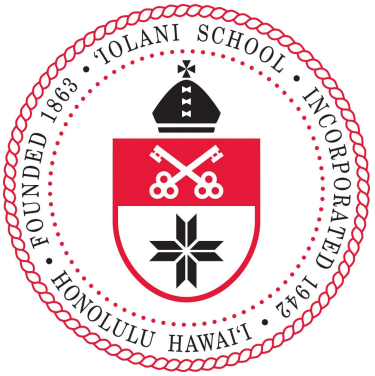
ʻIolani School
Physics Club



FUNDAMENTAL QUANTUM MECHANICS

Raphael Esquivel

Honolulu
2024

Preface

This lecture series, organized for the 'Iolani Physics Club, is the primary focus for the 2024/25 academic year as far as fun meetings go. We will explore quantum mechanics, a topic that garnered the highest interest among members last year. Our approach will emulate the rigor and depth of a collegiate course.

Instead of resorting to vast simplifications of popular topics (which can be readily found in YouTube), we will thoroughly examine key topics such as the Schrödinger equation and the Heisenberg uncertainty principle. Our aim is to uncover the intrinsic beauty in the technical and mathematical foundations of quantum mechanics. While it is unrealistic to cover an entire course in this informal setting, the objective is to introduce you to quantum formalism and various intriguing phenomena, providing a comprehensive overview of this expansive field.

Due to the breadth of topics falling under the domain of quantum mechanics, we do not follow a specific textbook, although you can find references for each chapter in the last section. Instead, this curriculum was specifically designed to cover the most popular topics as derived from the fundamental postulates in a short amount of time. Of course, this is not the most rigorous way to do things, but it is in the interest of time that we take such an unconventional approach. Nevertheless, the topics are presented in a way that is continuous and relies on minimal prior knowledge.

Contents

1 Introduction: The Double Slit Experiment **3**
1.1 Observations . . . 3
1.2 Theory . . . 5
1.3 Quantum particles . . . 6
1.4 Wave-particle duality . . . 7

2 Abstract Vector Spaces **9**
2.1 Ket space and underlying linear algebra . . . 9
2.2 Bra space and inner products . . . 11
2.3 Base kets . . . 15
2.4 Matrix representations . . . 17

3 First postulates **18**
3.1 Measurements . . . 18
3.2 Compatible observables . . . 19
3.3 The uncertainty relation . . . 22

4 Field Trip to Classical Mechanics **24**
4.1 Euler-Lagrange equation: a proof that $y = mx + b$ . . . 24
4.2 Principle of least action . . . 25
4.3 Defining the Lagrangian . . . 26
4.4 Hamiltonian mechanics . . . 27
4.5 Poisson brackets . . . 30
4.6 Canonical transformations . . . 30

5 Time **34**
5.1 Change of basis . . . 34
5.2 Generalization to continuous spectra . . . 35
5.3 Position . . . 36
5.4 Translation . . . 37
5.5 The classical analogy: the Heisenberg picture . . . 39
5.6 Ehrenfest's theorem: Newton's second law of quantum mechanics 40
5.7 The Schrodinger picture . . . 41

6 Wavefunctions **42**
6.1 The position-space wavefunction . . . 42
6.2 The momentum operator in the position basis . . . 43
6.3 The momentum-space wavefunction . . . 44

7 Bibliography **47**

1 Introduction: The Double Slit Experiment

In classical mechanics, Newton's second law tells us that

$$F = ma = m\frac{d^2x}{dt^2} \tag{1}$$

which gives us an equation of motion that can be solved for the position $x(t)$ as a function of the time t. That gives us the trajectory of the particle and tells us precisely where it will be at any time.

Quantum mechanics is fundamentally different from this. If we're told that a quantum particle was found at a position x_i at the initial time t_i, then all we can predict for when we measure its position later is the *probability* that we will find it at position x_f. The particle generally does not even have a well-defined position until we go measure it. This means that, unlike in classical mechanics, a quantum particle does not follow a well-defined trajectory in getting from one point to another.

1.1 Observations

Imagine we were to shoot lots of "classical" particles (such as, for example, BB gun pellets) at a tiny hole in a wall. Some pellets will pass through the hole. We place a backstop behind the wall, so that when the pellets hit they get stuck, and we can count how many were hit and where. If we were to make a histogram recording where each of the pellets hits the backstop, we would expect to get a distribution.

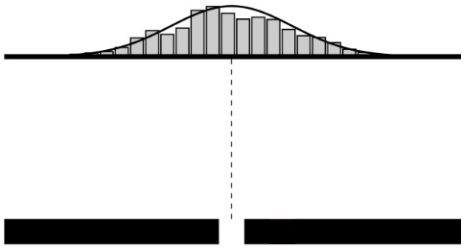


Figure 1: Distribution of classical pellets behind a single hole

If we have two holes, the total distribution will look like the sum of those two individual distributions, because each pellet either goes through one hole or the other.

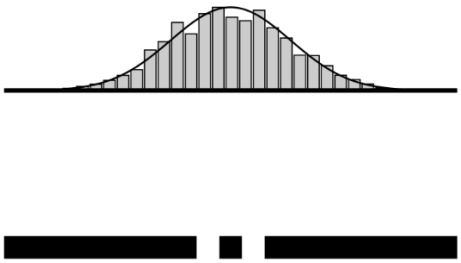


Figure 2: Distribution of classical pellets behind two holes

Now, consider the scenario where we have a classical wave (such as light from a laser) going straight into a wall with a small hole, just like the BB gun. On the outgoing screen, we will look at the intensity of the light.

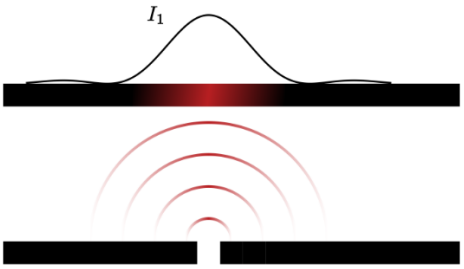


Figure 3: Wave propagates through a single hole.

However, unlike the classical particle, when we shoot the classical wave through two holes, there are two separate waves emanating from each hole that **interfere** with each other, creating what we call an **interference pattern** (or **diffraction pattern**) on the back screen.

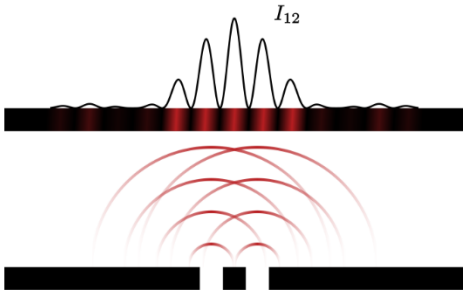


Figure 4: Wave propagates through two holes, creating an interference pattern.

The bright spots appear when the waves interfere *constructively*, meaning they add together to make a bigger wave, and the dark spots are where the interference is *destructive* and the two waves have canceled each other out. Let's look at the math behind this.

1.2 Theory

We'll suppose that the gaps in the barrier are very narrow. An incoming plane wave ψ_{inc} strikes the barrier and breaks up into two spherical waves ψ_1 and ψ_2. The total outgoing wave is the sum of those two spherical waves. Because the holes in the barrier are separated a little bit, the two waves have to travel slightly different distances to arrive at a given point on the observation screen– call them r_1 and r_2. In the middle of the screen, $r_1 = r_2$, and the waves are perfectly in sync, combining constructively to make a bright spot. However, away from the middle, they won't necessarily be synced. Any time the quantity $r_1 - r_2$ is a multiple of the wavelength λ, the waves will sync up again and we'll get another bright spot. Mathematically, we can write the incoming wave as

$$\psi_{inc}(y,t) = A\cos\phi(y,t) \tag{2}$$

Here, A is the amplitude of the wave and $\phi(y,t)$ is the phase, which is a function of the vertical coordinate y and the time t given by

$$\phi(y,t) = \frac{2\pi}{\lambda}(y - vt) \tag{3}$$

This is the standard form for the phase of a wave of wavelength λ that travels in the y direction at a speed v. λ tells us how far apart the successive peaks are, and A tells us how tall they are.

Writing everything in terms of cosines, however, can be a bit of a pain; it is much more convenient to write down a *complex* wave,

$$\psi_{inc} = A\exp\left[i\phi(y,t)\right] \tag{4}$$

In our notation, $\exp[x]$ is the same as e^x. The idea here is that we're using Euler's identity

$$\exp[i\phi] = \cos\phi + i\sin\phi \tag{5}$$

We can think of our original cosine wave as the real part of this complex wave:

$$A\cos\phi = \mathrm{Re}\left(A\exp[i\phi]\right) \tag{6}$$

On the opposite side of the barrier, we can write the outgoing waves in a similar way. Because they are spherical waves, they don't vary with the vertical coordinate y but rather with the radial distances:[1]

$$\psi_1(r_1,t) \propto A\exp\left[i\phi(r_1,t)\right], \quad \psi_2(r_2,t) \propto A\exp\left[i\phi(r_2,t)\right] \tag{7}$$

The total outgoing wave is given by the sum of these two contributions. The intensity is the square of the modulus of that wave:

$$I_{12} = |\psi_1 + \psi_2|^2 \tag{8}$$

The algebra works out to

$$I_{12} = 4A^2\cos^2\left(\frac{\pi}{\lambda}(r_1 - r_2)\right) \tag{9}$$

[1] I wrote $\propto$, meaning "proportional to", rather than $=$ because the spherical waves fall off with a factor of $\frac{1}{r}$ but that's not relevant for the shape of the interference pattern in the limit that we're interested in.

Let's write s for the separation between the holes and d for the distance between the barrier and the observation screen. Let's also label the points on the screen by a coordinate x, measured from the center of the screen. With a little geometry, we can use the assumptions that $d >> s$ and $d >> x$ to make the approximation

$$r_1 - r_2 \approx \frac{sx}{d} \tag{10}$$

The intensity pattern becomes $I_{12} \approx 4A^2 \cos^2\left(\frac{\pi sx}{\lambda d}\right)$ which is an oscillating pattern of bright and dark fringes along the screen, as observed.

1.3 Quantum particles

Let's run this experiment with the double slit one last time–this time, shooting tiny, quantum particles at the barrier, like electrons. With only one hole open, the distribution looks similar to that of the BB gun pellets. However, if we open two holes, it makes an interference pattern, just like what the light waves did.

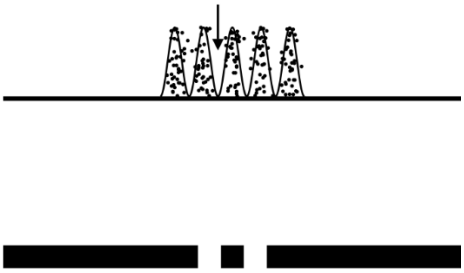


Figure 5: Double slit experiment with quantum particles, creating an interference pattern.

Notice the spot indicated by the arrow. If the electrons acted like classical particles, there would be electrons found here! Instead, it becomes a dead-zone, at the bottom of an interference pattern that we would have expected for waves emanating from these two holes. This completely contradicts our classical intuition about how particle-like objects should behave. We could try to come up with classical mechanisms to explain what's going on; for example, you might wonder if the many electrons passing through the apparatus at any given moment are colliding with one another, and it's those collisions that somehow conspire to produce this intricate interference pattern. However, if even if we fire one particle at a time, we see that over time they build up the same distribution of fringes, matching the interference pattern.

Evidently, each electron is somehow probing both slits at once, and interfering with itself. In particular, an electron does not have a well-defined trajectory in the sense we're familiar with from classical mechanics. The fact that a quantum particle doesn't follow a well-defined trajectory means we can't say for certain where the particle will wind up hitting the detector, even if we know everything there is to know about how the particle was fired at the barrier. All we can predict is the probability we'll find the particle, according to the shape of the interference curve.

1.4 Wave-particle duality

We've learned two key things from the double-slit experiment. First, electrons behave like particles in the sense that we're accustomed to, insofar as they always hit the detector in discrete lumps, but an electron can also behave like a wave in that it exhibits interference. The same actually goes by our laser beam experiment as well. If we turn the intensity of the light way down, we'll find that what looked like a continuous light wave hitting our detector was actually a stream of individual particles called photons. If we send one photon through at a time, we'll once again gradually build up the same interference pattern. The fact that a quantum particle can exhibit both of these sorts of properties–sometimes what we classically think of as a particle and sometimes like a classical wave–is what **wave-particle duality** refers to. However, an electron is neither a classical particle nor a classical wave–it's a quantum particle.

