\alpha|\hat{O}_2|\beta\rangle$ for any $|\alpha\rangle, |\beta\rangle$ in $\mathcal{V}$. This condition can be relaxed further. Two operators are the same if

$$\langle\alpha|\hat{O}_1|\alpha\rangle = \langle\alpha|\hat{O}_2|\alpha\rangle$$

for any $|\alpha\rangle$ in $\mathcal{V}$.

Proof. First part first. $\langle\alpha|\hat{O}_1|\beta\rangle = \langle\alpha|\hat{O}_2|\beta\rangle$ for any $|\alpha\rangle, |\beta\rangle$ means $\hat{O}_1|\beta\rangle = \hat{O}_2|\beta\rangle$ for any $|\beta\rangle$, by Definition 14. Using the other part of the same definition, we conclude $\hat{O}_1 = \hat{O}_2$. Next the second part. For any given $|\alpha_1\rangle$ and $|\beta_1\rangle$, form $|\alpha\rangle = |\alpha_1\rangle + |\beta_1\rangle$. Applying this to $\langle\alpha|\hat{O}_1|\alpha\rangle = \langle\alpha|\hat{O}_2|\alpha\rangle$, we get $\langle\alpha_1|\hat{O}_1|\beta_1\rangle + \langle\beta_1|\hat{O}_1|\alpha_1\rangle = \langle\alpha_1|\hat{O}_2|\beta_1\rangle + \langle\beta_1|\hat{O}_2|\alpha_1\rangle$. Next form $|\alpha\rangle = |\alpha_1\rangle + i|\beta_1\rangle$ and repeat the procedure to obtain $\langle\alpha_1|\hat{O}_1|\beta_1\rangle - \langle\beta_1|\hat{O}_1|\alpha_1\rangle = \langle\alpha_1|\hat{O}_2|\beta_1\rangle - \langle\beta_1|\hat{O}_2|\alpha_1\rangle$. Adding these two equations, we obtain $\langle\alpha_1|\hat{O}_1|\beta_1\rangle = \langle\alpha_1|\hat{O}_2|\beta_1\rangle$ for *any* $|\alpha_1\rangle$ and $|\beta_1\rangle$. Thus, $\hat{O}_1 = \hat{O}_2$ by the first part. □

Note. ***The Hilbert Space is All We Care About.***

Establishing the above equality within the Hilbert space $\mathcal{W}$, or more precisely speaking the largest Hilbert space containing all physical states, is enough for all physics.

Definition 28. ***Expectation Value Notation***

For a given $|\alpha\rangle$, we define $\langle\hat{O}\rangle \equiv \langle\alpha|\hat{O}|\alpha\rangle$.

Axiom 29. ***II. Observables in Quantum Mechanics***
An observable dynamical quantity in Quantum Mechanics corresponds to a Hermitian operator:

$$\hat{A}^\dagger = \hat{A}$$

A Hermitian operator is also called a "self-adjoint" or "real" operator.

Note that a linear transformation can be Hermitian only if it is an operator ("square matrix"). For instance, an N-dimensional column vector ($N > 1$) cannot be Hermitian, as it is a different object dimensionally from an N-dimensional row vector.

Example 30. ***$\hat{p}$ is a Hermitian operator, indeed, and so is $\hat{x}$***

Let us check whether $\hat{p} = -i\hbar\frac{\partial}{\partial x}$ is a Hermitian operator. It better be. Here is how it goes. Take any two wave functions Ψ_1 and Ψ_2. Form $\langle\Psi_1|\hat{p}|\Psi_2\rangle$. In the function space, $\langle f|g\rangle = \int dx f^* g$, and so $\langle\Psi_1|\hat{p}|\Psi_2\rangle = \int dx\Psi_1^*\left(-i\hbar\frac{\partial}{\partial x}\right)\Psi_2$. Taking complex conjugate, we obtain

$$\begin{aligned}\langle\Psi_1|\hat{p}|\Psi_2\rangle^* &= \int dx\Psi_1\left(i\hbar\frac{\partial}{\partial x}\right)\Psi_2^* \\ &= \int dx\Psi_2^*\left(-i\hbar\frac{\partial}{\partial x}\right)\Psi_1 \\ &= \langle\Psi_2|\hat{p}|\Psi_1\rangle\end{aligned}$$

By definition, this is equal to $\langle\Psi_2|\hat{p}^\dagger|\Psi_1\rangle$, for *any* $|\Psi_1\rangle$ and $|\Psi_2\rangle$. And thus, $\hat{p}^\dagger = \hat{p}$, which means that $\hat{p}$ is a Hermitian operator, indeed. It is trivial to show that $\hat{x}$ is also Hermitian.

It is clear what $\hat{p}|\Psi\rangle$ is: it is $\left(-i\hbar\frac{\partial}{\partial x}\right)\Psi$. You may ask, however, what $\langle\Psi|\hat{p}$ is. The easy answer is that it can be considered as an *operator* $\Psi^*\left(-i\hbar\frac{\partial}{\partial x}\right)$. Alternatively, note that it should be possible to consider $\hat{p}$ as acting on $\langle\Psi|$, according to Definition 51. Noting that for any $|\Psi_1\rangle$, $\langle\Psi|\hat{p}|\Psi_1\rangle = \int dx i\hbar\frac{\partial\Psi^*}{\partial x}\Psi$ (by integration by parts), we can see that $\langle\Psi|\hat{p} = i\hbar\frac{\partial\Psi^*}{\partial x}$. This is also the same as $(\hat{p}|\Psi\rangle)^\dagger = \left(-i\hbar\frac{\partial\Psi}{\partial x}\right)^*$ as it should be (see Definition 17).

Fact 31. *Sum of Hermitians is a Hermitian.*

If $\hat{A}$ and $\hat{B}$ are Hermitian operators, then $\hat{A}+\hat{B}$ is also a Hermitian operator.

Proof. $\langle\alpha|\hat{A}+\hat{B}|\beta\rangle^* = \langle\alpha|\hat{A}|\beta\rangle^* + \langle\alpha|\hat{B}|\beta\rangle^* = \langle\beta|\hat{A}|\alpha\rangle + \langle\beta|\hat{B}|\alpha\rangle = \langle\beta|\hat{A}+\hat{B}|\alpha\rangle$. □

Fact 32. *Real Function of a Hermitian is a Hermitian.*

If $\hat{A}$ is a Hermitian operator, then any real analytic function of $\hat{A}$, $f(\hat{A}) \equiv \sum_n \frac{f^{(n)}(0)}{n!}\hat{A}^n$, is also Hermitian. Here, "real" function means that $f^{(n)}(0)$, the n-th derivative of $f(x)$ at $x=0$, is real for all n.

Proof. First, $\hat{A}^{n\dagger} = \hat{A}^{\dagger n}$ (due to Fact 21) $= \hat{A}^n$. Since $f^{(n)}(0)$ is real, then each term of $\sum_n \frac{f^{(n)}(0)}{n!}\hat{A}^n$ is Hermitian. By Fact 31, any sum of Hermitian operators is also Hermitian. □

Example 33. *Hamiltonian is Hermitian.*

The Hamiltonian operator, $\hat{H} = \hat{T}+\hat{V}$, where $\hat{T} = \frac{\hat{p}^2}{2m}$ and $\hat{V} =$ a real function of $\hat{x}$, is Hermitian.

Definition 34. *Anti-Hermitian, or imaginary, operator*

$\hat{O}$ is called anti-Hermitian, if $\hat{O}^\dagger = -\hat{O}$.

Fact 35. *Anti-commutator and Commutator*

If $\hat{A}$ and $\hat{B}$ are Hermitian, then the *anti-commutator*

$$\left\{\hat{A},\hat{B}\right\} \equiv \hat{A}\hat{B}+\hat{B}\hat{A}$$

is Hermitian and the *commutator*

$$\left[\hat{A},\hat{B}\right] \equiv \hat{A}\hat{B}-\hat{B}\hat{A}$$

is anti-Hermitian. One can write $\hat{A}\hat{B} = \frac{1}{2}\left\{\hat{A},\hat{B}\right\} + \frac{1}{2}\left[\hat{A},\hat{B}\right]$, a sort of decomposition of $\hat{A}\hat{B}$ into the real and the imaginary parts.

Proof. Follows directly from the reverse chain rule (Fact 21) of the Hermitian conjugate operation. □

Definition 36. ***Canonical Conjugate Variables***

Observables $\hat{A}$ and $\hat{B}$ are called canonical conjugate variables if

$$\left[\hat{A},\hat{B}\right] = \pm i\hbar$$

The $\pm$ sign makes the definition work even if $\hat{A}$ and $\hat{B}$ are swapped. $\hat{x}$ and $\hat{p}$ are one example. Another example is angle and angular momentum (see Example 48).

Definition 37. ***Inverse Operator***

The inverse of an operator $\hat{O}$ is written as $\hat{O}^{-1}$ and is defined by

$$\hat{O}\hat{O}^{-1} = \hat{O}^{-1}\hat{O} = 1$$

The necessary and sufficient condition for an inverse to exist can be stated in a few equivalent ways: (i) the determinant of $\hat{O}$ is non-zero, (ii) the image of $\hat{O}$ has the same dimension as the domain space, (iii) the column vectors of $\hat{O}$ are linearly independent, (iv) the row vectors of $\hat{O}$ are linearly independent.

Note. We need only one condition $\hat{O}\hat{O}^{-1} = 1$ or $\hat{O}^{-1}\hat{O} = 1$, as either one implies the other. Please refer to your favorite Linear Algebra book for the explanation.