In the light of the interference pattern, we introduce a new quanitity for each particle that we fire at the barrier, called its **wavefunction** $\psi(\vec{r}, t)$. Then just like we had for classical waves, after the incoming wavefunction strikes the barrier, it diffracts, and two spherical waves ψ_1 and ψ_2 emerge from the holes on the opposite side. The outgoing wavefunction is their sum, and just like before, the modulus of that total wave squared takes the shape of the interference curve.

That's the wave side of wave-particle duality. Then again, we always detect a localized lump on the backstop. What the interference curve is telling us in this case is not the intensity or brightness of some laser–it's the *probability* of where we'll find the particle when it hits the backstop. After we've fired many quantum particles at the barrier, we therefore find that the spots where they end up are distributed according to that probability curve. The wavelength λ that we assign to the particle when we write down its wavefunction can be determined experimentally by measuring the separation between the pekas of the interference pattern, which is proportional to λ. The result is called the **de Broglie wavelength**, and if we fired off particles with momentum p, it's given by

$$\lambda = 2\pi\frac{\hbar}{p} \tag{11}$$

where $\hbar$ is **Planck's constant**.

In other words, the incoming electrons are described the wavefunction

$$\psi_{inc}(y,t) = A\exp\left[i\phi(y,t)\right] \tag{12}$$

where $\phi(y,t) = \frac{2\pi}{\lambda}\left(y - \frac{v}{2}t\right)$, similar to a complex classical wave, except that λ is now given by de Broglie's formula (and v has been replaced by $v/2$, for reasons related to the difference between the classical wave equation and the quantum Schrodinger equation).

In summary, each quantum particle is described by a wavefunction that explains the distribution of particles we observe on the backstop of the double-slit experiment. $|\psi(y,t)|^2$ is interpreted as the probability amplitude of the particle's presence. Since the possible positions of the particle form a continuum, the probability $dP(y,t)$ of the particle being, at time t, in a spatial element dy situated at the point y must be proportional to dy. It is therefore infinitesimal, and the modulus squared of the wavefunction is interpreted as the corresponding *probability density*, with

$$dP(y,t) = C\left|\psi(y,t)\right|^2 dy \tag{13}$$

where C is a normalization constant. Of course, we can generalize this to three dimensions, where we take $r = (x, y, z)$ and $d^3r = dxdydz$, yielding $dP(r,t) = C\left|\psi(r,t)\right|^2 d^3r$. For a system composed of only one particle, the total probability of finding the particle anywhere in space at time t is equal to 1:

$$\int dP(r,t) = 1 \tag{14}$$

That is, if we integrate over all the infinitesimal dP at every point in space and add up all those infinitesimal probabilities, there's a 100% chance that the particle will be *somewhere*. We conclude that the wavefunction must be square-integrable; that is,

$$\int \left|\psi(r,t)\right|^2 d^3r \quad \text{is finite.} \tag{15}$$

The normalization constant C is given by

$$\frac{1}{C} = \int \left|\psi(r,t)\right|^2 d^3r \tag{16}$$

2 Abstract Vector Spaces

Staying loyal to my goal to keep this self-contained, the prerequisite math to this is mostly just a general understanding of "normal" matrices and vectors, and maybe some intuitive understanding of other general math concepts. Otherwise, *technically* you don't need anything that you wouldn't learn in high school to *understand* the following content. Of course, it's not easy math–after all, this is quantum mechanics! However, the underlying mathematical framework is essential to understanding quantum mechanics, and cannot be overlooked.

2.1 Ket space and underlying linear algebra

You may be familiar with the notion of a **Euclidean vector**; a geometric object with magnitude and direction. However, we can generalize these to *abstract vector spaces*. For this, lets do some math![2]

Definition 2.1. *A **vector space** $\mathbb{V}$ is a collection of objects $|1\rangle, |2\rangle, ..., |V\rangle, ..., |W\rangle, ...$, called **vectors**, for which there exists*

1. *A definite rule for forming the vector sum, denoted $|V\rangle + |W\rangle$, with the following features:*
   - *Addition is commutative: $|V\rangle + |W\rangle = |W\rangle + |V\rangle$*
   - *Addition is associative: $|V\rangle + (|W\rangle + |Z\rangle) = (|V\rangle + |W\rangle) + |Z\rangle$*
   - *There exists a **null vector** $|0\rangle$ obeying $|V\rangle + |0\rangle = |V\rangle$*
   - *For every vector $|V\rangle$ there exists an inverse under addition, $|-V\rangle$ such that $|V\rangle + |-V\rangle = |0\rangle$*
2. *A definite rule for multiplication by scalars $a, b, ...$ denoted $a|V\rangle$ with the following features:*
   - *Scalar multiplication is distributive in the vectors: $a(|V\rangle + |W\rangle) = a|V\rangle + a|W\rangle$*
   - *Scalar multiplication is distributive in the scalars: $(a + b)|V\rangle = a|V\rangle + b|V\rangle$*

*Furthermore, the result of these operations must be another element of the space, a feature called **closure**.*

Other properties that follow from the definition are:

- $|0\rangle$ is unique; i.e., if $|0'\rangle$ has all the properties of $|0\rangle$, then $|0\rangle = |0'\rangle$
- $0|V\rangle = |0\rangle$
- $|-V\rangle = -|V\rangle$
- $|-V\rangle$ is the unique additive inverse of $|V\rangle$

The way to remember all of these is just to do what comes naturally.

[2] I've made some simplifications to keep things relevant to physics.

Definition 2.2. *A set is called a field if on it are defined two operations: addition and multiplication, both commutative and associative and also related by the distributive law* $(a + b)c = ac + bc$. *There exists an element* 0 *for which an element of the set added by 0 returns itself, and an element* 1 *for which an element multiplied by* 1 *returns itself. For every element of the set there also exist additive and multiplicative inverses.*

Trivial examples of fields include the set of **real numbers** $\mathbb{R}$ and the set of **complex numbers** $\mathbb{C}$. The numbers $a, b, ...$ of definition 2.1 are called the field *over which the vector space is defined.*

Definition 2.3. *A* ***complex vector space*** *is a vector space* $\mathbb{V}$ *for which the field over which it is defined is the set of complex numbers* $\mathbb{C}$.

Observe that we are using the new symbol $|V\rangle$ to denote a generic vector. This object is called a **ket** and this nomenclature is from **Dirac notation**–one of the most widely used mathematical "languages" for quantum mechanics. We purposely refrain from using the symbol $\vec{V}$ to denote vectors to avoid the limited concept of the vector as an arrow. The set of Euclidean vectors forms a vector space, but not all vector spaces are made of Euclidean vectors; they simply have to follow the axioms specified in the definition.

Definition 2.4. *A set of vectors* $|1\rangle, |2\rangle, ... |n\rangle$ *is said to be* ***linearly independent*** *if*

$$\sum_{i=1}^{n} a_i |i\rangle = |0\rangle \tag{17}$$

implies all $a_i = 0$. *In other words, no vector in the set can be written as a linear combination of others.*

Definition 2.5. *A vector space has* ***dimension*** n *if it can accommodate a maximum of n linearly independent vectors. Alternatively, the dimension is the number of elements in a minimal generating set for the space; that is, every vector in the space can be described as a linear combination of the elements of this set of vectors in the vector space, and the elements of the set are linearly independent.*[3]

If any vector in the vector space can be described as a linear combination of a set of linearly independent vectors, that set is called a **basis**. We can write

$$|V\rangle = \sum_{i=1}^{n} v_i |i\rangle \tag{18}$$

where the vectors $|i\rangle$ form a basis. The coefficients of expansion in v_i of a vector in terms of a linearly independent basis are called the **components** of the vector *in that basis.*

We consider a complex vector space whose dimensionality is specified according to the nature of a physical system under consideration. In quantum mechanics, a physical state is represented by a **state vector** $|\alpha\rangle$ in a complex vector space.[4] This state ket is postulated to contain complete information

[3] Note that vector spaces in quantum mechanics can possibly be infinite-dimensional!

[4] As mentioned, a vector in the context of quantum mechanics is called a ket. We will now begin using this nomenclature.

about the physical state; everything we are allowed to ask about the state is contained in the ket.

An **observable**–a physical property or quantity that can be measured–can be represented by an **operator**, such as A, in the vector space in question. Generally, an operator acts on a ket from the left,

$$A \cdot (|\alpha\rangle) = A\,|\alpha\rangle \tag{19}$$

which is yet another ket. In general, $A\,|\alpha\rangle$ is *not* a constant times $|\alpha\rangle$. However, there are particular kets of importance, known as **eigenkets** of operator A, denoted by

$$|a'\rangle\,, |a''\rangle\,, |a'''\rangle\,, ... \tag{20}$$

with the property

$$A\,|a'\rangle = a'\,|a'\rangle \tag{21}$$

and so on, where $a', a'', ...$ are just numbers. Notice that applying A to an eigenket just reproduces the same ket apart from a multiplicative number. The set of numbers $\{a', a'', a''', ...\}$, more compactly denoted $\{a'\}$, is called the set of **eigenvalues** of operator A. When it becomes necessary to order eigenvalues in a specific manner, $\{a^{(1)}, a^{(2)}, a^{(3)}, ...\}$ can be used in place of $\{a', a'', a''', ...\}$.

Suppose there are two (or more) linearly independent eigenkets of A having the same eigenvalue; then the eigenvalues of the two eigenkets are said to be **degenerate**. In such a case the notation $|a'\rangle$ that labels the eigenket by its eigenvalue alone does not give a complete description.

The physical state corresponding to an eigenket is called an **eigenstate**.

2.2 Bra space and inner products

In linear algebra, a **map** is like a function. If you have a map L from a set X to a set Y, it assigns to each element of X an element of Y. We denote this by $L : X \to Y$.

Definition 2.6. *A **linear map** is a mapping between two vector spaces that preserves the operations of vector addition and scalar multiplication.*

In other words, a linear map $L : \mathbb{V} \to \mathbb{U}$ must satisfy

$$\begin{aligned} L\,(|V\rangle + |W\rangle) &= L\,|V\rangle + L\,|W\rangle \\ L\,(c\,|V\rangle) &= cL\,|V\rangle \end{aligned} \tag{22}$$

where V and W are arbitrary elements of $\mathbb{V}$.

Definition 2.7. *A **linear functional** on a vector space $\mathbb{V}$ over a field $\mathbb{K}$ is any linear map $L : \mathbb{V} \to \mathbb{K}$.*

Consider a linear functional L_x that returns the x-component of a vector in $\mathbb{R}^2$.[5] If we write the vector in column vector notation, this looks a little like

$$L_x \begin{bmatrix} 3 \\ -1 \end{bmatrix} = 3 \tag{23}$$

[5] In our language, $\mathbb{R}^2$ is a real vector space of dimensionality 2; that is, the field over which it is defined is the set of real numbers $\mathbb{R}$. This is just like any two-dimensional space we usually work with.

Formally, this maps an element of $\mathbb{R}^2$ to an element of $\mathbb{R}^1$; that is, $L_x : \mathbb{R}^2 \to \mathbb{R}^1$. The matrix that defines this linear functional is[6]

$$L_x = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{24}$$

If we were to look at a general linear functional $L : \mathbb{R}^2 \to \mathbb{R}^1$, we see that, by the rules of matrix multiplication, L must be a 1x2 row matrix. This means that the set of all linear functional in $\mathbb{R}^2$ consists of the set of all 1x2 row matrices. Much like column vectors, this set of row matrices forms its own vector space. This vector space is called the *dual space*. This has been specific to $\mathbb{R}^2$, but the dual space is a very useful concept in abstract vector spaces. Remember that, since we work in complex vector spaces, our linear functionals return an element of $\mathbb{C}$ in our context.

Definition 2.8. *Given a vector space $\mathbb{V}$, the **dual space** $\mathbb{V}^*$ is the vector space of all linear functionals in $\mathbb{V}$.*

Let $\mathbb{V}$ be an n-dimensional vector space, with the basis $\{|1\rangle, |2\rangle, ..., |n\rangle\}$. Define n linear functionals $\{L_1, L_2, ..., L_n\}$ by

$$L_i(e_j) = \delta_{ij} \tag{25}$$

where we have used the **Kronecker delta**

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{26}$$

The linear functionals $\{L_1, ..., L_n\}$ form a basis of $\mathbb{V}^*$, called the **dual basis** to $\{|i\rangle\}$. Note that these equations uniquely define each linear functional L_i, since their values are assigned on each basis vector e_j in turn.

We now introduce the notion of a **bra space**, a vector space "dual to" the ket space. We postulate that corresponding to every ket $|\alpha\rangle$ there exists a bra, denoted $\langle\alpha|$, in this dual, or bra, space. The bra space is spanned by eigenbras $\{\langle a'|\}$ which correspond to the eigenkets $\{|a'\rangle\}$. There is a one-to-one correspondence between a ket space and a bra space. Roughly speaking, we can regard the bra space as some kind of mirror image of the ket space.

The bra dual to $c|\alpha\rangle$ is postulated to be $c^* \langle\alpha|$, where the asterisk denotes complex conjugation, *not* $c\langle\alpha|$, which is a very important point.

Definition 2.9. *The **inner product** of a bra and a ket is defined to be*

$$\langle\beta|\alpha\rangle = (\langle\beta|) \cdot (|\alpha\rangle) \tag{27}$$

In general, the product is a complex number. Notice that in forming an inner product we always take one vector from the bra space and one vector from the ket space.

We postulate two fundamental properties of inner products. First,

$$\langle\beta|\alpha\rangle = \langle\alpha|\beta\rangle^* \tag{28}$$

In other words, $\langle\beta|\alpha\rangle$ and $\langle\alpha|\beta\rangle$ are complex conjugates of each other. The second postulate on the inner products is

$$\langle\alpha|\alpha\rangle \geq 0 \tag{29}$$

[6] Try it for yourself!

where equality holds if and only if $|\alpha\rangle = |0\rangle$. This is known as the postulate of **positive definite metric**.

Definition 2.10. *Two kets $|\alpha\rangle$ and $|\beta\rangle$ are said to be **orthogonal**[7] if*

$$\langle\alpha|\beta\rangle = 0 \tag{30}$$

Given a ket which is not a null ket, we can form a **normalized ket** $|\tilde{\alpha}\rangle$ where

$$|\tilde{\alpha}\rangle = \left(\frac{1}{\sqrt{\langle\alpha|\alpha\rangle}}\right)|\alpha\rangle \tag{31}$$

with the property

$$\langle\tilde{\alpha}|\tilde{\alpha}\rangle = 1 \tag{32}$$

Quite generally, $\sqrt{\langle\alpha|\alpha\rangle}$ is known as the **norm** of $|\alpha\rangle$, analogous to the magnitude of a vector in Euclidean vector space. Because $|\alpha\rangle$ and $c\,|\alpha\rangle$ represent the same physical state, we might as well require that the kets we use for physical states be normalized.[8]

As we remarked earlier, observables are to be represented by operators that can act on kets. We can consider a more general class of operators that can act on kets; they will be denoted by X, Y, and so forth, while A, B, and so on will be used for a restrictive class of operators that correspond to observables. Two operators X and Y are said to be equal if

$$X\,|\alpha\rangle = Y\,|\alpha\rangle \tag{33}$$

for an arbitrary ket in the ket space in question. Operator X is said to be the **null operator** if, for any arbitrary ket $|\alpha\rangle$, we have

$$X\,|\alpha\rangle = 0 \tag{34}$$

Operators can be added. Addition operations are commutative and associative. The operators that we will consider are all linear, that is,

$$X\left(c_\alpha\,|\alpha\rangle + c_\beta\,|\beta\rangle\right) = c_\alpha X\,|\alpha\rangle + c_\beta X\,|\beta\rangle \tag{35}$$

An operator X always acts on a bra from the right side:

$$(\langle\alpha|)\cdot X = \langle\alpha|\,X \tag{36}$$

and the resulting product is another bra. The ket $X\,|\alpha\rangle$ and the bra $\langle\alpha|\,X$ are in general not dual to each other. We define the symbol $X^\dagger$, known as the **Hermitian adjoint** of X, to be the operator for which $X\,|\alpha\rangle$ is dual to $\langle\alpha|\,X$. In other words, let's represent $X\,|\alpha\rangle$ by $|X\alpha\rangle$; the Hermitian adjoint $X^\dagger$ satisfies $\langle\beta|X\alpha\rangle = \langle X^\dagger\beta|\alpha\rangle$. An operator X is said to be **Hermitian** if $X = X^\dagger$.

Operators can be multiplied. Multiplication operations are, in general, non-commutative; that is, generally, $XY \neq YX$. However, operators are associative:

$$X(YZ) = (XY)Z = XYZ \tag{37}$$

[7] That is, perpendicular; remember, we can have as many mutually perpendicular vectors as we want to, because our ket space can be many dimensions!

[8] For eigenkets of observables with continuous spectra, different normalization conventions will be used.

We also have

$$X\left(Y\left|\alpha\right\rangle\right)=\left(XY\right)\left|\alpha\right\rangle=XY\left|\alpha\right\rangle,\quad\left(\left\langle\beta\right|X\right)Y=\left\langle\beta\right|\left(XY\right)=\left\langle\beta\right|XY \tag{38}$$

Notice that

$$\left(XY\right)^{\dagger}=Y^{\dagger}X^{\dagger} \tag{39}$$

because $XY\left|\alpha\right\rangle=X\left(Y\left|\alpha\right\rangle\right)$ is dual to $\left(\left\langle\alpha\right|Y^{\dagger}\right)X^{\dagger}=\left\langle\alpha\right|Y^{\dagger}X^{\dagger}$. This is a very important property.

Definition 2.11. *The **outer product** of a ket and a bra is defined to be*

$$\left(\left|\beta\right\rangle\right)\cdot\left(\left\langle\alpha\right|\right)=\left|\beta\right\rangle\left\langle\alpha\right| \tag{40}$$

As is clear from Equation 37, multiplication operations among operators are associative; the associative property is actually postulated to hold quite generally as long as we are dealing with "legal" multiplications among kets, bras, and operators. Dirac calls this important postulate the **associative axiom of multiplication**. By this axiom, if we consider an outer product acting on a ket,

$$\left(\left|\beta\right\rangle\left\langle\alpha\right|\right)\cdot\left|\gamma\right\rangle \tag{41}$$

this equals

$$\left|\beta\right\rangle\cdot\left(\left\langle\alpha|\gamma\right\rangle\right) \tag{42}$$

Of course, $\left\langle\alpha|\gamma\right\rangle$ is just a number, which, multiplied by the ket $\left|\beta\right\rangle$, returns just another ket; in other words, an outer product can be regarded as an operator, for when it acts upon a ket, it returns a ket. Notice that the operator $\left|\beta\right\rangle\left\langle\alpha\right|$ rotates $\left|\gamma\right\rangle$ into the direction of $\left|\beta\right\rangle$. It is easy to see that if

$$X=\left|\beta\right\rangle\left\langle\alpha\right| \tag{43}$$

then

$$X^{\dagger}=\left|\alpha\right\rangle\left\langle\beta\right| \tag{44}$$

There are also "illegal products". For example, operators must stand on the left of a ket or right of a bra: $\left|\alpha\right\rangle X$ and $X\left\langle\alpha\right|$ are examples of illegal products. They are neither kets, nor bras, nor operators; they are simply nonsensical. Products like $\left|\alpha\right\rangle\left|\beta\right\rangle$ and $\left\langle\alpha\right|\left\langle\beta\right|$ are also illegal when $\left|\alpha\right\rangle$ and $\left|\beta\right\rangle$ ($\left\langle\alpha\right|$ and $\left\langle\beta\right|$) are ket (bra) vectors belonging to the same ket (bra) space. By associativity, we could write the expression in Equation 41 or 49 as $\left|\beta\right\rangle\left\langle\alpha|\gamma\right\rangle$; however, if we had $\left(\left\langle\alpha|\gamma\right\rangle\right)\cdot\left|\beta\right\rangle$ we cannot afford to omit the dot and brackets because the resulting expression would look illegal.

Another important illustration of the associative axiom is that

$$\left(\left\langle\beta\right|\right)\cdot\left(X\left|\alpha\right\rangle\right)=\left(\left\langle\beta\right|X\right)\cdot\left(\left|\alpha\right\rangle\right) \tag{45}$$

Because the two sides are equal, we might as well use the more compact notation

$$\left\langle\beta|X|\alpha\right\rangle \tag{46}$$

Recall that $\left\langle\alpha\right|X^{\dagger}$ is the bra dual to $X\left|\alpha\right\rangle$, so

$$\begin{aligned}\left\langle\beta|X|\alpha\right\rangle&=\left\langle\beta\right|\cdot\left(X\left|\alpha\right\rangle\right)\\&=\left[\left(\left\langle\alpha\right|X^{\dagger}\right)\cdot\left|\beta\right\rangle\right]^{*}\\&=\left\langle\alpha|X^{\dagger}|\beta\right\rangle^{*}\end{aligned} \tag{47}$$

For *Hermitian* X we have

$$\left\langle\beta|X|\alpha\right\rangle=\left\langle\alpha|X|\beta\right\rangle^{*} \tag{48}$$

2.3 Base kets

Theorem 2.12. *The eigenvalues of a Hermitian operator A are real; the eigenkets of A corresponding to different eigenvalues are orthogonal.*

I'll include the proof, as it utilizes a lot of our concepts. However, *for theorems for which the proof is not provided, you are encouraged to try to prove them yourself.*

First, recall that

$$A\left|a'\right\rangle = a'\left|a'\right\rangle \tag{49}$$

Because A is Hermitian,

$$\left\langle a''\right| A = a''^{*}\left\langle a''\right| \tag{50}$$

where $a', a''...$ are eigenvalues of A. If we multiply both sides of Equation 49 by $\langle a''|$ on the left,[9]

$$\begin{aligned}\langle a''|A|a'\rangle &= \langle a''|a'|a'\rangle \\ &= a'\langle a''|a'\rangle\end{aligned} \tag{51}$$

and both sides of 50 by $|a'\rangle$ on the right:

$$\langle a''|A|a'\rangle = a''^{*}\langle a''|a'\rangle \tag{52}$$

we obtain

$$(a' - a''^{*})\langle a''|a'\rangle = 0 \tag{53}$$

If we choose a' and a'' to be the same, then we deduce the reality condition

$$a' = a'^{*} \tag{54}$$

where we have used the fact that $|a'\rangle$ is not a null ket. If we assume them to be different, by the reality condition the difference $a' - a''^{*}$ is equal to $a' - a''$, which cannot vanish by assumption. The inner product must then vanish, proving the orthogonality property.

We expect on physical grounds that an observable has real eigenvalues. The theorem just proved guarantees the reality of eigenvalues whenever the operator is Hermitian.

It is conventional to normalize $|a'\rangle$ so the $\{|a'\rangle\}$ form an orthonormal set:

$$\langle a''|a'\rangle = \delta_{a''a'} \tag{55}$$

An arbitrary ket in the ket space can be expanded in terms of eigenkets A; in other words, the eigenkets of A are to be used as base kets in much the same way as a set of mutually orthogonal unit vectors is used as base vectors in Euclidean space. Given an arbitrary ket $|\alpha\rangle$ in the ket space spanned by the eigenkets of A, let's try to expand it as follows:

$$\left|\alpha\right\rangle = \sum_{a'} c_{a'}\left|a'\right\rangle \tag{56}$$

[9]Note that we can bring the a' outside of the inner product because, being an eigenvalue, it's just a complex number.