Fact 38. ***Reverse Chain Rule for Inverse***

$\left(\hat{O}_1\hat{O}_2...\hat{O}_n\right)^{-1} = \hat{O}_n^{-1}...\hat{O}_2^{-1}\hat{O}_1^{-1}$

Proof. Because $\left(\hat{O}_1\hat{O}_2...\hat{O}_n\right)\hat{O}_n^{-1}...\hat{O}_2^{-1}\hat{O}_1^{-1} = 1.$ □

Definition 39. ***Unitary Operator***

$\hat{U}$ is *unitary* if $\hat{U}^\dagger = \hat{U}^{-1}$.

Theorem 40. ***Unitary Operator Conserves Norm and Vice Versa***

For a given vector $|\alpha\rangle$ and a unitary operator $\hat{U}$, define $|\beta\rangle = \hat{U}|\alpha\rangle$. Then $\langle\beta|\beta\rangle = \langle\alpha|\alpha\rangle$. Conversely, if an operator $\hat{U}$ conserves norm for any vector $|\alpha\rangle$, then it is unitary.

Proof. $\langle\beta| = \langle\alpha|\hat{U}^\dagger = \langle\alpha|\hat{U}^{-1}$. $|\beta\rangle = \hat{U}|\alpha\rangle$. So, $\langle\beta|\beta\rangle = \langle\alpha|\alpha\rangle$. To prove the converse, let us assume that $\langle\beta|\beta\rangle = \langle\alpha|\hat{U}^\dagger\hat{U}|\alpha\rangle = \langle\alpha|\alpha\rangle$ for any $|\alpha\rangle$. By Theorem 27, we see that $\hat{U}^\dagger\hat{U} = 1$, which means that $\hat{U}^\dagger = \hat{U}^{-1}$. □

Definition 41. ***Normal Operator***

$\hat{N}$ is *normal*, if $[\hat{N}, \hat{N}^\dagger] = 0$.

Definition 42. ***Eigenvectors and Eigenvalues***

If $\hat{O}|O\rangle = O|O\rangle$ where $|O\rangle \neq |null\rangle = 0$, we call O an *eigenvalue* and $|O\rangle$ an *eigenvector*. Note that $\hat{O}|O\rangle = O|O\rangle$ means $\langle O|\hat{O}^\dagger = \langle O|O^*$. We say that $|O\rangle$ is an *eigenket* of $\hat{O}$ with eigenvalue O and $\langle O|$ is an *eigenbra* of $\hat{O}^\dagger$ with eigenvalue O^*.

Definition 43. ***Degeneracy***

Suppose $\hat{O}|O_1\rangle = O|O_1\rangle$ and $\hat{O}|O_2\rangle = O|O_2\rangle$ and that $|O_1\rangle$ and $|O_2\rangle$ are linearly independent of each other. These two states $|O_1\rangle$ and $|O_2\rangle$ are then called *degenerate* eigenstates of $\hat{O}$. The number of linearly independent states that share the same eigenvalue is called the *degeneracy* of that eigenvalue.

Theorem 44. ***Normal Operators can be Diagonalized***

Consider a normal operator $\hat{N}$ defined in a finite dimensional inner product space $\mathcal{V}$. There exist mutually orthogonal eigenvectors $|N\rangle$ of $\hat{N}$, i.e. $\hat{N}|N\rangle = N|N\rangle$, and the number of eigenvectors is equal to the dimension of $\mathcal{V}$. Note that Hermitian operator and unitary operator are both normal operators (please prove this to yourself), so this theorem means that they can always be diagonalized.

Proof. Please consult your favorite Linear Algebra book for the proof of this main result of usual Linear Algebra. □

Note. ***How Useful Is This Big-Deal Theorem?***

The above standard result holds only in a *finite dimensional* inner product space. For our unusual Linear Algebra, it is difficult to state any such general theorem like that. So how useful *is* this theorem? In practice, it is actually quite useful. First of all, any numerical work deals with finite dimensional Hilbert spaces and there this theorem is directly applicable. In fact, if one realizes that any physics problem has upper and lower bounds in length (or momentum or energy) scales, then it is reasonable to view any physics problem as being defined in a *finite* dimensional inner product space, whose dimension may be very large but still finite. Second, in the analytical approach, we do have to deal with infinite dimensional vector spaces. In this case, "diagonalizing a matrix" amounts to solving a differential equation. A handful of differential equations can be exactly solved. Some may not be solved analytically. The majority of "real world" problems are not solvable, and they need to be approached with approximate methods, such as perturbation theory. More often than not, approximations lead to finite size matrices for which this theorem is directly applicable.

Theorem 45. ***Eigenvalues of a Hermitian Operator $\hat{A}$ are Real***

Proof. The eigenvalue equation $\hat{A}|A\rangle = A|A\rangle$ means $\langle A|\hat{A}|A\rangle = A\langle A|A\rangle$. On the other hand, taking the Hermitian conjugate of the entire identity, $\hat{A}|A\rangle = A|A\rangle$, we get $\langle A|\hat{A}^\dagger = \langle A|A^*$, which since $\hat{A}$ is Hermitian, means $\langle A|\hat{A} = \langle A|A^*$. Multiplying $|A\rangle$ on this, we get $\langle A|\hat{A}|A\rangle = A^*\langle A|A\rangle$. So, we got two expressions for $\langle A|\hat{A}|A\rangle$ and they must be the same: $A\langle A|A\rangle = A^*\langle A|A\rangle$. Eigenvectors cannot be the null vector, by definition, and so $A = A^*$, that is, A must be real. □

Theorem 46. ***Eigenvalues of a Unitary Operator $\hat{U}$ belong to $\{e^{i\theta}, \theta = real\}$***

Proof. Say $\hat{U}|U\rangle = U|U\rangle$. We already proved that $\hat{U}$ conserves norms, in Theorem 40. Thus $|U| = 1$. □

Theorem 47. ***Orthogonality for Eigenvectors of a Hermitian Operator***
Eigenvectors of a Hermitian operator are orthogonal if they have different eigenvalues. A finite number of eigenvectors that share the same eigenvalue can be always diagonalized, if they belong in the Hilbert space.

Proof. Consider two eigenvectors $|A\rangle$ and $|A'\rangle$ for a Hermitian operator $\hat{A}$. $\langle A|\hat{A}|A'\rangle = A\langle A|A'\rangle$, due to the reality of A. Also, $= A'\langle A|A'\rangle$. Thus, if $A \neq A'$, $\langle A|A'\rangle = 0$. What if $A = A'$? In that case, we can collect all eigenvectors $\{|A\rangle, |A'\rangle, |A''\rangle, ...\}$ with the same eigenvalue. In this case, any linear combination of the vectors in $\{|A\rangle, |A'\rangle, |A''\rangle, ...\}$ is also an eigenvector. These vectors can be chosen to be orthogonal to each other using the standard method, as long as there are finite number of them and they have finite norms, as will be discussed later in this document[13]. □

Example 48. ***Generators and Unitary Operators (Advanced Topic)***
It makes a lot of common sense to say that "momentum/energy makes things go." Also, angular momentum things go around. These concepts are neatly summarized in the notion of generators. A *generator* in QM is an observable that makes certain Unitary operation happen. So, how does $\hat{H}$ make things go in QM? This is what we call the deterministic character of the Schrödinger equation: $\hat{H}|\Psi\rangle = i\hbar\frac{\partial}{\partial t}|\Psi\rangle$. This equation reads $\hat{H}dt\Psi(x,t) = i\hbar\left[\Psi(x,t+dt) - \Psi(x,t)\right]$. Thus,

$$\Psi(x,t+dt) = \left(1 - i\frac{\hat{H}}{\hbar}dt\right)\Psi(x,t) \equiv \hat{\mathbb{U}}(dt)\Psi(x,t)$$

This is how $\hat{H}$ *generates* the time evolution[14] $\hat{\mathbb{U}}(dt)$. Is $\hat{\mathbb{U}}(dt)$ unitary? To linear order in dt, we have $\hat{\mathbb{U}}(dt)^\dagger\hat{\mathbb{U}}(dt) = \left(1 + i\frac{\hat{H}}{\hbar}dt\right)\left(1 - i\frac{\hat{H}}{\hbar}dt\right) = 1$, and so the answer is yes. It follows that $\hat{\mathbb{U}}(\Delta t)$ for any finite time Δt is also unitary.

[13] The standard procedure, Graham Schmidt orthogonalization procedure (Definition 56), may not be useful for uncountably infinite or possibly countably infinite number of eigenvectors in general. In real problems, the number of degenerate states tend to be finite. What happens typically is that a simultaneous eigenstates of more than one observable are sought for. Even if one operator (e.g. parity) has an infinite degeneracy, there is another commuting operator (e.g. Hamiltonian, momentum) that reduces the degeneracy to a finite number. However, one may still have the problem of diverging norms, which will make the Graham Schmidt orthogonalization procedure useless.

[14] Conventional notation for the time evolution operator is $\hat{U}$. On the other hand, $\hat{U}$ is also the conventional notation for a unitary operator. So, here, we use $\hat{\mathbb{U}}$ for the time evolution operator and $\hat{U}$ for a unitary operator.

Similarly, momentum $\hat{p}$ generates translation $\hat{T}$. $\hat{p}|\Psi\rangle = -i\hbar\frac{\partial}{\partial x}|\Psi\rangle$ means

$$\Psi(x+dx,t) = \left(1 + i\frac{\hat{p}}{\hbar}dx\right)\Psi(x,t) \equiv \hat{T}(dx)\Psi(x,t)$$

We have yet to deal with three dimensional problems to discuss angular momentum fully, but for the purpose of this discussion it suffices to consider a one dimensional problem for which a particle moves on a circle in the xy plane. We can simply measure the particle position using the angle coordinate ϕ. This particle has angular momentum along the z axis. It is quite straightforward to see that that angular momentum, which we will denote as $\hat{L}_\phi$, is to $\hat{\phi}$ in this problem what the linear momentum $\hat{p}$ is to $\hat{x}$ in the ordinary one dimensional problem. Thus, $\hat{L}_\phi = -i\hbar\frac{\partial}{\partial\phi}$, and we see that $\hat{L}_\phi$ generates rotation $\hat{R}$.

$$\Psi(\phi+d\phi,t) = \left(1 + i\frac{\hat{L_\phi}}{\hbar}d\phi\right)\Psi(\phi,t) \equiv \hat{R}(d\phi)\Psi(\phi,t)$$

This can be viewed as an operator equation, since it is valid for *any* wave function Ψ: $1 + i\frac{\hat{L_\phi}}{\hbar}d\phi = \hat{R}(d\phi) = \hat{R}(0) + d\hat{R} = 1 + d\hat{R}$, where in the last step we made use of the fact that $\hat{R}(0)$, no rotation, is the identity operator. So, we obtain $i\frac{\hat{L_\phi}}{\hbar}d\phi = d\hat{R}$, which can be readily integrated to give $\hat{R}(\Delta\phi) = \exp(i\hat{L}_\phi\Delta\phi/\hbar)$, where $\Delta\phi$ is now any finite amount.