Multiplying by $\langle a''|$ on the left and using the orthonormality property, we can immediately find the expansion coefficient,[10]

$$c_{a'} = \langle a'|\alpha\rangle \tag{57}$$

In other words, we have

$$|\alpha\rangle = \sum_{a'} |a'\rangle \langle a'|\alpha\rangle \tag{58}$$

which is analogous to an expansion of a vector V in real Euclidean space:

$$V = \sum_i \hat{e}_i \left(\hat{e}_i \cdot V\right) \tag{59}$$

where $\{\hat{e}_i\}$ form an orthogonal set of unit vectors. We now recall the associative axiom of multiplication. $\sum_{a'} |a'\rangle \langle a'|\alpha\rangle$ can be regarded either as the number $\langle a'|\alpha\rangle$ multiplying $|a'\rangle$ or, equivalently, as the operator $|a'\rangle \langle a'|$ acting on $|\alpha\rangle$. Because $|\alpha\rangle$ is arbitrary, the **completeness relation** (or **closure**) follows:

$$\sum_{a'} |a'\rangle \langle a'| = 1 \tag{60}$$

where the 1 on the right hand side is to be understood as the identity operator. Consider, for example, $\langle\alpha|\alpha\rangle$. By inserting the identity operator between $\langle\alpha|$ and $|\alpha\rangle$,

$$\begin{aligned} \langle\alpha|\alpha\rangle &= \langle\alpha| \cdot \left(\sum_{a'} |a'\rangle \langle a'|\right) \cdot |\alpha\rangle \\ &= \sum_{a'} |\langle a'|\alpha\rangle|^2 \end{aligned} \tag{61}$$

If $|\alpha\rangle$ is normalized, then the expansion coefficients in Equation 56 must satisfy

$$\sum_{a'} |c_{a'}|^2 = \sum_{a'} |\langle a'|\alpha\rangle|^2 = 1 \tag{62}$$

If we operate the outer product $|a'\rangle \langle a'|$ on $|\alpha\rangle$,

$$\left(|a'\rangle \langle a'|\right) \cdot |\alpha\rangle = |a'\rangle \langle a'|\alpha\rangle = c_{a'} |a'\rangle \tag{63}$$

We see that $|a'\rangle \langle a'|$ selects that portion of the ket $|\alpha\rangle$ parallel to $|a'\rangle$, so $|a'\rangle \langle a'|$ is known as the **projection operator** along the base ket $|a'\rangle$ and is denoted by $\Lambda_{a'}$:

$$\Lambda_{a'} \equiv |a'\rangle \langle a'| \tag{64}$$

The completeness relation can now be written as

$$\sum_{a'} \Lambda_{a'} = 1 \tag{65}$$

[10] The left side becomes $\langle a'|\alpha\rangle$, and, by the orthogonality condition, the right side vanishes except for when $a' = a''$, at which point it reduces to $\langle a''|c_{a'}|a'\rangle = c_{a'} \langle a''|a'\rangle = c_{a'}$.

2.4 Matrix representations

Let's represent an operator X by a square matrix. Using the completeness relation twice,

$$X = \sum_{a''} \sum_{a'} |a''\rangle \langle a''|X|a'\rangle \langle a'| \tag{66}$$

There are N^2 numbers of form $\langle a''|X|a'\rangle$, where N is the dimensionality of the ket space. We may arrange them into an $N \times N$ square matrix such that the eigenbras on the left of X are the row indices and the eigenkets on the right of X are the column indices; in other words,

$$X \doteq \begin{pmatrix} \langle a^{(1)}|X|a^{(1)}\rangle & \langle a^{(1)}|X|a^{(2)}\rangle & \cdots \\ \langle a^{(2)}|X|a^{(1)}\rangle & \langle a^{(2)}|X|a^{(2)}\rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{67}$$

where we use the symbol $\doteq$, meaning "is represented by", because the particular form of a matrix representation depends on the particular choice of base kets used. By Equation 47,

$$\langle a''|X|a'\rangle = \left\langle a'|X^\dagger|a''\right\rangle^* \tag{68}$$

At last, the Hermitian adjoint operation has been related to the concept of a *complex conjugate transposed*. That is, if you were to represent an operator as a matrix, you can get its Hermitian adjoint by transposing it and taking the complex conjugate of each of its elements.

3 First postulates

3.1 Measurements

"A measurement always causes the system to jump into an eigenstate of the dynamical variable that is being measured" (P. A. M. Dirac, 1958)

You may be familiar with this statement. In layman's terms, it's like how Schrodinger's cat is in a superposition of both aliveness and death before the box is open. Before the measurement of an observable A is made, the system is assumed to be represented by some linear combination

$$|\alpha\rangle = \sum_{a'} c_{a'} |a'\rangle = \sum_{a'} |a'\rangle \langle a'|\alpha\rangle \tag{69}$$

When the measurement is performed, the system is "thrown into" one of the eigenstates, say $|a'\rangle$ of observable A. In other words, $|\alpha\rangle \to |a'\rangle$. A measurement usually changes the state; the only exception is when the state is already in one of the eigenstates of the observable being measured, in which case $|a'\rangle \to |a'\rangle$ with certainty. When the measurement causes $|\alpha\rangle$ to change into $|a'\rangle$, it is said that A is measured to be a'. In this sense, the result of a measurement yields one of the eigenvalues of the observable being measured.

We do not know in advance into which of the various eigenstates the system will be thrown into. We do postulate, however, that the probability of jumping into some particular $|a'\rangle$ is given by

$$\text{Probability for } a' \ = |\langle a'|\alpha\rangle|^2 \tag{70}$$

provided that $|\alpha\rangle$ is normalized. This probabilistic interpretation for the squared inner product is one of the fundamental postulates of quantum mechanics, so it cannot be proven.

Definition 3.1. *The **expectation value** of A taken with respect to state $|\alpha\rangle$ is*

$$\langle A\rangle \equiv \langle \alpha|A|\alpha\rangle \tag{71}$$

To make sure that we are referring to state $|\alpha\rangle$, the notation $\langle A\rangle_\alpha$ is sometimes used.

This definition agrees with our intuitive notion of average measured value because it can be written as

$$\begin{aligned} \langle A\rangle &= \sum_{a'} \sum_{a''} \langle \alpha|a''\rangle \langle a''|A|a'\rangle \langle a'|\alpha\rangle \\ &= \sum_{a'} a' |\langle a'|\alpha\rangle|^2 \end{aligned} \tag{72}$$

wherein, in the second expression, a' represents a measured value, and $|\langle a'|\alpha\rangle|^2$ represents the probability for obtaining a'.

To clarify further the meaning of measurements in qantum mechanics we introduce the notion of a **selective measurement** or *filtration*. We imagine a measurement process with a device that selects only one of the eigenkets of A, say $|a'\rangle$, and rejects all others. Mathematically we can say that such a selective measurement amounts to applying the projection operator $\Lambda_{a'}$ to $|\alpha\rangle$:

$$\Lambda_{a'} |\alpha\rangle = |a'\rangle \langle a'|\alpha\rangle \tag{73}$$

3.2 Compatible observables

Definition 3.2. *The **commutator** operation $[,]$ between two matrices A and B is defined by*

$$[A, B] \equiv AB - BA \tag{74}$$

*and the **anticommutator** $\{,\}$ is*

$$\{A, B\} \equiv AB + BA \tag{75}$$

*Two matrices are said to **commute** if $[A, B] = 0$ or **anticommute** if $\{A, B\} = 0$.*

A few properties of the commutator, which can all be proven with a little bit of linear algebra:

- Antisymmetry:
$$[A, B] = -[B, A] \tag{76}$$
- Linearity:
$$[A, B + C + D + ...] = [A, B] + [A, C] + [A, D] + ... \tag{77}$$
- Leibniz identity:
$$[A, BC] = [A, B]\, C + B\, [A, C] \tag{78}$$
- Hermitian conjugate:
$$[A, B]^\dagger = \left[B^\dagger, A^\dagger\right] \tag{79}$$
- Jacobi identity:
$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \tag{80}$$

Additionally, the commutator of two Hermitian operators is anti-Hermitian, and an operator always commutes with a scalar.

Observables A and B are defined to be **compatible** when the corresponding operators commute, and **incompatible** when $[A, B] \neq 0$. In the case of compatible observables, we assume that the ket space is spanned by the eigenkets of A as usual, and regard the same ket space as being spanned by the eigenkets of B. We may now ask how the A eigenkets are related to the B eigenkets for compatible A and B.

Recall the definition of eigenket degeneracy. You may have noticed that our earlier theorem on the orthogonality of different eigenkets was proven under the assumption of no degeneracy; furthermore, the whole concept that the ket space is spanned by $\{|a'\rangle\}$ appears to run into difficulty when the dimensionality of the ket space is larger than the number of distinct eigenvalues of A. Fortunately, in practical applications, it is usually the case that in such a situation the eigenvalues of some other commuting observable, say B, can be used to label degenerate eigenkets.

Theorem 3.3. *Suppose that A and B are compatible observables, and the eigenvalues of A are nondegenerate. Then the matrix elements of $\langle a''|B|a'\rangle$ are all diagonal.*[11]

[11] Again, you are encouraged to prove theorems by yourself. Fortunately, this one is quite simple!

We can write the matrix elements of B as

$$\langle a''|B|a'\rangle = \delta_{a'a''}\langle a'|B|a'\rangle \tag{81}$$

So both A and B can be represented by diagonal matrices with the same set of base kets. We can then write B as

$$B = \sum_{a''} |a''\rangle\langle a''|B|a''\rangle\langle a''| \tag{82}$$

If we make this operator act on an eigenket of A,

$$B|a'\rangle = \sum_{a''} |a''\rangle\langle a''|B|a''\rangle\langle a''|a'\rangle = (\langle a'|B|a'\rangle)|a'\rangle \tag{83}$$

But this is nothing other than the eigenvalue equation for the operator B with eigenvalue

$$b' \equiv \langle a'|B|a'\rangle \tag{84}$$

The ket $|a'\rangle$ is therefore a *simultaneous eigenket* of A and B. To be impartial to both operators, we may use $|a', b'\rangle$ to characterize the simultaneity. This simultaneous eigenket has the property

$$\begin{aligned} A|a', b'\rangle &= a'|a', b'\rangle \\ B|a', b'\rangle &= b'|a', b'\rangle \end{aligned} \tag{85}$$

Quite often the **collective index** K' is used to stand for (a', b'), so that

$$|K'\rangle = |a', b'\rangle \tag{86}$$

We can generalize our considerations to a situation where there are more than two mutually compatible observables. That is,

$$[A, B] = [B, C] = [A, C] = ... = 0 \tag{87}$$

Assume we have found a **maximal** set of commuting observables; that is, we cannot add any more observables to our list without violating Equation 87. The eigenvalues of individual operators may have degeneracies, but if we specify a combination $(a', b', c', ...)$ then the corresponding simultaneous eigenket of $A, B, C, ...$ is uniquely specified. We can again use a collecive index K' to stand for $(a', b', c', ...)$. The orthonormality relation for $|K'\rangle$ reads

$$\langle K''|K'\rangle = \delta_{K'K''} = \delta_{aa'}\delta_{bb'}\delta_{cc'}..., \tag{88}$$

while the completeness relation can be written as

$$\sum_{K'} |K'\rangle\langle K'| = \sum_{a'}\sum_{b'}\sum_{c'} ...|a', b', c', ...\rangle\langle a', b', c', ...| = 1 \tag{89}$$

Consider the measurements of two compatible observables A and B. Suppose we measure A first and get a', and subsequently measure B and get b'. Finally, we measure A again. The third measurement always gives a' with certainty; that is, the second (B) measurement does not destroy the previous

information obtained in the first (A) measurement. This is rather obvious when the eigenvalues of A are nondegenerate:

$$|\alpha\rangle \xrightarrow{\text{A measurement}} |a', b'\rangle \xrightarrow{\text{B measurement}} |a', b'\rangle \xrightarrow{\text{A measurement}} |a', b'\rangle \tag{90}$$

When there is degeneracy, after the first A measurement yielding a' the system is thrown into some linear combination

$$\sum_i^n c_{a'}^{(i)} \left|a', b^{(i)}\right\rangle \tag{91}$$

where the degeneracy of a' is n-fold and the kets $\left|a', b^{(i)}\right\rangle$ all have the same eigenvalue a' as far as operator A is concerned. The second measurement may select just one of the terms in the linear combination, but the third measurement applied to it still yields a'. Whether or not there is degeneracy, A measurements and B measurements do not interfere.

However, measurements of incompatible measurements do not yield the same happy results. Suppose we have a beam of particles that pass through three filters that each select some particular eigenstate and reject all others; the first (A) filter only lets in some particular $|a'\rangle$, the second (B) filter only lets in $|b'\rangle$ and the third (C) filter only lets in $|c'\rangle$. We are interested in the probability of obtaining $|c'\rangle$ from a beam of particles coming out of a'. Because probabilities are multiplicative, this would be

$$|\langle c'|b'\rangle|^2 |\langle b'|a'\rangle|^2 \tag{92}$$

Let's sum over b' to consider the total probability for going through all possible b' routes. Operationally this means that we first record the probability of obtaining c' with all but the first b' route blocked, then we repeat the procedure with all but the second b' blocked, and so on. We sum the probabilities at the end and obtain

$$\sum_{b'} |\langle c'|b'\rangle|^2 |\langle b'|a'\rangle|^2 = \sum_{b'} \langle c'|b'\rangle \langle b'|a'\rangle \langle a'|b'\rangle \langle b'|c'\rangle \tag{93}$$

We now compare this with a different arrangement, where we simply do not have a B filter. The probability is then just $|\langle c'|a'\rangle|^2$, which can be expanded using the completeness relation as

$$|\langle c'|a'\rangle|^2 = \left|\sum_{b'} \langle c'|b'\rangle \langle b'|a'\rangle\right|^2 = \sum_{b'}\sum_{b''} \langle c'|b'\rangle \langle b'|a'\rangle \langle a'|b''\rangle \langle b''|c'\rangle \tag{94}$$

Notice that Equation 93 and Equation 94 are different! This is remarkable because in both cases the pure $|a'\rangle$ beam can be regarded as being made up of B eigenkets

$$|a'\rangle = \sum_{b'} |b'\rangle \langle b'|a'\rangle \tag{95}$$

where the sum is over all the possible values of b'. Crucially, the result coming out of the C filter depends on whether or not B measurements have actually been carried out. It is only under the condition that $[A, B]$ or $[B, C]$ that equality between the two expressions actually holds.

3.3 The uncertainty relation

Given an observable A, we define an operator

$$\Delta A \equiv A - \langle A \rangle \tag{96}$$

where the expectation value is to be taken for a certain physical state under consideration.

Definition 3.4. *The **dispersion** (or **variance** or **mean square deviation**) of an observable A is the expectation value of $(\Delta A)^2$.*

Because we have

$$\langle (\Delta A)^2 \rangle = \langle \big(A^2 - 2A\langle A \rangle + \langle A \rangle^2 \big) \rangle = \langle A^2 \rangle - \langle A \rangle^2 \tag{97}$$

we can also define dispersion as $\langle A^2 \rangle - \langle A \rangle^2$.

Clearly, the dispersion vanishes when the state in question is an eigenstate of A. Roughly speaking, the dispersion of an observable characterizes "fuzziness".

We now state the uncertainty relation, which is the generalization of the well-known Heisenberg uncertainty principle, which we'll discuss later.

Theorem 3.5. *(**Uncertainty relation**) Let A and B be Hermitian operators. Then for any state we must have the following inequality:*

$$\langle (\Delta A)^2 \rangle \langle (\Delta B)^2 \rangle \geq \frac{1}{4} \left| \langle [A, B] \rangle \right|^2 \tag{98}$$

The proof of this is instructive and insightful. First, we first state three lemmas.

Lemma 3.6. *(**Schwarz inequality**)*[12]

$$\langle \alpha | \alpha \rangle \, \langle \beta | \beta \rangle \geq \left| \langle \alpha | \beta \rangle \right|^2 \tag{99}$$

This is analogous to

$$|a|^2 |b|^2 \geq |a \cdot b|^2 \tag{100}$$

in Euclidean space for two vectors a and b.

Lemma 3.7. *The expectation value of a Hermitian operator is purely real.*

Lemma 3.8. *The expectation value of an anti-Hermitian operator, defined by $C = -C^\dagger$, is purely imaginary.*

We now seek to prove the uncertainty relation. Suppose that $\langle A \rangle$ and $\langle B \rangle$ represent the expectation values with respect to some state $|\psi\rangle$; $\langle A \rangle = \langle \psi | A | \psi \rangle$, $\langle B \rangle = \langle \psi | B | \psi \rangle$. We have $(\Delta A)^2 = A^2 - 2A\langle A \rangle + \langle A \rangle^2$ hence

$$\big\langle \psi | (\Delta A)^2 | \psi \big\rangle = \langle (\Delta A)^2 \rangle \tag{101}$$

[12] The proof is easy enough; just use Equation 29, where you choose the ket you're taking the inner product of with itself to be $|\alpha\rangle + \lambda \, |\beta\rangle$, where λ is any complex number. Hint: remember that inner products are complex numbers and can be used in your guess for λ to complete the proof.

and similarly for B. Let us write the action of the operators ΔA and ΔB on $|\psi\rangle$ as

$$\begin{aligned} |\chi\rangle &= \Delta A\,|\psi\rangle \\ |\phi\rangle &= \Delta B\,|\psi\rangle \end{aligned} \tag{102}$$

The Schwarz inequality for $|\chi\rangle$ and $|\phi\rangle$ is

$$\langle\chi|\chi\rangle\,\langle\phi|\phi\rangle \geq |\langle\chi|\phi\rangle|^2 \tag{103}$$

Since A and B are Hermitian, ΔA and ΔB must also be Hermitian: $\Delta A^\dagger = A^\dagger - \langle A\rangle^\dagger$, and the adjoint of the expectation value is $\langle A\rangle$ itself since the adjoint of a complex number is its conjugate–the expectation value of a Hermitian operator is real and thus is equal to its conjugate because it has no imaginary part. By the Hermiticity of ΔA, we have

$$\begin{aligned} \langle\chi|\chi\rangle &= \big\langle\psi|\Delta A^\dagger \Delta A|\psi\big\rangle \\ &= \big\langle\psi|(\Delta A)^2|\psi\big\rangle \\ &= \langle(\Delta A)^2\rangle \end{aligned} \tag{104}$$

We can use a similar process to show that $\langle\phi|\phi\rangle = \langle(\Delta B)^2\rangle$ and $\langle\chi|\phi\rangle = \langle\psi|\Delta A|\Delta B\rangle\,\psi = \langle\Delta A\Delta B\rangle$, so the Schwarz inequality becomes

$$\langle(\Delta A)^2\rangle\langle(\Delta B)^2\rangle \geq |\langle\Delta A\Delta B\rangle|^2 \tag{105}$$

We can write $\Delta A\Delta B$ in terms of the commutators and anticommutators:

$$\begin{aligned} \Delta A\Delta B &= \frac{1}{2}\,[\Delta A, \Delta B] + \frac{1}{2}\,\{\Delta A, \Delta B\} \\ &= \frac{1}{2}\,[A, B] + \frac{1}{2}\,\{\Delta A, \Delta B\} \end{aligned} \tag{106}$$

Since $[A, B]$ is anti-Hermitian and $\{\Delta A, \Delta B\}$ is Hermitian and since the expectation value of a Hermitian operator is real and that the expectation value of an anti-Hermitian operator is imaginary, the expectation value $\langle\Delta A\Delta B\rangle$ becomes equal to the sum of a real part $\langle\{\Delta A, \Delta B\}\rangle/2$ and an imaginary part $\langle[A, B]\rangle/2$ hence

$$|\langle\Delta A\Delta B\rangle|^2 = \frac{1}{4}\,|\langle[A, B]\rangle|^2 + \frac{1}{4}\,|\langle\{\Delta A, \Delta B\}\rangle|^2 \tag{107}$$

Since the last term is a positive real number we can infer that

$$|\langle\Delta A\Delta B\rangle|^2 \geq \frac{1}{4}\,|\langle[A, B]\rangle|^2 \tag{108}$$

and hence

$$\langle(\Delta A)^2\rangle\langle(\Delta B)^2\rangle \geq \frac{1}{4}\,|\langle[A, B]\rangle|^2 \tag{109}$$

Later, we will show that $[x, p] = i\hbar$, which leads to the famous **Heisenberg uncertainty relation**

$$\langle(\Delta x)^2\rangle\langle(\Delta p_x^2)\rangle \geq \frac{\hbar^2}{4} \tag{110}$$

which is often simplified to

$$\Delta x\Delta p \geq \frac{\hbar}{2} \tag{111}$$

Physically, this sets a limit to the precision with which we can simultaneously know the position and momentum of a particle. The more accurately position is measured, the less accurately momentum is known, and vice versa.

4 Field Trip to Classical Mechanics

J. Schwinger once said, "for the fundamental properties [of quantum mechanics] we will borrow only names fromm classical physics." Before heading on to the next section, we need to make a very important analogy with the quantum commutator. While this section is rather lengthy,[13] it is thorough and complete.

4.1 Euler-Lagrange equation: a proof that $y = mx + b$

It is a well-known fact that the shortest path between two points on a plane is a straight line. To prove this, we can use the *calculus of variations*. Say we have two points, (x_1, y_1) and (x_2, y_2), and a path, $y = y(x)$, joining them. Our task is to find the path that has the shortest length. The length of a short segment of the path is $ds = \sqrt{dx^2 + dy^2}$, which we can rewrite as $ds = \sqrt{1 + \dot{y}(x)^2}dx$ since $dy = \frac{dy}{dx}dx \equiv \dot{y}(x)dx$. This gives us the arc length

$$L = \int_1^2 ds = \int_{x_1}^{x_2} \sqrt{1 + \dot{y}(x)^2}dx. \tag{112}$$

We wish to minimize this integral. This entails finding a particular path such that the integral has a value that is *stationary* relative to paths differing infinitesimally from the correct one (think of stationary points and relative minima and maxima).

For now, consider an integral J for which we wish to find a stationary value for the correct path relative to *any* neighboring path. The variation must be zero relative to some particular set of neighboring paths labeled by an infinitesimal parameter α. Such a set of paths might be denoted by $y(x, \alpha)$, with $y(x, 0)$ the correct path. Assuming our simple case of finding the shortest path between x_1 and x_2, this time generalizing to any function f between x_1 and x_2,

$$J(\alpha) = \int_{x_1}^{x_2} f[(y(x, \alpha), \dot{y}(x, \alpha), x]dx \tag{113}$$

From minimization in calculus,

$$\left(\frac{dJ}{d\alpha}\right)_{\alpha=0} = 0. \tag{114}$$

When we differentiate both sides of Equation 113 with respect to α using the rules of integration under the integral sign, we find that

$$\frac{dJ}{d\alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y}\frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial \dot{y}}\frac{\partial \dot{y}}{\partial \alpha}\right) dx. \tag{115}$$

We can say that

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial \dot{y}}\frac{\partial \dot{y}}{\partial \alpha}dx \equiv \int_{x_1}^{x_2} \frac{\partial f}{\partial \dot{y}}\frac{\partial^2 y}{\partial x \partial \alpha}dx, \tag{116}$$

so integrating by parts,

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial \dot{y}}\frac{\partial^2 y}{\partial x \partial \alpha}dx = \frac{\partial f}{\partial \dot{y}}\frac{\partial y}{\partial \alpha}\,|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}\right)\frac{\partial y}{\partial \alpha}dx \tag{117}$$

[13] Analytical mechanics is my specialty–sorry!

The conditions on all the varied curves is that they pass through the points (x_1, y_1) and (x_2, y_2), so $\partial y/\partial\alpha$ at x_1 and x_2 must vanish, therefore the first term of Equation 117 vanishes, so 115 reduces to

$$\frac{dJ}{d\alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial \dot{y}} \right) \frac{\partial y}{\partial \alpha} dx, \tag{118}$$

which is zero at $\alpha = 0$.

Lemma 4.1. *(Fundamental lemma of the calculus of variations) If*

$$\int_{x_1}^{x_2} M(x)\eta(x)dx = 0 \tag{119}$$

for all arbitrary functions $\eta(x)$ continuous through the second derivative, then $M(x)$ must identically vanish in the interval (x_1, x_2).

While a formal mathematical proof of the lemma can be found textbooks on calculus of variations, we're interested in the physics, namely what follows by applying the lemma to Equation 118.

Theorem 4.2. *(Euler-Lagrange equation) For an integral of a function $f(y(x), \dot{y}(x), x)$ from points x_1 to x_2 to have a stationary value,*

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial \dot{y}} = 0 \tag{120}$$

for all x in the relevant interval. If there are n dependent variables within the original integral, there are n Euler-Lagrange equations.

Applying this to our minimization of length, we let

$$f(y, \dot{y}, x) = \sqrt{1 + \dot{y}^2}. \tag{121}$$

Evaluating the partial derivatives,

$$\frac{\partial f}{\partial y} = 0, \quad \frac{\partial f}{\partial \dot{y}} = \frac{\dot{y}}{\sqrt{1 + \dot{y}^2}} \tag{122}$$

$$\Rightarrow \frac{d}{dx}\frac{\partial f}{\partial \dot{y}} = 0 \tag{123}$$

This implies that $\partial f/\partial \dot{y}$ is a constant, C, therefore

$$\dot{y}^2 = C^2(1 + \dot{y}^2). \tag{124}$$

It is quickly apparent that $\dot{y}$ is constant–we'll let $\dot{y} = m$ and integrating, $y = mx + b$.