Doing similarly for other operators, assuming no explicit time dependence in $\hat{H}$ in the case of $\hat{\mathbb{U}}$, we obtain these standard relations:

$$\hat{\mathbb{U}}(\Delta t) = \exp\left(\frac{-i\hat{H}\Delta t}{\hbar}\right),\ \hat{T}(\Delta x) = \exp\left(\frac{i\hat{p}\Delta x}{\hbar}\right),\ \hat{R}(\Delta\phi) = \exp\left(\frac{i\hat{L}_\phi\Delta\phi}{\hbar}\right)$$

where all Δ quantities are *finite.* Notice that all of these operators are unitary, as expected, and they all have the form of $\hat{U} = \exp(i\hat{h})$, where $\hat{h}$ is a Hermitian operator. So, all eigenvectors of $\hat{h}$ are also eigenvectors of $\hat{U}$ and thus diagonalize $\hat{U}$, with eigenvalues consistent with Theorem 46.

Example 49. ***Projection Operator***

For a given state $|\alpha\rangle$, $\hat{P}_\alpha \equiv |\alpha\rangle\langle\alpha|$ is a Hermitian operator, since $\hat{P}_\alpha^\dagger = (|\alpha\rangle\langle\alpha|)^\dagger = \langle\alpha|^\dagger|\alpha\rangle^\dagger = |\alpha\rangle\langle\alpha| = \hat{P}_\alpha$. This is called a projection operator if $|\alpha\rangle$ is normalized (or Dirac-normalized – cf. Theorem 61). Applying on a ket $|\beta\rangle$, it returns $\hat{P}_\alpha|\beta\rangle = |\alpha\rangle\langle\alpha|\beta\rangle = \langle\alpha|\beta\rangle|\alpha\rangle$. Acting on a bra $|\beta\rangle$, it returns $\langle\beta|\hat{P}_\alpha = \langle\beta|\alpha\rangle\langle\alpha|$. This is analogous to calculating the component of a vector $\vec{a}$ along a unit vector $\vec{e}$, by calculating the projection of $\vec{a}$ on $\vec{e}$: $\vec{e}\cdot\vec{a}$.

4. Different Types/Descriptions of Hilbert Space

Different types/descriptions of Hilbert space result depending on whether basis vectors are finite, countably infinite, or uncountably infinite.

Definition 50. ***Discrete, Continuous, and Mixed Set of States***

A set of states $\{|\alpha_n\rangle\}$ with integer index n are called *a discrete set of states.* A set of states $\{|\alpha_r\rangle\}$, where the index r covers a continuous range of real numbers, are called *a continuous set of states.* Discrete states will also be called *quantized states* or *bound states*, and continuous states will also be called *continuum states* or *scattering states.* When a set of states consist of some discrete states and some continuous states, they will be called *a mixed set.* Note that the number of elements of a discrete set can be finite or infinite. The number of elements of a continuum set is always uncountably infinite.

Definition 51. ***Linear Combination***

A linear combination, for given quantized states $\{|\alpha_n\rangle\}$ and continuum states $\{|\alpha_r\rangle\}$, is defined as

$$\sum_n c_n|\alpha_n\rangle + \int dr\phi(r)|\alpha_r\rangle$$

where

$$|c_n| < \infty \text{ for any } n$$

$$\left|\int_r^{r+0^+} dr'\phi(r')\right| < \infty \text{ for any } r$$

Note. To understand the origin of the above condition $|\int_r^{r+0^+} dr'\phi(r')| < \infty$, let us divide the real axis into small but finite segments of length h. Definite $|\alpha_n\rangle = \frac{1}{h}\int_{nh}^{nh+h} dr|\alpha_r\rangle$. Note that the starting point and the ending point of n is determined by the range of r. Then

$$\int dr\phi(r)|\alpha_r\rangle = \sum_n \int_{nh}^{nh+h} dr\phi(r)|\alpha_r\rangle \equiv \lim_{h\to 0}\sum_n \left[\int_{nh}^{nh+h} dr\phi(r)\right]|\alpha_n\rangle \tag{4.1}$$

In this way, we have reduced continuous states to discrete states, for which $|c_n| < \infty$ seems so reasonable. We identify $c_n = \int_{nh}^{nh+h} dr\phi(r)$, and thus we require that this be finite for any n, which is equivalent to saying $|\int_r^{r+h} dr'\phi(r')| < \infty$ for any r.

Fact 52. *Hermitian Conjugate of Linear Combinations*
If $|\beta\rangle = \sum_n c_n|\alpha_n\rangle$, then $\langle\beta| = \sum_n c_n^*\langle\alpha_n|$. If $|\beta\rangle = \int dr\phi(r)|\alpha_r\rangle$, then $\langle\beta| = \int dr\phi(r)^*\langle\alpha_r|$.

Proof. Left as exercise. □

Definition 53. *Linear Dependence*
A vector $|\alpha\rangle$ is said to be linearly dependent on a set of vectors, if $|\alpha\rangle$ can be expressed as a linear combination of the vectors of the set. Otherwise, it is linearly independent of the set.

Definition 54. *Linearly Independent Set of Vectors*
A set of vectors is called linearly independent, if *every* element of the set is linearly independent of the rest of the set. A linearly independent set of vectors has an important geometric meaning. For a given vector, an infinite length line can be generated by simply scaling that vector. No point of the line, except the origin, can be accessed by adding or scaling the rest of the vectors, i.e. by any linear combination of the rest of the vectors. In this way, each vector in a linearly independent set of vectors "owns" one dimension, belonging to that axis[15].

Definition 55. *Basis*
A set of linearly independent vectors in a vector space $\mathcal{V}$ are called *basis vectors*, or simply *basis,* of $\mathcal{V}$, if the set *spans* $\mathcal{V}$, i.e. if any vector of $\mathcal{V}$ can be expressed as a linear combination of the vectors of the set.

Definition 56. ***Graham-Schmidt Orthogonalization Procedure***
Suppose we found such a set of basis vectors, and discovered that there are a finite number of basis vectors, N. Then it follows that any other set of basis vectors for the same vector space should have the same number of basis vectors, N, which naturally defines the dimension of the vector space. As implied just now, for a given vector space $\mathcal{V}$, there is no unique way to define basis vectors. Indeed, for an N dimensional vector space, *any* linearly independent set of vectors of size N is a valid set of basis vectors. For finite dimensional vectors with finite norms, in an inner product space, a standard procedure exists to turn any basis vectors into orthogonal basis set, with each basis vector normalized, say one starts with a basis set $\{|e_1\rangle, ..., |e_N\rangle\}$. Define $|e_1'\rangle = A_1|e_1\rangle$ and $|e_n'\rangle = A_n\left[|e_n\rangle - \sum_{i=1}^{n-1}|e_i'\rangle\langle e_i'|e_n\rangle\right]$ for $n > 1$. Due to finite norms, A_n's can be, and are, chosen so that $\langle e_n'|e_n'\rangle = 1$. [Exercise: prove that $\langle e_n'|e_{n'}'\rangle = \delta_{nn'}$ for any n and $n' = 1, ..., N$.] This procedure is called the *Graham-Schmidt orthogonalization procedure.* While this procedure is useful,

[15]You might say that if we are considering complex numbers, which we are, then each "axis" is really two dimensions. While this is correct in a mathematical sense, it makes much more physical sense to call a set of complex numbers a set of one dimensional vectors.

its usefulness in infinite dimensional cases is doubtful. Luckily, Nature readily provides orthogonal basis vectors in QM, so it is not necessary to apply this type of procedure for infinitely many states.

Definition 57. ***Complete Basis***

A set of basis vectors are called *complete*, if they span the Hilbert space $\mathcal{W}$. Also, we will require that the basis vectors should not be "over-complete," in the sense that any basis vector $|e\rangle$ should have *some* overlap with *some* physical state, i.e. it cannot be that $\langle e|\Psi\rangle = 0$ for *all* states $|\Psi\rangle$ in $\mathcal{W}$[16].

Note. One thing worth noting is that basis vectors themselves may not belong in the Hilbert space. A simple example is the set of delta functions, $|e, x_0\rangle = \delta(x - x_0)$ for all x_0 values. Since for any function $f(x)$, $f(x) = \int dx_0 f(x_0)|e, x_0\rangle$, it is clear that $|e, x_0\rangle$ does span the Hilbert space of wave functions $\int dx|f(x)|^2 < \infty$. However, $\delta(x - x_0)$ itself is not a physical state, since $\int dx\delta(x - x_0)^2 = \delta(x_0 - x_0) = \infty$.

Definition 58. ***Natural Basis***

A set of basis vectors are called *natural,* if they satisfy these properties. The third property is a standard normalization procedure, which is always possible given the first two properties.