4.2 Principle of least action

In the Lagrangian formulation of mechanics, the independent variable is time. The dependent variables are the coordinates that specify the position or "configuration" of a system, and are usually denoted by $q_1, q_2, ..., q_n$. The number n of coordinates depends on the nature of the system; for a single particle in three

dimensions, q_1, q_2, q_3 could be x, y, z, could be r, θ, ϕ, or something else. For N particles moving freely in three dimensions, n is $3N$. Because the coordinates $q_1, ..., q_n$ can take so many forms, they are called **generalized coordinates**. Mathematically, we can think of a system of n generalized coordinates as defining a point in an n-dimensional **configuration space** Q, each of whose points labels a unique configuration of the system. Each position $\mathbf{r}$ as a function of $q_1, q_2, ..., q_n$, and possibly t, and conversely, each q_i can be expressed in terms of the positions and possible t. In Lagrangian mechanics, one forms the **tangent space** T_qQ at every point $q \in Q$, which is the set of all possible velocity vectors at that point. The set of all tangent spaces is called the **velocity phase space** TQ (also called the **tangent bundle** of Q); coordinates on TQ are denoted

$$(q_1, ..., q_n, \dot{q}_1, ..., \dot{q}_n) \tag{125}$$

The integral whose stationary value determines the evolution of the mechanical system is called the action integral, and its integrand is the Lagrangian $\mathcal{L} = \mathcal{L}(q_1, \dot{q}_1, ..., q_n, \dot{q}_n, t)$. In other words, Lagrangian is a function $\mathcal{L} : TQ \to \mathbb{R}$.

Definition 4.3. *The **action integral** is defined as*

$$S = \int_{t_1}^{t_2} \mathcal{L}(q_1, \dot{q}_1, ..., q_n, \dot{q}_n, t)dt. \tag{126}$$

Theorem 4.4. *(Hamilton's principle) The actual path which a particle follows between two points in a given time interval is such that the action integral is stationary when taken along the actual path.*

4.3 Defining the Lagrangian

Corollary 4.5. *(Lagrange's equations) For the Lagrangian, the Euler-Lagrange equations are called Lagrange's equations:*

$$\frac{\partial \mathcal{L}}{\partial q_k} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_k} \tag{127}$$

for $k = 1, ..., n$.

Definition 4.6. *The **Lagrangian function**, or just **Lagrangian**, is defined as*

$$\mathcal{L} = T - U, \tag{128}$$

where T is the kinetic energy and U is the potential energy.

Equation 128 is specifically for the *non-relativistic Lagrangian for a system of particles in the absence of an electromagnetic field.*

Consider a particle that moves unconstrained in three dimensions, subject to a net force $\mathbf{F}(\mathbf{r})$. The particle's kinetic energy is of course

$$T = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{\mathbf{r}}^2 \tag{129}$$

and its potential energy is

$$U = U(\mathbf{r}). \tag{130}$$

Consider the partials of the Lagrangian

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= -\frac{\partial U}{\partial x} = F_x, \\ \frac{\partial \mathcal{L}}{\partial \dot{x}} &= \frac{\partial T}{\partial \dot{x}} = m\dot{x} = p_x. \end{aligned} \tag{131}$$

Equation 127 leads us to

$$F_x = \frac{d}{dt}(p_x), \tag{132}$$

which is Newton's second law! **The Lagrangian formulation of mechanics has no *fundamental* difference to Newtonian mechanics; they're both equally valid models of the same phenomena.**

Definition 4.7. *The **generalized force** is defined as*

$$\frac{\partial \mathcal{L}}{\partial q_i} = \textit{ith component of generalized force} \tag{133}$$

*and the **generalized momentum** or **conjugate momentum** is*

$$\frac{\partial \mathcal{L}}{\partial \dot{q}_i} = \textit{ith component of conjugate momentum.} \tag{134}$$

If we work in polar coordinates, the ϕ component of generalized force is torque, and the ϕ component of generalized momentum is angular momentum. Generalized forces/momenta need not necessarily have the dimensions of force/momentum.

Definition 4.8. *When the Lagrangian is independent of a coordinate q_i, that coordinate is sometimes said to be **ignorable** or **cyclic**.*

What we mean by this is that $\mathcal{L}$ is unchanged, or *invariant*, when q_i varies (with all other generalized coordinates held fixed). Thus we can say that if $\mathcal{L}$ is invariant under variations of a coordinate q_i, then the corresponding generalized momentum p_i is conserved. It is a good idea to choose coordinates so that as many as possible are ignorable and their corresponding momenta are constant. In fact, this is perhaps the main criterion in choosing generalized coordinates for any given problem.

4.4 Hamiltonian mechanics

In Hamiltonian mechanics, we used the conjugate momenta instead of the generalized velocity: that is, we make the change of variables from $(q_i, \dot{q}_i)$ to (q_i, p_i). For each tangent space T_qQ, we can define a **cotangent space** T_q^*Q, which is dual to the tangent space. The set of all cotangent spaces on Q is called the **phase space** or **cotangent bundle** with coordinates $(q_1, ...q_n, p_1, ...p_n)$. These coordinates are called the **canonical coordinates**.

Definition 4.9. *The **Hamiltonian** is*

$$H(q_1, ..., q_n, p_1, ...p_n, t) = \sum_{j=1}^{n} p_j \dot{q}_j - \mathcal{L}(q_1, ..., q_n, \dot{q}_1, ...\dot{q}_n, t) \tag{135}$$

The kinetic energy of a system of N particles can be expressed in the form

$$T = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{1}{2}\sum_{k=1}^{3N} m_k \dot{x}_k^2 \tag{136}$$

To express this in terms of generalized coordinates and generalized velocities, we use the chain rule to obtain

$$\dot{x}_k = \sum_{i=1}^{n} \frac{\partial x_k}{\partial q_i}\dot{q}_i + \frac{\partial x_k}{\partial t} \tag{137}$$

We find that the kinetic energy is thus

$$T = \frac{1}{2}\sum_{k=1}^{3N} m_k \left(\sum_{i=1}^{n} \frac{\partial x_k}{\partial q_i}\dot{q}_i + \frac{\partial x_k}{\partial t}\right)^2 \tag{138}$$

We can expand this in a quadratic form:

$$T = T_2 + T_1 + T_0 \tag{139}$$

where the quadratic portion is

$$T_2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\dot{q}_i\dot{q}_j \tag{140}$$

The mass coefficients m_{ij} are

$$m_{ij} = \sum_{k=1}^{3N} m_k \frac{\partial x_k}{\partial q_i}\frac{\partial x_k}{\partial q_j} = \frac{\partial^2 T}{\partial \dot{q}_i \partial \dot{q}_j} \tag{141}$$

The portion that is linear in the $\dot{q}$s is

$$T_1 = \sum_{i=1}^{n} a_i \dot{q}_i \tag{142}$$

where

$$a_i = \sum_{i=1}^{3N} m_k \frac{\partial x_k}{\partial q_i}\frac{\partial x_k}{\partial t} \tag{143}$$

And finally the portion of kinetic energy which is not a function of the $\dot{q}$s is

$$T_0 = \frac{1}{2}\sum_{k=1}^{3N} m_k \left(\frac{\partial x_k}{\partial t}\right)^2 \tag{144}$$

For a system that is *scleronomic*–that is, the constraints on the particle don't change with time–the term $\partial x_k/\partial t = 0$, vanishing T_1 and T_0 and leaving T_2.[14]

[14]Most systems are scleronomic, but an example is a pendulum with an oscillating pivot point, because the constraints change with time. Non-scleronomic systems are said to be rheonomic.

For the time being, we introduce the function

$$T' = \sum p_j \dot{q}_j - T \tag{145}$$

In the case that the potential energy is *not* velocity dependent, which is in practically all cases except for electromagnetic fields (Lorentz force), this reduces to

$$\sum \frac{\partial T}{\partial \dot{q}_j} - T = (2T_2 + T_1) - (T_2 + T_1 + T_0) = T_2 - T_0 \tag{146}$$

That is, in the case of non-velocity dependent potentials in sceronomic systems (which is certainly the case for a free particle), $T' = T$, and we find that the Hamiltonian reduces to

$$H = T + U \tag{147}$$

which is just the *total energy*! It is for this reason that we regard the Hamiltonian to be sort of a generalized form of energy.[15]

Theorem 4.10. *Lagrange's equations are equivalent to* ***Hamilton's equations***

$$\begin{aligned} \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i} \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i} \end{aligned} \tag{148}$$

Let for a system with n degrees of freedom, let η be a column matrix with $2n$ elements such that

$$\eta = \begin{pmatrix} q_1 \\ \vdots \\ q_n \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \tag{149}$$

Moreover, let J be the $2n \times 2n$ square matrix composed of the $n \times n$ zero and unit matrices according to the scheme

$$J = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix} \tag{150}$$

where $\mathbf{0}$ is the $n \times n$ matrix all of whose elements are zero, and $\mathbf{1}$ is the standard $n \times n$ unit matrix. Hamilton's equations can be written in compact form as

$$\dot{\eta} = J \frac{\partial H}{\partial \eta} \tag{151}$$

this is known as *symplectic notation.*

[15] Actually, for a charged particle in an electromagnetic field, the Hamiltonian is still the total energy–just not $T + U$.

4.5 Poisson brackets

Hamilton's equations can be recast in **Poisson bracket form**

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i} = [q_i, H] \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i} = [p_i, H] \end{aligned} \tag{152}$$

where the **Poisson bracket** is given by

$$[F, G] = \sum_{i=1}^{n} \left(\frac{\partial F}{\partial q_i} \frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial G}{\partial q_i} \right) \tag{153}$$

and F and G are functions of the p_i and q_i. By closer examination, you'll find that the following general properties is guaranteed:

$$\dot{F} = [F, H] \tag{154}$$

It's quite easy to verify this using the chain rule to expand the time derivative of F. This result has very deep and beautiful implications. In fact, a version of the conservation of energy follows by considering that the Hamiltonian obviously Poisson-commutes with itself and is thus invariant with time.

You can also easily verify the following fundamental Poisson bracket relations:

$$[q_i, q_j] = 0 \quad [p_i, p_j] = 0 \quad [q_i, p_j] = \delta_{ij} \tag{155}$$

Of even more fascinating nature, it can be verified that the Poisson bracket has the same properties as the quantum commutator: antisymmetry, linearity, the Jacobi identity, and everything else that follows. In fact, Dirac discovered that the Poisson bracket is a classical analogue of the quantum commutator, a crucial result that we will use later.

In symplectic notation, we have

$$[F, G] = \widetilde{\frac{\partial F}{\partial \eta}} J \frac{\partial G}{\partial \eta} \tag{156}$$

The transpose sign on the first matrix on the right-hand side indicates that this matrix must be treated as a single row matrix in the multiplication.

Hamilton's equations become

$$\dot{\eta} = [\eta, H] = J \frac{\partial H}{\partial \eta} \tag{157}$$

We can rewrite the fundamental relations in Equation 155 as

$$[\eta, \eta] = J \tag{158}$$

4.6 Canonical transformations

Often in Hamiltonian mechanics, there will be a need to change the coordinates of the phase space for simplicity. For a set of original coordinates (q_i, p_i), we wish to transform our problem onto a new set of coordinates (Q_i, P_i), where the Q_i

and P_i are functions of the q's, p's, and time. Such a transformation between two sets of canonical coordinates is called a **canonical transformation**. For this transformation to be valid, we require there to be a transformed Hamiltonian K satisfying the transformed Hamilton's equations

$$\begin{aligned} \dot{Q}_i &= \frac{\partial K}{\partial P_i} \\ \dot{P}_i &= -\frac{\partial K}{\partial Q_i} \end{aligned} \tag{159}$$

This implies a modified Hamilton's principle

$$\delta \int_{t_1}^{t_2} P_i Q_i - K(Q, P, t) dt = 0 \tag{160}$$

While it is true that

$$\delta \int_{t_1}^{t_2} p_i q_i - H(q, p, t) dt = 0 \tag{161}$$

this does not imply equality in the integrands. In fact, they are related by the following relation:

$$\lambda(p_i \dot{q}_i - H) = P_i \dot{Q}_i - K + \frac{dF}{dt} \tag{162}$$

λ is a constant, and F is a function in phase space called the **generating function**, and acts as a bridge between the sets of canonical coordinates. A trivial example of a canonical transformation is the scale transformation, where we just scale the canonical coordinates (e.g. change the units). If we have

$$\begin{aligned} Q_i &= \mu q_i \\ P_i &= \nu p_i \end{aligned} \tag{163}$$

where μ, ν are constants, then $\lambda = \mu\nu$ and F vanishes. Since this is trivial, we'll only consider cases in which $\lambda = 1$.