(1) Basis vectors are mutually orthogonal.
(2) Basis vectors are complete.
(3) Each basis vector is normalized in the following way.
    (a) If the basis vector $|e_n\rangle$ belongs in a discrete set of basis states, then $\langle e_m|e_n\rangle = \delta_{mn}$, i.e. the basis is *orthonormalized.*
    (b) If the basis vector $|e_r\rangle$ belongs in a continuous set of basis states, then $\langle e_r|e_s\rangle = \delta(r - s)$, i.e. the basis is *Dirac-orthonormalized.*

Theorem 59. ***Orthonormalization of a Discrete Natural Basis (no need to read if you just accept Definition 58)***

Suppose that we have a quantized basis set, $\{|e_n\rangle\}$, mutually orthogonal and complete. Then it follows that the basis vectors can be rescaled so that the new set $\{|e'_n\rangle = c_n|e_n\rangle\}$ satisfies the usual normalization rule:

$$\langle e'_m|e'_n\rangle = \delta_{mn}$$

We call the basis "orthonormalized."

Proof. Since the basis vectors are already mutually orthogonal, it suffices to show that $\langle e_n|e_n\rangle < \infty$, so that they are normalizable to 1. Suppose that there is a basis state $|e_m\rangle$ s.t. $\langle e_m|e_m\rangle = \infty$. Any state $|\alpha\rangle = c|e_m\rangle + \sum_{n\neq m} c_n|e_n\rangle$ for which $c \neq 0$ would be unphysical since $\langle\alpha|\alpha\rangle = |c|^2\langle e_m|e_m\rangle + \sum_{n\neq m} |c_n|^2\langle e_n|e_n\rangle$ (using the orthogonality) $\geq |c|^2\langle e_m|e_m\rangle = \infty$. Therefore, the basis set is "over-complete," clearly violating the assumption of a complete basis set (Definition 57). □

Corollary 60. ***Probability Sum Rule in the case of a Discrete Basis***

For a physical state $|\Psi\rangle = \sum_n c_n|e_n\rangle$ where $\{|e_n\rangle\}$ is a discrete natural basis, the following holds true.

$$\sum_n |c_n|^2 = 1$$

Proof. Direct consequence of the above theorem and Axiom 10. □

[16]This is not a serious condition at all. If there ever was such a basis vector, then we can simply throw away that basis vector without losing any physics. In reality, we never encounter such a "useless" basis vector in Quantum Mechanics.

Theorem 61. ***Dirac Orthonormalization of a Continuous Natural Basis (no need to read if you just accept Definition 58)***

Suppose that we have a continuum basis set, $\{|e_r\rangle\}$, mutually orthogonal and complete. Then it follows that the basis vectors cannot be normalized to $\langle e'_r|e'_s\rangle = \delta_{rs}$. However, it can be scaled so that the new set $\{|e'_r\rangle = c_r|e_r\rangle\}$ satisfies the following rule:

$$\langle e'_r|e'_s\rangle = \delta(r-s)$$

where $\delta(x)$ is the Dirac delta function, a function that satisfies $\int_{-\infty}^{\infty} dx\delta(x) = 1$ while non-zero only at $x = 0$. We call the basis "Dirac orthonormalized."

Proof. Due to the orthogonality, $\langle e_r|e_s\rangle \geq 0$ since the inner product is nonzero only when $r = s$. So, $\int ds\langle e_r|e_s\rangle = A_r \geq 0$. What we need to prove is that $A_r \neq 0$ so that $\left\{|e_r\rangle/\sqrt{A_r}\right\}$ forms the required Dirac normalized basis set $\{|e'_r\rangle\}$. We prove this by "reductio ad absurdum." Suppose $A_r = 0$ for some r. By virtue of completeness, for any physical state $|\Psi\rangle$ we have $|\Psi\rangle = \int ds\phi(s)|e_s\rangle$. Consider $\langle e_r|\Psi\rangle = \int ds\phi(s)\langle e_r|e_s\rangle$. Since the only non-zero point of the function $f(s) = \langle e_r|e_s\rangle$ is at $s = r$, it is clear that $\langle e_r|\Psi\rangle = \phi(r)\int ds\langle e_r|e_s\rangle = \phi(r)A_r$. This means that, if $A_r = 0$, then $\langle e_r|\Psi\rangle = 0$ for *any* physical state $|\Psi\rangle$. But this violates the assumption that the set $\{|e_r\rangle\}$ is complete but not over-complete. □

Note. The above theorem does not necessarily mean, nor did it assume, that A_r is finite for each r. For instance, it could be that $A_r \propto 1/r$, diverging at $r = 0$. Regardless, the Dirac normalization is possible. Related to this discussion, the procedure outlined in the proof can be viewed from a different angle. Instead of re-scaling basis vectors, one can simply change the variable from r to $z = \int^r A_s ds$ so that $\int ds\frac{\langle e_r|e_s\rangle}{A_r} = \int ds\frac{\langle e_r|e_s\rangle}{A_s} = \int dz'\langle e_z|e_{z'}\rangle = 1$, so that $\langle e_z|e_{z'}\rangle = \delta(z-z')$.

Corollary 62. ***Probability Sum Rule in the case of a Continuous Basis***

For a physical state $|\Psi\rangle = \int dr\phi(r)|e_r\rangle$ where $\{|e_r\rangle\}$ is a continuum natural basis, the following holds true:

$$\int dr|\phi(r)|^2 = 1$$

Proof. Direct consequence of the above theorem and Axiom 10. □

Corollary 63. ***Probability Sum Rule in the case of a Mixed Basis***

For a physical state $|\Psi\rangle = \sum_n c_n|e_n\rangle + \int dr\phi(r)|e_r\rangle$, where $\{|e_n\rangle\} + \{|e_r\rangle\}$ forms a mixed natural basis, the following holds true:

$$\sum_n |c_n|^2 + \int dr|\phi(r)|^2 = 1$$

Proof. Just like those of the above two corollaries. □

Theorem 64. ***Completeness as Operator Identity ("Resolutions of Identity")***

Let $\{|e_n\rangle\} + \{|e_r\rangle\}$ be a mixed natural basis. Then, the completeness means the following. For any state $|\alpha\rangle$ in the Hilbert space $\mathcal{W}$

$$|\alpha\rangle = \sum_n |e_n\rangle\langle e_n|\alpha\rangle + \int dr|e_r\rangle\langle e_r|\alpha\rangle$$

This is equivalent to the operator identity

$$\sum_n |e_n\rangle\langle e_n| + \int dr|e_r\rangle\langle e_r| = 1$$

In other words, the sum of all projection operators (see Example 49) for the natural basis is identity.

Proof. The completeness means that any state $|\alpha\rangle$ in the Hilbert space can be written as $|\alpha\rangle = \sum_n c_n \langle e_n| + \int dr \phi(r) \langle e_r|$. By multiplying $\langle e_m|$ or $\langle e_s|$ from the left, and using the orthonormality and the Dirac-orthonormality, $c_n = \langle e_n|\alpha\rangle$ and $\phi(r) = \langle e_r|\alpha\rangle$. This leads to the first result, which then leads to the second result immediately. □

Example 65. *Finite dimensional Hilbert space*

Recall the polarization measurement of photon with a linear polarizer. In that case, one can define the polarization vector as a two dimensional vector. More specifically, let us use notation $|X\rangle$ for the state of photon with an x polarization, and $|Y\rangle$ for the state of photon with a y polarization. Then, it follows that for transverse light the photon state with any polarization can be written as

$$|\alpha\rangle = c|X\rangle + d|Y\rangle$$

An explicit column vector representation, familiar from Classical Optics, can be used also. The following representation is standard.

$$|\alpha\rangle = \begin{bmatrix} c \\ d \end{bmatrix}, |X\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |Y\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The inner product can be thought of as the matrix multiplication between a row vector and a column vector.

$$\langle\alpha|\beta\rangle \equiv c^* e + d^* f$$

since

$$\langle\alpha| = \begin{bmatrix} c^* & d^* \end{bmatrix}, |\beta\rangle = \begin{bmatrix} e \\ f \end{bmatrix}$$

Note that the basis vectors are orthonormalized, $\langle X|X\rangle = \langle Y|Y\rangle = 1$ and $\langle X|Y\rangle = 0$, so they are natural basis vectors.

Example 66. ***Countably Infinite Dimensional Basis***

Let us consider the Harmonic Oscillator problem in one dimension. Let us define the vector $|n\rangle$ to mean the stationary state corresponding to the wave function $\psi_n(x)$, $n = 0, 1, 2, ...$ Our Hilbert space in this case is defined as the vector space spanned by these stationary states $\{|\alpha\rangle |\, |\alpha\rangle = \sum_n c_n|n\rangle\}$. We then define

$$\langle m|n\rangle = \int dx \psi_m^* \psi_n$$

which completely specifies our Hilbert space. We already know that

$$\langle m|n\rangle = \delta_{mn}$$

For any state $|\alpha\rangle = \sum_n c_n|n\rangle$ and $|\beta\rangle = \sum_n d_n|n\rangle$, the rule for the inner product (linearity) and this orthonormality leads to

$$\langle\alpha|\beta\rangle = \sum_n c_n^* d_n$$

The basis states are normalizable to 1, consistent with Theorem 59.

Example 67. ***Uncountably Infinite Dimensional Basis – Momentum Basis***

Let us consider the free particle problem in one dimension. Let us define the vector $|p\rangle$ to mean the stationary state corresponding to the wave function

$$|p\rangle = \psi_p(x) = \frac{\exp(ipx/\hbar)}{\sqrt{2\pi\hbar}}$$

p =any real number. These are eigenstates of the momentum operator $\hat{p}$ since

$$\hat{p}\psi_p(x) = p\psi_p(x)$$

Due to the Fourier theorem, the basis vectors $|p\rangle$ can span any (piecewise continuous) function. Thus, we have a vector space whose basis vectors satisfy the Dirac orthonormality, consistent with Theorem 61,

$$\langle p_1|p_2\rangle = \delta(p_1 - p_2)$$

Any (piecewise continuous) function $f(x)$ can be expressed as

$$|f\rangle = f(x) = \int dp\tilde{f}(p)\psi_p(x) = \int dp|p\rangle\langle p|f\rangle$$

where

$$\tilde{f}(p) = \langle p|f\rangle = \int dx\psi_p(x)^* f(x)$$

which are simply statements of Fourier and inverse Fourier transformations. Note that in this case, the basis states are *not* physical. Note that the inner product can be viewed as:

$$\langle f|g\rangle = \int dp\langle f|p\rangle\langle p|\int dp'|p'\rangle\langle p'|g\rangle = \int dp\langle f|p\rangle\langle p|g\rangle = \int dp\tilde{f}(p)^*\tilde{g}(p)$$

This is a hint that the real space wave functions $f(x)$ and $g(x)$ are not special. The *momentum space wave functions* $\tilde{f}(p)$ and $\tilde{g}(p)$ do an equally good job, and they could have been used to define vectors and the inner product in the first place. The Hilbert space itself is not uncountably infinite dimensional. It is only countably infinite dimensional. It is just the basis that is uncountably infinite dimensional. The origin of the uncountably infinite dimension is the continuum spectrum of $\hat{p}$.

Example 68. ***Uncountably Infinite Dimensional Basis – Position Basis***

Let us take $|x_0\rangle$ to mean the wave function $\delta(x - x_0)$. Note that

$$\hat{x}|x_0\rangle = x\delta(x - x_0) = x_0\delta(x - x_0) = x_0|x_0\rangle$$

so $|x_0\rangle$ is a position eigenket. Clearly this wave function is not physical, but if one collects all $|x_0\rangle$'s then they form a fine Dirac-orthonormalized basis set.

$$\langle x_0|x_0'\rangle = \int dx\delta(x - x_0)\delta(x - x_0') = \delta(x_0 - x_0')$$

and for any function $f(x)$

$$|f\rangle = f(x) = \int dx_0 f(x_0)\delta(x - x_0) = \int dx_0|x_0\rangle\langle x_0|f\rangle$$

Note that in this case the inner product of the two functions is given by

$$\langle f|g\rangle = \int dx_0 f(x_0)^* g(x_0)$$

which is what we started with! The Hilbert space description using the position basis is of course what we have been dealing with all along. The Hilbert space itself is not uncountably infinite dimensional. It is only countably infinite dimensional. It is just the basis that is uncountably infinite dimensional. The origin of the uncountably infinite dimension is the continuum spectrum of $\hat{x}$.

Definition 69. *Subspace*

Given a Hilbert space $\mathcal{W}$, let us say that a basis set is given. If one takes a subset of the basis set, and consider the span of that basis set, one obtains a subspace of the original Hilbert space. This is a rather trivial concept, but a useful one to keep in mind nonetheless. It is because often it is sufficient to consider a small subspace of the original Hilbert space, as some operators are confined only within that subspace.

Definition 70. ***Product Space***

Suppose one has a Hilbert space $\mathcal{W}_1$ with orthogonal basis vectors $|e_1\rangle, |e_2\rangle, ..., |e_N\rangle$ and another Hilbert space $\mathcal{W}_2$ with orthogonal basis vectors $|\varepsilon_1\rangle, |\varepsilon_2\rangle, ..., |\varepsilon_M\rangle$. We assume both are derived from the same field $\mathcal{C}$. Let us then define new $N \times M$ basis vectors by direct products $|i,j\rangle \equiv |e_i\rangle \otimes |\varepsilon_j\rangle$, or simply $|e_i\rangle|\varepsilon_j\rangle$ to mean the same thing, and then span a vector space $\mathcal{W}$ by forming linear combinations $|\alpha\rangle = \sum_{i,j} c_{i,j}|i,j\rangle$. Clearly $\mathcal{W}$ is another inner product space if one defines

$\langle i, j|i'j'\rangle \equiv \langle e_i|e_{i'}\rangle\langle\varepsilon_j|\varepsilon_{j'}\rangle = \delta_{i,i'}\delta_{j,j'}$, and then *use* the linear property of the inner product rule #3 to define the general inner product $\langle\alpha|\beta\rangle$. Thus one defines a *tensor product space*, or a *direct product space*,$\mathcal{W} \equiv \mathcal{W}_1 \otimes \mathcal{W}_2$.

A product space can be generated just as easily from all types of Hilbert spaces discussed so far. A product space is more a rule than an exception in Quantum Mechanics. There is the spatial wave function space, and then there is another space for "spin" etc. etc. Or, there are more than one particle. For such things as *entanglement* to occur it is necessary that the Hilbert space is a product space.

5. Law of Measurement

We have laid grounds to ingredients – states and observables – but we do not have any law of Nature yet. By making contacts with measurement, the mathematical framework discussed so far becomes part of scientific laws.

Definition 71. ***Measurement***

Usually, the term measurement is taken for granted. That seems OK, since we kind of think we know what it means. But let me make some important points.

1. Measurement *does not require any human* involvement. This is obvious but may be worth emphasizing. Once the setup is prepared, either by Nature herself or by people, it is perfectly OK even if no person cares to look at the outcome of the measurement.
2. Measurement *requires equipment*. Here I use the term equipment to mean not just man-made objects but also relevant objects in Nature. Maybe a better term is a "large system," consisting of many, on the order of 10^{23}, particles. Actually, let us *define* measurement as an event[17] in which the overall state of the large system changes due to its interaction with a Quantum Mechanical system. The sense of amplification is contained in this definition.
3. We assume, in our theorist hats, that measurement equipment is as sophisticated as possible, so that measurement can discern any discrete quantum levels.

Axiom 72. ***III. Law of Measurement or General Statistical Interpretation***

Measurement of an observable $\hat{A}$ on a physical state $|\Psi\rangle$ returns an eigenvalue or a range of eigenvalues according to the following law. We distinguish the discrete case, in which the eigenvalues of $\hat{A}$ are quantized, and the continuous case, in which the eigenvalues of $\hat{A}$ form a continuum.

In the discrete case, the measurement results in any one of the eigenvalues, A. The probability for the measurement to yield the result A is given by $|\langle A|\Psi\rangle|^2$. After the measurement, the new wave function is the eigenstate $|A\rangle$.

In the continuous case, the measurement has the probability $|\langle A|\Psi\rangle|^2\delta A$ to return a value from A to $A + \delta A$ (or equivalently from $A - \delta A/2$ to $A + \delta A/2$, if you prefer), assuming δA is a small but finite number. After the measurement, the new wave function is a "sqquare pulse" wave packet in the following sense: $|\Psi_{new}\rangle = \int dA'\phi(A')|A'\rangle$ where $\phi(A') = 1/\sqrt{\delta A}$ if $A < A' < A + \delta A$ and 0 otherwise. Here, for simplicity we assumed an ideal "square shaped measurement aperture" for our equipment. In reality, $\phi(A')$ may be a more general function, determined by the shape of the"resolution function" of the equipment. In any case, the width δA of $\phi(A')$ characterizes the resolution of the equipment. The compact way to generally describe the probability is $|\langle\Psi_{new}|\Psi\rangle|^2$, where $|\Psi_{new}\rangle = \int dA'|A'\rangle\phi(A')$.

In fact, the expression for the probability $|\langle\Psi_{new}|\Psi\rangle|^2$ is valid for both discrete or continuous cases, with $|\Psi_{new}\rangle = |A\rangle$ for the discrete case and $|\Psi_{new}\rangle = \int dA'|A'\rangle\phi(A')$ for the continuous case.

[17]Indeed, Heisenberg preferred the term "event" to terms like "observation" and "measurement." I agree. I think "measurement" and "observable" are serious misnomers if they make people think incorrectly that human involvement to "observe" or "measure" is essential for understanding the foundation of Quantum Mechanics.

Note. The fact that $|\Psi\rangle$ becomes $|\Psi_{new}\rangle$ is described as a "wave function collapse." This sounds overly dramatic, and it really is. Remember that during measurements big events happen due to the interaction of a large system with a QM system, like a light bulb lights up or a ball rolls down or a Geiger counter clicks etc. These events involve a truly huge number of particles. The fact that the small QM system goes through a change in the wave function in the same process is almost nothing compared to these big events, and so the wave function collapse is hardly a surprise!

Axiom 73. ***IV. Eigenstates of Any Observable are Complete***
The set of eigenvectors of any observable $\hat{A}$ is complete, in the sense of Definition 57.

Note. The meaning of this axiom is very clear physically. Suppose that the set of eigenvectors of a certain observable $\hat{A}$ was not complete. Then, it means that there is a physical state $|\Psi\rangle = |W\rangle + |X\rangle$, where $|W\rangle = \sum_n c_n|A_i\rangle + \int dr\phi(r)|A_r\rangle$ is the part that can be expressed in terms of eigenvectors of $\hat{A}$ and $|X\rangle$ is the part that *cannot* be expressed in that fashion. This means $\langle W|X\rangle = 0$. Thus, $\langle\Psi|\Psi\rangle = \langle W|W\rangle + \langle X|X\rangle$. By definition, $|\Psi\rangle$ represents a physical state, and thus it is normalized to 1, and so $\langle X|X\rangle, \langle W|W\rangle < 1$. Thus, both $|X\rangle$ and $|W\rangle$ are normalizable, and so $|\Psi\rangle - |W\rangle = |X\rangle$ is a separate physical state when normalized to 1. But $|X\rangle$ is a state that cannot be measured! If $\hat{A} = \hat{x}$, what this means is that $|X\rangle$ is a state that *does not exist* at any x, since measurement of $\hat{x}$ will produce absolutely no result. Well, if a particle does not exist at any x, then it is by definition a non-existent particle, and therefore $|X\rangle = |null\rangle = 0$. This is the meaning of the above completeness axiom.