If $F = F_1(q, Q, t)$, then

$$p_i \dot{q}_i - H = P_i \dot{Q}_i + \frac{\partial F_1}{\partial t} + \frac{\partial F_1}{\partial q_i} \dot{q}_i + \frac{\partial F_1}{\partial Q_i} \dot{Q}_i \tag{164}$$

Because $\dot{q}_i$ and $\dot{Q}_i$ are independent of each other, for equality to hold we must have

$$\begin{aligned} p_i \dot{q}_i &= \frac{\partial F_1}{\partial q_i} \dot{q}_i \\ 0 &= P_i \dot{Q}_i + \frac{\partial F_1}{\partial Q_i} \dot{Q}_i \end{aligned} \tag{165}$$

which implies

$$\begin{aligned} p_i &= \frac{\partial F_1}{\partial q_i} \\ P_i &= -\frac{\partial F_1}{\partial Q_i} \\ K &= H + \frac{\partial F}{\partial t} \end{aligned} \tag{166}$$

If our generating function is $F_2(q_i, P_i, t)$, then we set $F = F_2 - Q_i P_i$ so that we obtain

$$p_i\dot{q}_i - H = -Q_i\dot{P}_i - K + \frac{\partial F_2}{\partial t} + \frac{\partial F_2}{\partial q_i}\dot{q}_i + \frac{\partial F_2}{\partial P_i}\dot{P}_i \tag{167}$$

so that

$$\begin{aligned} p_i &= \frac{\partial F_2}{\partial q_i} \\ Q_i &= \frac{\partial F_2}{\partial P_i} \\ K &= H + \frac{\partial F_2}{\partial t} \end{aligned} \tag{168}$$

We are now ready to consider the **infinitesimal canonical transformation**; that is, one for which

$$\begin{aligned} Q_i &= q_i + \delta q_i \\ P_i &= p_i + \delta p_i \end{aligned} \tag{169}$$

where δq_i and δp_i denote infinitesimal changes in the respective canonical coordinates. The generating function would look something like $F = q_i P_i + \epsilon G(q, P, t)$ where ϵ is an infinitesimal parameter, and G is differentiable in the q's, p's, and t. The transformation equations for the momenta are found to be of the form

$$p_j = \frac{\partial F}{\partial q_j} = P_j + \epsilon\frac{\partial G}{\partial q_j} \tag{170}$$

which means

$$\delta p_j = -\epsilon\frac{\partial G}{\partial q_j} \tag{171}$$

Similarly,

$$Q_j = \frac{\partial F}{\partial p_j} = q_j + \epsilon\frac{\partial G}{\partial P_j} \tag{172}$$

Since the second term is linear in ϵ and P differs from p only by an infinitesimal, it is consistent to the first order to replace P_j in the derivative function by p_j. We may then consider G as a function of q, p only, and possibly t. The transformation equation can be written as

$$\delta q_j = \epsilon\frac{\partial G}{\partial p_j} \tag{173}$$

Consider a simple example of a transformation where we have only one generalized coordinate, x, that changes by an amount dx (our "δq), and the momentum is left invariant. Since the difference in momentum δp_x is zero,

$$\epsilon\frac{\partial G}{\partial x} = 0 \tag{174}$$

while

$$\epsilon\frac{\partial G}{\partial p_x} = dx \neq 0 \tag{175}$$

The derivative of G with respect to x vanishes, but not with respect to momentum. In other words, G is a function of p only. For this reason we say that **momentum *generates* infinitesimal spatial translations**. While the proof of the following statement is lengthy and will not be displayed, we can generalize to get a very important result:

Theorem 4.11. *The constants of the motion are the generating functions of those infinitesimal canonical transformations that leave the Hamiltonian invariant.*

5 Time

5.1 Change of basis

Suppose we have two incompatible observables A and B. The ket space in question can be viewed as being spanned by either the set $\{|a'\rangle\}$ or $\{|b'\rangle\}$. Chainging the set of base kets is referred to as a **change of basis** or a **change of representation**. The basis in which the eigenkets are given by $\{|a'\rangle\}$ is called the A representation. Our basic task is to construct a transformation operator that connects the old orthonormal set $\{|a'\rangle\}$ and the new orthonormal set $\{|b'\rangle\}$. We begin with the following.

Theorem 5.1. *Given two sets of base kets, both satisfying orthonormality and completeness, there exists a unitary operator U such that*

$$\left|b^{(1)}\right\rangle = U\left|a^{(1)}\right\rangle, \quad \left|b^{(2)}\right\rangle = U\left|a^{(2)}\right\rangle, \quad \ldots \quad \left|b^{(N)}\right\rangle = U\left|a^{(N)}\right\rangle \tag{176}$$

*By a **unitary operator** we mean an operator fulfilling the conditions*

$$\begin{aligned} U^\dagger U &= 1 \\ UU^\dagger &= 1 \end{aligned} \tag{177}$$

It is instructive to study the matrix representation of the U operator in the old $\{|a'\rangle\}$ basis. We have

$$\left\langle a^{(k)}|U|a^{(l)}\right\rangle = \left\langle a^{(k)}|b^{(l)}\right\rangle \tag{178}$$

In other words, the matrix elements of the U operator are built up of the inner products of old base bras and new base kets. The square matrix made up of $\left\langle a^{(k)}|U|a^{(l)}\right\rangle$ is referred to as the **transformation matrix** from the $\{|a'\rangle\}$ basis to the $\{|b'\rangle\}$ basis. This may be reminiscent of the rotation matrix that changes one set of unit base vectors into another set–and fundamentally, it is, except written in Dirac notation and generalized to quantum mechanics.

Given an arbitrary ket $|\alpha\rangle$ whose expansion coefficients $\langle a'|\alpha\rangle$ are known in the old basis:

$$|\alpha\rangle = \sum_{a'} |a'\rangle \langle a'|\alpha\rangle \tag{179}$$

we obtain the expansion coefficients in the new basis $\langle b'|\alpha\rangle$ by multiplying both sides of Equation 179 by $\left\langle b^{(k)}\right|$ and substituting the transformation matrix:

$$\begin{aligned} \left\langle b^{(k)}|\alpha\right\rangle &= \sum_l \left\langle b^{(k)}|a^{(l)}\right\rangle \left\langle a^{(l)}|\alpha\right\rangle \\ &= \sum_l \left\langle a^{(k)}|U^\dagger|a^{(l)}\right\rangle \left\langle a^{(l)}|\alpha\right\rangle \end{aligned} \tag{180}$$

In matrix notation, this just states that the column matrix for $|\alpha\rangle$ in the new basis can be obtained just by applying the square matrix $U^\dagger$ to the column matrix in the old basis. This is the **similarity transformation**

$$X' = U^\dagger X U \tag{181}$$

Definition 5.2. *The **trace** of an operator X is defined as the sum of diagonal elements:*

$$tr(X) = \sum_{a'} \langle a'|X|a'\rangle \tag{182}$$

Importantly, the trace is independent of representation. We can also prove

$$\begin{aligned} \mathrm{tr}\,(XY) &= \mathrm{tr}\,(YX) \\ \mathrm{tr}\left(U^\dagger XU\right) &= \mathrm{tr}\,(X) \\ \mathrm{tr}\left(|a'\rangle\langle a''|\right) &= \delta_{a'a''} \\ \mathrm{tr}\left(|b'\rangle\langle a'|\right) &= \langle a'|b'\rangle \end{aligned} \tag{183}$$

Theorem 5.3. *Consider again two sets of orthonormal basis $\{|a'\rangle\}$ and $\{|b'\rangle\}$ connected by the U operator. Knowing U, we may construct a **unitary transform** of A, UAU^{-1}; then A and UAU^{-1} are said to be **unitary equivalent observables**. The eigenvale equation for A*

$$A\left|a^{(l)}\right\rangle = a^{(l)}\left|a^{(l)}\right\rangle \tag{184}$$

implies that

$$UAU^{-1}U\left|a^{(l)}\right\rangle = a^{(l)}U\left|a^{(l)}\right\rangle \tag{185}$$

which can be rewritten

$$\left(UAU^{-1}\right)\left|b^{(l)}\right\rangle = a^{(l)}\left|b^{(l)}\right\rangle \tag{186}$$

This tells us that the $|b'\rangle$ are eigenkets of UAU^{-1} with exactly the same eigenvalues as the A eigenvalues. In other words, *unitary equivalent observables have identical spectra.*

5.2 Generalization to continuous spectra

We've been considering discrete eigenvalue spectra. However, there are also observables with continuous eigenvalues, such as momentum. The vector space by eigenkets that exhibit a continuous spectrum is infinitely dimensional. While this may seem alarming, most generalizations to continuous spectra are relatively straightforward.

Firstly, consider an operator ξ. We wish to find the corresponding expressions for the continuous basis $\{\xi'\}$. The eigenvalue equation is written

$$\xi\,|\xi'\rangle = \xi'\,|\xi'\rangle \tag{187}$$

In pursuing this analogy, we replace sums with integrals. We also replace the Kronecker delta with the Dirac delta function.

Definition 5.4. *The "function" $\delta(x-x')$ with the properties*

$$\begin{aligned} \delta(x-x') &= \begin{cases} +\infty, & x = x' \\ 0, & x \neq x' \end{cases} \\ \int_a^b \delta(x-x')dx' &= 1, \quad a < x < b \end{aligned} \tag{188}$$

is called the ***Dirac delta function*** *and fixes the normalization of the basis vectors:*

$$\langle \xi' | \xi'' \rangle = \delta(\xi' - \xi'') \tag{189}$$

The completeness relation $\sum_{a'} |a'\rangle \langle a'| = 1$ becomes

$$\int d\xi' \, |\xi'\rangle \langle \xi'| = 1 \tag{190}$$

We can use this to derive the expansion of a state ket

$$|\alpha\rangle = \int d\xi' \, |\xi'\rangle \langle \xi'|\alpha\rangle \tag{191}$$

and the definition of the inner product between two states

$$\langle \beta|\alpha\rangle = \int d\xi' \, \langle \beta|\xi'\rangle \langle \xi'|\alpha\rangle \tag{192}$$

Additionally, the sum of the probabilities $\sum_{a'} |\langle a'|\alpha\rangle|^2 = 1$ becomes

$$\int d\xi' \, |\langle \xi'|\alpha\rangle|^2 = 1 \tag{193}$$

Finally, the matrix representation of $|\xi\rangle$ becomes

$$\langle \xi''|\xi|\xi'\rangle = \xi' \delta(\xi'' - \xi') \tag{194}$$

5.3 Position

Let's consider the position operator x in one dimension with eigenkets $|x'\rangle$. The state ket for an arbitrary physical state can be expanded in terms of $\{|x'\rangle\}$:

$$|\alpha\rangle = \int_{-\infty}^{\infty} dx' \, |x'\rangle \langle x'|\alpha\rangle \tag{195}$$

Suppose we place a very tiny detector that clicks only when the particle is precisely at x' and nowehere else. Immediately after the detector clicks, we say that the state in question is represented by $|x'\rangle$–that $|\alpha\rangle$ abruptly "jumps into" $|x'\rangle$. In practice, the best the detector can do is to locate the particle within a narrow interval $(x' - \Delta/2, x' + \Delta/2)$, and when a count is registered, the state ket changes abruptly into

$$|\alpha\rangle \to \int_{x'-\Delta/2}^{x'+\Delta/2} dx'' \, |x''\rangle \langle x''|\alpha\rangle \tag{196}$$

Assuming $|x''\rangle \, \alpha$ does not change appreciably within the narrow interval, the probability for the detector to click is given by

$$|\langle x'|\alpha\rangle|^2 \, dx' \tag{197}$$

where we have written dx' for Δ. Given that $|\alpha\rangle$ is normalized (that is, $\langle \alpha|\alpha\rangle = 1$) the probability of recording the particle somewhere between $-\infty$ and ∞ is given by

$$\int_{-\infty}^{\infty} dx' \, |\langle x'|\alpha\rangle|^2 \tag{198}$$

Generalizing this to three dimensions, we can expand the state ket in terms of $\{|r'\rangle\}$ as follows:

$$|\alpha\rangle = \int d^3x' \, |r'\rangle \, \langle r'|\alpha\rangle \tag{199}$$

where r' stands for x', y', and z'; in other words, $|r'\rangle$ is a simultaneous eigenket of the observables x, y, z. To be able to consider such a simultaneous eigenket at all, we are implicitly assuming that the three components of the position vector can be measured simultaneously to arbitrary degrees of accuracy; hence, we must have

$$[x_i, x_j] = 0 \tag{200}$$

where x_1, x_2, x_3 stand for x, y, z respectively.