Theorem 74. ***Measured Quantity***

For a physical state $|\Psi\rangle$, all physically measured quantities can be expressed as $\langle\Psi|\hat{h}|\Psi\rangle$ for some Hermitian operator $\hat{h}$. The short-hand notation for this is $\langle\hat{h}\rangle$.

Proof. By Axiom 72, what we measure is simply the probability $|\langle\Psi_{new}|\Psi\rangle|^2$, which is $\langle\Psi|\Psi_{new}\rangle\langle\Psi_{new}|\Psi\rangle$ and so we can take $\hat{h}$ to be the projection operator $|\Psi_{new}\rangle\langle\Psi_{new}|$, which is Hermitian. □

6. Time Evolution and Pictures

6.1. ***Time Evolution Operator.*** The Schrödinger equation $i\hbar\frac{\partial}{\partial t}\Psi(x,t) = \hat{H}\Psi(x,t)$ can be written in Dirac notation as

$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle \tag{6.1}$$

Note that the partial derivative for the wave function becomes a total derivative for the Dirac ket, since an unrepresented state (cf. Section 9) does not depend on any other variable than t. We saw in one homework problem that the time evolution operator is, for $\hat{H}$ that does not explicitly depend on t, $\hat{\mathbb{U}}(t) = \exp(-i\hat{H}t/\hbar)$ and $|\Psi(t)\rangle = \hat{\mathbb{U}}(t)|\Psi(0)\rangle$. For small time dt, we see that

$$\hat{\mathbb{U}}(dt) = 1 - i\frac{\hat{H}}{\hbar}dt$$

This expression for small dt is valid for *any* $\hat{H}$, even when it has an explicit time dependence (see Example 48). This expression for $\hat{\mathbb{U}}(dt)$ follows immediately from the above Schrödingerequatition, also, as it should, if one donsiders the Taylor expansion of $|\Psi(t+dt)\rangle$. When applied to $|\Psi\rangle$, this operator advances time by dt

$$\hat{\mathbb{U}}(dt)|\Psi(t)\rangle = |\Psi(t+dt)\rangle$$

Note that $\hat{\mathbb{U}}(dt)\hat{\mathbb{U}}^\dagger(dt) = 1$ to linear order in dt, and so $\hat{\mathbb{U}}(dt)$ is unitary, and actually this proves that $\hat{\mathbb{U}}(t)$ is unitary for *any* t and *any* Hermitian $\hat{H}$.

6.2. ***Schrödinger Picture.*** One of the first things that we learn in QM is the Schrödinger equation, which says that wave functions evolve in time. In contrast, operators such as $\hat{x}$ and $\hat{p}$ are viewed as independent of time. This convention goes by the name of *Schrödinger picture.* In this picture, all operators are time-independent, unless e.g. there is an external field coupling to the system to give an *explicitly* time dependent dynamical variable, for instance the Hamiltonian under the influence of an external AC field.

In this picture, the solution of a QM problem can be written as

$$|\Psi(t)\rangle = \hat{\mathbb{U}}(t)|\Psi(0)\rangle \tag{6.2}$$

whether or not $\hat{H}$ is explicitly t-dependent. If $\hat{H}$ is explicitly t-dependent, $\hat{\mathbb{U}}(t)$ is very complicated. In any case $\hat{\mathbb{U}}(t)$ is always unitary as long as $\hat{H}$ is Hermitian.

6.3. ***Time Evolution of*** $\langle\hat{O}\rangle$. The following identity holds for any state $|\Psi\rangle$ and for any operator $\hat{O}$:

$$\frac{d\left\langle\hat{O}\right\rangle}{dt} = \frac{i}{\hbar}\left\langle\left[\hat{H},\hat{O}\right]\right\rangle + \left\langle\frac{\partial\hat{O}}{\partial t}\right\rangle \tag{6.3}$$

To derive this important relation, we note first

$$d\langle\hat{O}\rangle = \langle\Psi(t+dt)|\hat{O}(t+dt)|\Psi(t+dt)\rangle - \langle\Psi(t)|\hat{O}(t)|\Psi(t)\rangle$$

Here, the time dependence $\hat{O}(t)$ is to note any *explicit* time dependence that it might have. *By definition*, we write $\hat{O}(t+dt) = \hat{O}(t) + dt\left(\partial\hat{O}/\partial t\right)$, using *partial t-derivative.* See Note below. We use $\hat{\mathbb{U}}(dt)|\Psi(t)\rangle = |\Psi(t+dt)\rangle$ with $\hat{\mathbb{U}}(dt) = 1 - i\frac{\hat{H}}{\hbar}dt$. Also, keep in the back of your mind that $\hat{H}$ may depend explicitly on t as well, so a full notation for it is $\hat{H}(t)$. So, we get

$$d\langle\hat{O}\rangle = \langle\Psi(t)|\left[1+\frac{i\hat{H}dt}{\hbar}\right]\left[\hat{O}(0) + dt\left(\partial\hat{O}/\partial t\right)\right]\left[1-\frac{i\hat{H}dt}{\hbar}\right]|\Psi(t)\rangle - \langle\Psi(t)|\hat{O}(t)|\Psi(t)\rangle$$

which means, by keeping terms in linear order in dt,

$$d\langle\hat{O}\rangle = \langle\Psi(t)|\frac{\partial\hat{O}}{\partial t}|\Psi(t)\rangle dt + \langle\Psi(t)|\frac{i}{\hbar}\left(\hat{H}\hat{O}(t) - \hat{O}(t)\hat{H}\right)|\Psi(t)\rangle dt$$

which can be rewritten as

$$\frac{d\langle\hat{O}\rangle}{dt} = \frac{i}{\hbar}\langle\Psi(t)|[\hat{H}(t),\hat{O}(t)]|\Psi(t)\rangle + \langle\Psi(t)|\frac{\partial\hat{O}}{\partial t}|\Psi(t)\rangle$$

which is identical to Eq. 6.3 that we set out to derive.

Note. **What in the world does $\partial\hat{O}/\partial t$ mean?**

It means an *explicit* and *extrinsic* time dependence. It is caused by a "meddling hand" from outside. Consider as an example the harmonic potential $\hat{O} = \frac{1}{2}m\omega^2\hat{x}^2$ of a QM spring. Suppose during your experiment, day became night and the room temperature droped by whopping 10 degrees. The natural frequency ω of your spring changed due to this annoyance. We can express this as $\frac{d\omega}{dt}$. In the discussion so far, if an operator has any time dependence, this type of time dependence is all there is. So, why not write this as $d\hat{O}/dt = m\omega\frac{d\omega}{dt}\hat{x}^2$. In due respect to Heisenberg, we do not do that. Instead we *define* this as $\partial\hat{O}/\partial t = m\omega\frac{d\omega}{dt}\hat{x}^2$. Heisenberg viewed any dynamical variable, e.g. $\hat{x},\hat{p},\hat{L}_\phi,...$, as *intrinsically* time dependent, in the sense that it is driven by $\hat{H}$ of the system (see the commutator part in Eq. 6.6). The fact that ω may be changing from $\omega(t)$ to $\omega(t+dt)$ has nothing to do with $\hat{H}(t)$ of the system and has everything to do with outside weather, and so that part is *extrinsic.* In this example, note that $\hat{a}_-$ and $\hat{a}_+$ will have non-zero partial time derivatives as well, since they are dependent on the parameter ω. Another common example is the situation where a charged particle is driven by an external AC field, e.g. the potential $\hat{O} = -eE\hat{x}\cos\omega t$. In this case, $\partial\hat{O}/\partial t = eE\hat{x}\sin\omega t$ (reminder: $\hat{x}$ here is the x operator, not the unit vector along $\hat{x}$!). In standard QM textbooks, this latter example is the standard, and probably the only, problem where a non-zero $\partial\hat{O}/\partial t$ is encountered. Note that our Iron Man problem of midterm also involved an explicitly time dependent potential.

6.4. ***Heisenberg Picture.*** Quantum Mechanics can be described in several different ways. Here we describe only Heisenberg picture as an alternative view. Classical Mechanics also has several different pictures: Newton mechanics, Lagrangian mechanics, and Hamiltonian mechanics.

Note that due to Theorem 74, what we care about in the end are quantities of the form $\langle\Psi|\hat{h}|\Psi\rangle$, where $\hat{h}$ is a Hermitian operator. *Heisenberg picture* takes the view that $|\Psi\rangle$ does not evolve in time, but the operator $\hat{h}$ does. For this reason, this picture is described as Heisenberg "matrix mechanics" as opposed to Schrödinger "wave mechanics." Either choice is fine as long as $\langle\Psi|\hat{h}|\Psi\rangle$ comes out right.

Note that in the discussions above, when an operator had an explicit time dependence $\hat{O}$ we wrote, by definition, the involved time differentiation as $\partial\hat{O}/\partial t$. This does not make much sense in the Schrödinger picture, where that type of time dependence is all time dependence that an operator will ever have. The reason for that notation was in anticipation of Heisenberg picture, in which an operator is totally time dependent in general and a state is totally time independent. We set the reference time to be an arbitrary time, say $t = 0$, and define

$$|\Psi\rangle_H = |\Psi(t=0)\rangle_S$$

and then the Heisenberg picture expression corresponding to Equation 6.2 is

$$\hat{O}_H(t) = \hat{\mathbb{U}}(t)^\dagger \hat{O}_S(t)\hat{\mathbb{U}}(t) \tag{6.4}$$

The subscripts H and S are used to distinguish the two pictures and to avoid any confusion. Then, Eq. 6.3 can be rewritten as

$$\langle\Psi(0)|\frac{d\hat{O}_H}{dt}|\Psi(0)\rangle = \langle\Psi(0)|\frac{i}{\hbar}\left[\hat{H}_H,\hat{O}_H\right] + \left(\frac{\partial\hat{O}_S}{\partial t}\right)_H|\Psi(0)\rangle \tag{6.5}$$

where for the commutator part, the following has been used, using the unitary property $\hat{\mathbb{U}}(t)\hat{\mathbb{U}}(t)^\dagger = 1$:

$$\begin{aligned}
\hat{\mathbb{U}}(t)^\dagger[\hat{H}_S,\hat{O}_S]\hat{\mathbb{U}}(t) &= \hat{\mathbb{U}}(t)^\dagger(\hat{H}\hat{O}_S - \hat{O}_S\hat{H})\hat{\mathbb{U}}(t) \\
&= \hat{\mathbb{U}}(t)^\dagger\hat{H}\hat{\mathbb{U}}(t)\hat{\mathbb{U}}(t)^\dagger\hat{O}_S\hat{\mathbb{U}}(t) - \hat{\mathbb{U}}(t)^\dagger\hat{O}_S\hat{\mathbb{U}}(t)\hat{\mathbb{U}}(t)^\dagger\hat{H}\hat{\mathbb{U}}(t) \\
&= \hat{H}_H\hat{O}_H - \hat{O}_H\hat{H}_H \\
&= [\hat{H}_H,\hat{O}_H]
\end{aligned}$$

Eq. 6.5 is valid for any $|\Psi(0)\rangle$, which means an operator identity due to Theorem 27:

$$\frac{d\hat{O}_H}{dt} = \frac{i}{\hbar}\left[\hat{H}_H,\hat{O}_H\right] + \left(\frac{\partial\hat{O}_S}{\partial t}\right)_H \tag{6.6}$$

This is the celebrated *equation of motion in Heisenberg picture*, where any operator $\hat{O}_H$ is now considered fully time dependent. The cumbersome notation here for the last term is to avoid any confusion[18]. In any case, all operators, e.g. $\hat{x}$, $\hat{p}$, $\hat{a}_-$, $\hat{a}_+$, are time dependent in this picture. In

[18] To be fair, it really should be written as $\partial\hat{O}_H/\partial t$, and should be understood as the differentiation on t but not on fundamental dynamical variables such as $\hat{x}$, $\hat{p}$, etc. I just want to avoid any misunderstanding that $\hat{O}_H = \hat{\mathbb{U}}(t)^\dagger\hat{O}_S(t)\hat{\mathbb{U}}(t)$ and so $\partial\hat{O}_H/\partial t$ applies to all three operators on the right side. This is of course wrong, since that would be the total time derivative. The important point that needs to be understood is that Heisenberg did not start from the Schrödinger picture to derive his equation of motion. Take the position operator $\hat{x}$. He simply wrote $\hat{x}(t)$. The same for $\hat{p}(t)$. In one dimension (for a spin-less or "color-less" particle) any other dynamical quantity may be written as $D(\hat{x},\hat{p},t)$. If D is just a function of $\hat{x}$,$\hat{p}$, then $\partial D/\partial t = 0$, by definition, while D

some sense, this is a more natural view, since its equation of motion, Equation 6.6, directly poses questions like "how does the momentum change in time," or "how does the amplitude of harmonic oscillator change in time." As such, Heisenberg picture is more intuitive in discussing conservation rules (Section 8) and for certain problems. However, for the most part, we will stick with the the Schrödinger picture. This means that the subscript S is implied automatically, unless the explicit subscript H overrides it. Note that Eq. 6.3 is the equation for an expection value, and as such is valid in any picture.

7. Uncertainty Principle

Theorem 75. ***Uncertainty Principle for Two Observables***
Let $\hat{A}, \hat{B}$ be two observables. For a physical state,

$$\Delta A \Delta B \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right|$$

where $\Delta A \equiv \sqrt{\langle (\hat{A} - \langle \hat{A} \rangle)^2 \rangle}$ *and* $\Delta B \equiv \sqrt{\langle (\hat{B} - \langle \hat{B} \rangle)^2 \rangle}$.

Proof. Let $|\alpha\rangle = (\hat{A} - \langle \hat{A} \rangle)|\Psi\rangle$, and $|\beta\rangle = (\hat{B} - \langle \hat{B} \rangle)|\Psi\rangle$. Then, $\Delta A = \sqrt{\langle \alpha | \alpha \rangle}$ and $\Delta B = \sqrt{\langle \beta | \beta \rangle}$. Applying the Schwartz inequality (Theorem 5), we have $\Delta A \Delta B \geq |\langle \alpha | \beta \rangle|$. Now, $\langle \alpha | \beta \rangle = \langle \hat{A}\hat{B} \rangle - \langle \hat{A} \rangle \langle \hat{B} \rangle$. This means that $Re\langle \alpha | \beta \rangle = \frac{1}{2}\langle \{\hat{A}, \hat{B}\} \rangle - \langle \hat{A} \rangle \langle \hat{B} \rangle$ and $Im\langle \alpha | \beta \rangle = \frac{1}{2i}\langle [\hat{A}, \hat{B}] \rangle$, where the symbol $\{\hat{A}, \hat{B}\} = \hat{A}\hat{B} + \hat{B}\hat{A}$ is the *anti-commutator* and $[\hat{A}, \hat{B}]$ is the usual commutator. Taking the imaginary part to estimate the lower bound for $|\langle \alpha | \beta \rangle|$, we have $\Delta A \Delta B \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right|$. □

Definition 76. ***Compatible and Incompatible Observables***
Two observables $\hat{A}, \hat{B}$ for which $[\hat{A}, \hat{B}]$ is non-zero are called *incompatible observables.* Due to the uncertainty principle, for two incompatible variables, if one variable is measured with high accuracy then the other variable becomes poorly defined and vice versa. In contrast, if two observables $\hat{A}, \hat{B}$ are compatible then they can be measured as accurately as possible at the same time [see Section 8]. Another way of distinguishing the two cases is the following. Say $\hat{A}$ and $\hat{B}$ are two compatible observables. To make the discussion simple, let us assume discrete eigenvalues for both of them. One measures $\hat{A}$ first. Say we got result A. In general, measurement disturbs the wave function and the wave function changes to an eigenstate of $\hat{A}$. Now one measures $\hat{B}$. Say we got B. This time, there is some chance that the wave function was disturbed, but the chance is small in practice. Anyhow, let us now continue. Measure $\hat{A}$ again. This time the wave function is *not* disturbed at all, except for a disturbance in the overall phase. Furthermore, we get result A with 100 % certainty! Measure $\hat{B}$ again. Again, no wave function disturbance, and we get B with certainty. This game can go on indefinitely now, and we will be certain that the wave function is not disturbed and results A and B are all we will get. In contrast, if $\hat{A}$ and $\hat{B}$ are not compatible, and if we follow the same measurement sequence, then in general [excluding some fortuitous case] *every time* the measurement is done the wave function is disturbed and there is no certainty about the outcome of the measurement. Generally, we will get a sequence of results like $ABA'B'A''B''$.... The underlying reason for these different behaviors is Theorem 80.

Corollary 77. ***Uncertainty Principle for Canonical Conjugate Variables***
All canonical conjugate variables (Definition 36) correspond to incompatible variables. For them,

$$\Delta A \Delta B \geq \hbar/2$$

---

may have a full time dependence through $\hat{x}$ and $\hat{p}$. Schrödinger, in contrast, looked at $\hat{x}$, $\hat{p}$ as completely static objects. Both these views are fine, and we may use the equation $D(\hat{x}_H, \hat{p}_H, t) = \hat{\mathbb{U}}(t)^\dagger D(\hat{x}_S, \hat{x}_S, t)\hat{\mathbb{U}}(t)$ or $D(\hat{x}_S, \hat{p}_S, t) = \hat{\mathbb{U}}(t) D(\hat{x}_H, \hat{x}_H, t)\hat{\mathbb{U}}^\dagger(t)$ to go between the two views and that is helpful sometimes, but the point here is that it is actually totally unnecessary to consider these conversion formulae to establish the equation of motion in each wonderful and self-sufficient formulation of QM.

One example is $\hat{x}$ and $\hat{p}$. Thus, $\Delta x \Delta p \geq \frac{\hbar}{2}$ from above. Also, $\hat{\phi}$ and $\hat{L}_\phi = -i\hbar\frac{\partial}{\partial\phi}$ (see Example 48) are conjugate variables, since they satisfy $\left[\hat{\phi}, \hat{L}_\phi\right] = i\hbar$. Thus, $\Delta\phi\Delta L_\phi \geq \frac{\hbar}{2}$.

Fact 78. ***Minimum Uncertainty Wave Function***
In the Schwartz inequality, the equal sign holds when the two vectors are parallel. This gives the minimum uncertainty condition. $|\beta\rangle = \left(\hat{B} - \langle\hat{B}\rangle\right)|\Psi\rangle = C|\alpha\rangle = C\left(\hat{A} - \langle\hat{A}\rangle\right)|\Psi\rangle$. Also, $Re\langle\alpha|\beta\rangle$ needs to be zero for the equality in the above uncertainty relation to hold. Thus, C must be purely imaginary, so we write it as $C = ia$, where a is real. This means $\left(\hat{B} - \langle\hat{B}\rangle\right)|\Psi\rangle = ia\left(\hat{A} - \langle\hat{A}\rangle\right)|\Psi\rangle$. For $\hat{A} = \hat{x}$ and $\hat{B} = \hat{p}$, we have

$$\left(-i\hbar\frac{\partial}{\partial x} - \langle\hat{p}\rangle\right)\Psi(x) = ia(x - \langle\hat{x}\rangle)\Psi(x)$$

$$\left[x + \frac{\hbar}{a}\frac{\partial}{\partial x}\right]\Psi(x) = \left[\langle\hat{x}\rangle + \frac{i\langle\hat{p}\rangle}{a}\right]\Psi(x)$$

Not surprisingly, this equation has the same form as the equation for the coherent state of the Harmonic oscillator problem. Putting $y = \left[x - \langle\hat{x}\rangle - \frac{i\langle\hat{p}\rangle}{a}\right]^2$, we obtain $dy = 2dx\left[x - \langle\hat{x}\rangle - \frac{i\langle\hat{p}\rangle}{a}\right]$ and so we obtain

$$\frac{\partial\Psi}{\partial y} = -\frac{a}{2\hbar}\Psi$$

and so

$$\Psi(x) \propto \exp\left(-\frac{a}{2\hbar}\left[x - \langle\hat{x}\rangle - \frac{i\langle\hat{p}\rangle}{a}\right]^2\right)$$

This can be rewritten as

$$\Psi(x) = A\exp\left(-\frac{a}{2\hbar}\left[x - \langle\hat{x}\rangle\right]^2\right)\exp\left(\frac{i\langle\hat{p}\rangle x}{\hbar}\right)$$

which is a Gaussian wave packet, with a net momentum $\langle\hat{p}\rangle$. Note that $a > 0$ is required, and the t dependence of Ψ is not expressed explicitly since t is fixed throughout.