5.4 Translation

Suppose we start with a state that is well localized around r'. Let us consider an operation that changes this state into another well-localized state aroud $r' + dr'$ with everything else about the state unchanged. Such an operation is an **infinitesimal translation** by dr' and the operator that does that job is $\mathscr{T}(dr')$:

$$\mathscr{T}(dr') \, |dr'\rangle = |r + dr'\rangle \tag{201}$$

By expanding an arbitrary state ket $|\alpha\rangle$ in terms of the position eigenkets we can examine the effect of the infinitesimal translation on $|\alpha\rangle$:

$$\mathscr{T}(dr') \, |\alpha\rangle = \mathscr{T}(dr') \int d^3x' \, |r'\rangle \, \langle r'|\alpha\rangle = \int d^3x' \, |r' + dr'\rangle \, \langle r'|\alpha\rangle \tag{202}$$

It is reasonable to require that if the ket $|\alpha\rangle$ is normalized to unity, the translated ket $\mathscr{T}(dr') \, |\alpha\rangle$ also be normalized to unity, which requires that the infinitesimal translation be unitary:

$$\mathscr{T}^\dagger(dr')\mathscr{T}(dr') = 1 \tag{203}$$

Suppose we consider two successive infinitesimal translations: dr' followed by dr'', where dr' and dr'' need not be in the same direction. We expect the net result to be the same as a single translation operation by the vector sum $dr' + dr''$ so we demand that

$$\mathscr{T}(dr'')\mathscr{T}(dr') = \mathscr{T}(dr' + dr'') \tag{204}$$

Suppose we consider a translation in the opposite direction: this would be the inverse as the original translation,

$$\mathscr{T}(-dr') = \mathscr{T}^{-1}(dr') \tag{205}$$

Finally, we demand that as $dr' \to 0$, the translation operation reduce to the identity operation

$$\lim_{dr' \to 0} \mathscr{T}(dr') = 1 \tag{206}$$

If we take the infinitesimal translation operator to be

$$\mathscr{T}(dr') = 1 - iK \cdot dr' \tag{207}$$

where the components of K–K_x, K_y, K_z– are Hermitian operators, then all the properties listed are satisfied.[16] First note that

$$r\mathscr{T}(dr')\left|r'\right\rangle = r\left|r'+dr'\right\rangle = (r'+dr')\left|r'+dr'\right\rangle \tag{208}$$

and

$$\mathscr{T}(dr')r\left|r'\right\rangle = r'\mathscr{T}(dr')\left|r'\right\rangle = r'\left|r'+dr'\right\rangle \tag{209}$$

Hence,

$$[r,\mathscr{T}(dr')]\left|r'\right\rangle = dr'\left|r'+dr'\right\rangle \tag{210}$$

Cancelling terms of second order in dr', this reduces to $dr'\left|r'\right\rangle$. We must therefore have an operator identity

$$[r,\mathscr{T}(dr')] = dr' \tag{211}$$

or

$$-irK\cdot dr' + iK\cdot dr'r = dr' \tag{212}$$

By choosing dr' in the direction of x_j and forming the scalar product with x_i, we obtain

$$[x_i,K_j] = i\delta_{ij} \tag{213}$$

Now, remember generating functions from classical mechanics. Our infinitesimal translation can be analogously regarded as an active infinitesimal canonical translation,

$$\begin{aligned} R &= r'+dr' \\ P &= p' \end{aligned} \tag{214}$$

The generating function G is p since it is the constant of motion. This leaves us with the generating function for the translation itself,

$$F(r,P) = r\cdot P + p\cdot dr \tag{215}$$

This is of resemblance to the definition of the quantum translation operator in Equation 207, and we can actually associate K with p; to fit the dimensions, we set K to be

$$K = \frac{p}{\hbar} \tag{216}$$

to write

$$\mathscr{T}(dr') = 1 - ip\cdot dr'/\hbar \tag{217}$$

We can then rewrite Equation 213 as

$$[x_i,p_j] = i\hbar\delta_{ij} \tag{218}$$

Plugging it into the uncertainty relation, we finally get **Heisenberg's position-momentum uncertainty relation**:

$$\langle(\Delta x)^2\rangle\langle(\Delta p_x^2)\rangle \geq \frac{\hbar^2}{4} \tag{219}$$

[16] I encourage you to verify these yourself. Remember that terms of second order in dr' can be ignored for an infinitesimal translation.

One can obtain finite translations by successively compounding infinitesimal translations. Let us consider a finite translation by an amount $\Delta x'$ in the x-direction:

$$\mathscr{T}(\Delta x' x)\,|x'\rangle = |x' + \Delta x' x\rangle \tag{220}$$

By compounding N infinitesimal translations each $\Delta x'/N$ in the x-direction and letting $N \to \infty$, we obtain

$$\mathscr{T}(\Delta x' x) = \lim_{N\to\infty}\left(1 - \frac{ip_x \Delta x'}{N\hbar}\right)^N \tag{221}$$

But this is just the Taylor series expansion of the exponential function,

$$\mathscr{T}(\Delta x' x) = \exp\left[\frac{ip_x \Delta x'}{\hbar}\right] \tag{222}$$

Successive translations in different directions commute for obvious reasons. If we have two successive translations $\Delta x'$ and $\Delta y'$ (in the x- and y-directions), we treat them up to second order to obtain

$$\begin{aligned}[\mathscr{T}(\Delta y' y, \mathscr{T}(\Delta x' x] &= \left[\left(1 - \frac{ip_y \Delta y'}{\hbar} - \frac{p_y^2(\Delta y')^2}{2\hbar^2}\right), \left(1 - \frac{ip_x \Delta x'}{\hbar} - \frac{p_x^2(\Delta x')^2}{2\hbar^2}\right)\right] \\ &\approx -\frac{(\Delta x')(\Delta y')\,[p_y, p_x]}{\hbar^2} \\ &= 0\end{aligned} \tag{223}$$

This implies that

$$[p_x, p_y] = 0 \tag{224}$$

Here, we turn back to classical mechanics. Recall the Poisson bracket. It can easily be verified that

$$\begin{aligned}[q_i, q_i] &= 0 \\ [p_j, p_j] &= 0 \\ [q_i, p_j] &= \delta_{ij}\end{aligned} \tag{225}$$

These are the classical analogs of the **canonical commutation relations** in quantum mechanics

$$\begin{aligned}[x_i, x_i] &= 0 \\ [p_j, p_j] &= 0 \\ [x_i, p_j] &= i\hbar\delta_{ij}\end{aligned} \tag{226}$$

5.5 The classical analogy: the Heisenberg picture

Since the Poisson brackets and commutators have the same properties[17] Dirac concluded there is a correspondence principle between classical and quantum relations by replacing classical Poisson bracket by commutators, as follows:

$$[\ ,\ ]_{classical} \to \frac{[\ ,\ ]}{i\hbar} \tag{227}$$

[17]They are both *Lie brackets*, a general formalism of which is beyond the scope of this discussion.

There are two important differences between the two Lie brackets: first, the dimension of the Poisson bracket differs from the commutator because of the differentiations with respect to the q's and p's, and second, the Poisson bracket of two real functions is purely real, while the commutator of two Hermitian operators is anti-Hermitian. To take care of these differences, we inserted the factor $i\hbar$.

As we have discussed, for a function of spatial coordinates and momenta that does not involve time independently, its time derivative is given by its Poisson bracket with the Hamiltonian:

$$\frac{dA}{dt} = [A, H]_{classical} \tag{228}$$

We now derive a fundamental equation. First we introduce the **Heisenberg picture**. This is one of two important *dynamical pictures* which formulate the dynamics of a quantum mechanical system. In this approach, for unitary transformations, $X \to U^\dagger X U$ with state kets unchanged. In classical physics we do not introduce state kets, yet we talk about translation, time evolution, and so on. This is possible because these operations actually change the quantities such as position or momentum, which are observables of classical mechanics. The time evolution of a Heisenberg picture observable is defined by

$$A(t) = \mathscr{U}^\dagger(t) A(0) \mathscr{U}(t) \tag{229}$$

where the time evolution operator is given by

$$\mathscr{U}(t) = \exp\left(\frac{-iHt}{\hbar}\right) \tag{230}$$

Differentiating with respect to time,

$$\begin{aligned} \frac{dA}{dt} &= \frac{\partial \mathscr{U}^\dagger}{\partial t} A(0) \mathscr{U} + \mathscr{U}^\dagger A(0) \frac{\partial \mathscr{U}}{\partial t} \\ &= -\frac{1}{i\hbar} \mathscr{U}^\dagger H \mathscr{U} \mathscr{U}^\dagger A(0) \mathscr{U} + \frac{1}{i\hbar} \mathscr{U}^\dagger A(0) \mathscr{U} \mathscr{U}^\dagger H \mathscr{U} \\ &= \frac{1}{i\hbar} [A(0), \mathscr{U}^\dagger H \mathscr{U}] \end{aligned} \tag{231}$$

where we have used

$$\begin{aligned} \frac{\partial \mathscr{U}}{\partial t} &= \frac{1}{i\hbar} H \mathscr{U} \\ \frac{\partial \mathscr{U}^\dagger}{\partial t} &= -\frac{1}{i\hbar} \mathscr{U}^\dagger H \end{aligned} \tag{232}$$

The time evolution operator and the Hamiltonian commute, so $\mathscr{U}^\dagger H \mathscr{U} = H$, so we may write

$$\frac{dA}{dt} = \frac{1}{i\hbar} [A, H] \tag{233}$$

5.6 Ehrenfest's theorem: Newton's second law of quantum mechanics

Consider a particle in a potential field $V(x)$. The Hamiltonian is

$$H = \frac{p^2}{2m} + V(x) \tag{234}$$

$V(x)$ is to be understood as a function of the x, y, and z operators. We use the formula

$$[p_i, G(x)] = -i\hbar \frac{\partial G}{\partial x_i} \tag{235}$$

where $G(x)$ is a function that can be expanded in powers of x_j's; the proof follows from a repeated application of the Leibniz rule. Therefore, from the Heisenberg equation of motion,

$$\frac{dp_i}{dt} = \frac{1}{i\hbar}[p_i, V(x)] = -\frac{\partial}{\partial x_i} V(x) \tag{236}$$

We have $\frac{dx_i}{dt} = \frac{p_i}{m}$ because x_i commutes with the newly added term. We can use the Heisenberg equation of motion again to get

$$\frac{d^2 x_i}{dt^2} = \frac{1}{i\hbar}\left[\frac{dx_i}{dt}, H\right] = \frac{1}{i\hbar}\left[\frac{p_i}{m}, H\right] = \frac{1}{m}\frac{dp_i}{dt} \tag{237}$$

Then we final get in vectorial form

$$m\frac{d^2 x}{dt^2} = -\nabla V(x) \tag{238}$$

which looks exactly like Newton's second law! By taking the expectation values of both sides with respect to a Heisenberg state ket that does not move with time we obtain the **Ehrenfest theorem**:

$$m\frac{d^2}{dt^2}\langle x\rangle = \frac{d\langle p\rangle}{dt} = -\langle \nabla V(x)\rangle \tag{239}$$

5.7 The Schrodinger picture

In the Heisenberg picture the states are time independent, that is $\frac{\partial}{\partial t}|\psi\rangle = 0$. We had $A(t) = \mathscr{U}(t)^\dagger A(0)\mathscr{U}(t)$. Let us denote this time dependent $A(t)$ by $\hat{A}_H$ and the initial observable $A(0)$ with $\hat{A}$. Obviously $\frac{\partial \hat{A}}{\partial t} = 0$. Any observable corresponding to an operator $\hat{A}$ will be

$$A(t) = \left\langle \psi(0)|\hat{A}_H|\psi(0)\right\rangle = \left\langle \psi(0)|\mathscr{U}(t)\hat{A}\mathscr{U}(t)|\psi(0)\right\rangle \tag{240}$$

We define the Schrodinger picture as time-independent operators and time-dependent states:

$$\begin{aligned} |\psi(t)\rangle &= \exp\left(\frac{iHt}{\hbar}\right)|\psi(0)\rangle \\ A(t) &= \left\langle \psi(t)|\hat{A}|\psi(t)\right\rangle \end{aligned} \tag{241}$$

It follows that

$$\frac{\partial}{\partial t}|\psi(t)\rangle = \frac{1}{i\hbar}H|\psi(t)\rangle \tag{242}$$

which is **Schrodinger's equation**, usually written as

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle \tag{243}$$