Theorem 79. ***Uncertainty Principle for Energy and Time***
It is tempting to consider $\hat{H}$ and t as canonical conjugate variables, since $\hat{H} = i\hbar\frac{\partial}{\partial t}$. But unfortunately, this does not make sense yet since t is not an observable in non-relativistic Quantum Mechanics. However, we do have the uncertainty principle for time and energy that has the same form as that of two canonical conjugate observables:

$$\Delta E\Delta t \geq \frac{\hbar}{2}$$

where $\Delta E = \sqrt{\langle(\hat{H} - \langle\hat{H}\rangle)^2\rangle}$ and $\Delta t \equiv \frac{\Delta Q}{|d\langle\hat{Q}\rangle/dt|}$ for any observable $\hat{Q}$ which does not depend explicitly on time.

Proof. Putting $\hat{A} = \hat{H}$ and $\hat{B} = \hat{Q}$ where $\hat{Q}$ is some observable, we get $\Delta E\Delta Q \geq \left|\frac{1}{2i}\langle[\hat{H}, \hat{Q}]\rangle\right|$. There is really no point in considering an explicitly time dependent operator $\hat{Q}$ as that time dependence, driven by external fields for example, is controlled "by external hands" and thus would not be correlated with any intrinsic property of the system. Thus, assuming no explicit time dependence in $\hat{Q}$, Equation 6.3 shows that $\langle[\hat{H}, \hat{Q}]\rangle = -i\hbar d\langle\hat{Q}\rangle/dt$. So we get $\Delta E\Delta Q \geq \frac{\hbar}{2}\left|d\langle\hat{Q}\rangle/dt\right|$. Defining $\Delta t \equiv \frac{\Delta Q}{|d\langle\hat{Q}\rangle/dt|}$, i.e. the time scale in which $\langle\hat{Q}\rangle$ changes by ΔQ, or, more colloquially, the time scale in which $\langle\hat{Q}\rangle$ changes appreciably, we get the desired result. □

8. Symmetry and Conservation

Theorem 80. ***Compatible Observables and Simultaneous Eigenstates***

Suppose two observables $\hat{A}$ and $\hat{B}$ are compatible (Definition 76), i.e. $[\hat{A}, \hat{B}] = 0$. Assume that $\mathcal{V}_A$, the vector space spanned by eigenkets of $\hat{A}$, is contained in $\mathcal{V}_B$, the vector space spanned by eigenkets of $\hat{B}$. Then, there is a complete basis each member of which is a simultaneous eigenstate of $\hat{A}$ and $\hat{B}$.

Note. The above assumption about $\mathcal{V}_A$ and $\mathcal{V}_B$ is to avoid a weird case when $\mathcal{V}_A$ and $\mathcal{V}_B$ intersect only in the Hilbert space (which should be contained by both $\mathcal{V}_A$ and $\mathcal{V}_B$). Such a weird case probably never happens, but it seems hard to prove in general that it does not.

Proof. Due to the assumption, each eigenstate of $\hat{A}$ can be expressed as linear combinations of eigenstates of $\hat{B}$. Each of the eigenstates of $\hat{B}$ that are used to describe an eigenstate of $\hat{A}$ is then necessarily a simultaneous eigenstate of $\hat{A}$ and $\hat{B}$. Since $\hat{A}$ is an observable, any physical state is expressible as linear combinations of eigenstates of $\hat{A}$, which are in turn expressible as linear combinations of $\hat{A}$ and $\hat{B}$.– □

Fact 81. ***Natural Basis from Simultaneous Eigenstates***

Theorems 80 and 47, taken together, *almost* implies that **the simultaneous eigenstates of compatible observables can be used as a *natural* basis**. "Following Dirac," we will always assume that this is the case. The case of vectors of infinite lengths, associated with continuum spectra, makes the *mathematical proof* of this part quite difficult, if not undefined. What happens in reality is, though, that there are at least one different quantum number between two simultaneous eigenstates of two or more compatible observables so that the orthogonality is always guaranteed. Or, typically at least one of the observables has discrete quantum numbers [parity, angular momentum, etc.], and one ends up with a finite number of eigenvectors for that observable, which can be orthogonalized easily. In any case, when the orthogonality is guaranteed, then from Theorem 80 and (a general form of) Theorems 59,61 we get a natural basis. Well, if all this mathematical complication due to continuum eigenstates is a bit too much, then you can always trust it reasoning from the Numerical theorist point of view that *everything is finite,* in which case *everything is well* as discussed after theorem 47.

Definition 82. ***Conservation of an Observable Dynamical Variable (Very Important!!!)***

These are all equivalent statements for saying that an observable $\hat{A}$ is conserved. One definition leads to all other statements. The last statement is the direct consequence of the above theorem.

(1) In Heisenberg picture, $\frac{d\hat{A}_H}{dt} = 0$.
(2) $\frac{d}{dt}\langle \hat{A} \rangle = 0$ for any physical state $|\Psi\rangle$.
(3) $\frac{\partial \hat{A}}{\partial t} = 0$ and $[\hat{H}, \hat{A}] = 0$.
(4) $\frac{\partial \hat{A}}{\partial t} = 0$ and stationary states can be taken as eigenstates of $\hat{A}$.

Example 83. ***Parity Conservation***

Parity operator is defined as $\hat{P}\Psi(x) = \Psi(-x)$. Often $[\hat{H}, \hat{P}] = 0$, and this means that stationary states can be chosen to be parity eigenstates.

Example 84. ***Energy Conservation***

Taking $\hat{A} = \hat{H}$ and noting that any operator commutes with itself, it follows that energy is conserved as long as $\partial \hat{H}/\partial t = 0$, i.e. if Hamiltonian is not explicitly dependent on time. This fact is the most elegantly summarized by expressions such as "the homogeneity of time" or "any point in time being equivalent to any other point in time."

Example 85. ***Momentum Conservation***

The momentum operator is obviously not explicitly t dependent, and so for it to be conserved, we need only $[\hat{H}, \hat{p}] = 0$. This is for instance the case of the free particle problem, where $\hat{H} = \hat{p}^2/(2m)$.

9. Representations

State $|\Psi\rangle$ and wave function $\Psi(x)$ are different, as an arrow is different from its column vector representation. Strictly speaking, the wave function is given by $\Psi(x) = \langle x|\Psi\rangle$, where $|x\rangle$ is the position eigenstate, with *its* representation given by $\langle x'|x\rangle = \delta(x' - x)$. Then, for the same wave state, why not take $\langle p|\Psi\rangle$ for the momentum eigenstate $|p\rangle$ (Example 67)? The answer is yes, you can! This wave function is $\tilde{\Psi}(p)$ in the notation of Example 67. In Quantum Mechanics, canonically conjugate variables are incompatible. In Classical Mechanics, they are independent variables, but in Quantum Mechanics they are not. You can describe your states in either of the two conjugate variables, and there is in principle no superiority of one description to the other. When a state $|\Psi\rangle$ is turned into a list of numbers $\langle e|\Psi\rangle$ where $\langle e|$ is a basis bra, we say that the state is *represented* by the basis.

All different wave function *representations* should be treated on equal footings, and the state should be taken to be independent of wave functions. Due to the completeness of eigenstates of any observable, it is possible to use eigenstates of *any observable*, or, better still, *simultaneous eigenstates of any compatible observables*, as the basis to represent states.

One basic thing to note is that if states are represented by certain basis, then operators must be represented by the same basis. For instance, let us re-visit the problem of the free particle. $\hat{H} = \hat{p}^2/(2m)$ means that momentum eigenstates $|p\rangle$ are the best basis. In this basis, the Schrödinger equation simply reads

$$\frac{p^2}{2m}\tilde{\Psi}(p,t) = i\hbar\frac{\partial}{\partial t}\tilde{\Psi}(p,t)$$

where we used the fact that in the $|p\rangle$ basis, $\hat{p} = p$ (in this case $\hat{x}$ is a differential operator! $\hat{x} = i\hbar\partial/\partial p$). The solution is simple:

$$\tilde{\Psi}(p,t) = \exp\left(-i\frac{p^2}{2m\hbar}t\right)\tilde{\Psi}(p,0)$$

The momentum space wave function $\tilde{\Psi}(p,t)$ simply acquires a time dependent phase factor, and whatever wave packet is formed initially, the probability density in the momentum space $|\tilde{\Psi}(p,t)|^2$ is independent of time, which leads to the momentum conservation $\frac{d}{dt}\langle p\rangle = 0$ for any state. By choosing the right basis, it is much easier to figure these things out. Another easy thing to figure out is the basic fact that $\hat{p}$ is Hermitian. In Example 30, we went some lengths to demonstrate that $\hat{p}$ is Hermitian. If we used the momentum space representation of the Schrödinger equation, all we had to note was to realize that $\hat{p}$ is just a real number, i.e. a momentum eigenvalue p, so *of course* $\hat{p}$ is Hermitian.