Mathematical Foundations of Quantum Mechanics

BY ARNAV WADALKAR
NIT ROURKELA, DEPARTMENT OF PHYSICS AND ASTRONOMY

CONTENT

1. HILBERT SPACE FORMALISM

Quantum systems display features—superposition, interference, and intrinsic uncertainty—that cannot be captured within classical phase space. This chapter develops the mathematical framework that naturally accommodates these phenomena: the Hilbert space. We motivate the need for complex vector spaces and inner products from experimental principles, introduce completeness and the structure of dual vectors, and show how orthogonality and projections encode physical measurements. The chapter culminates in the formal description of quantum states as rays in a projective Hilbert space, providing the geometric and algebraic backbone of quantum mechanics.

1.1 Physical Motivation for Linearity and Superposition

In Classical Mechanics, the 'state' of a system is a **function in phase space**. The reason this couldn't be carried forward to Quantum Mechanics has two major reasons: **the superposition principle and the uncertainty relation.**

The mathematical *change occured in the probability theory* of both worlds. Classically, we know it to be true for mutually exclusive events that:

$$P(A \cup B) = P(A) + P(B)$$

Whereas, quantum mechanically, we see that this does not hold anymore, i.e. *probabilities no longer add up*.

$$P = |\psi_A + \psi_B|^2 \neq |\psi_A|^2 + |\psi_B|^2$$

$$P(x) = |\psi_1(x) + \psi_2(x)|^2 = = |\psi_1(x)|^2 + |\psi_2(x)|^2 + 2|\psi_1(x)||\psi_2(x)|\cos\delta(x)$$

Experimentally, this mathematical fact is realised in the **double slit experiment, where we see the amplitudes adding up and not the probability.** This brought forth a phenomenon known as interference.

The double slit experiment showed us patterns of intensity depicting variable intensities, which is possible via linear superposition with varying coefficients. This made a strong requirement that our new space not only needs to support linear superposition, but one that gives us constructive and destructive interference.

The defining property of a vector space is that it is closed under vector addition and scalar multiplication. Hence, if $|\psi_1\rangle$ and $|\psi_2\rangle$ are elements (vectors) in the space $\mathcal{H}$, then by the closure property of $\mathcal{H}$, their linear combination $\alpha|\psi_1\rangle + \beta|\psi_2\rangle$ must also be an element of $\mathcal{H}$. Hence, our search brought us to the conclusion of a linear vector space. To justify why we needed a **complex** linear vector space, we just need to analyse the possible values the probability density could accept if it were strictly real. In the double slit experiment, the wave functions are represented by

$$\psi_1 \to e^{i\phi_1} \text{ and } \psi_2 \to e^{i\phi_2}$$

For probability to be real valued, between 0 and 1, the only acceptable phase differences would be 0 and π, i.e. +1 and -1, in phase and out-of-phase, respectively. What we see experimentally is something different; we observe a continuous pattern of intensities corresponding to the existence of a **continuous range of phases, i.e. a continuous rotational symmetry**. A real vector space

cannot represent a continuous U(1) space. This gave rise to the need for *complex coefficients*. Therefore, our final requirement is constructed as a complex linear vector space.

Superposition Principle

If $|\psi_1\rangle$ and $|\psi_2\rangle$ are physically allowed states, then any linear combination of them:

$$|\psi\rangle = c_1 |\psi_1\rangle + c_2 |\psi_2\rangle$$

is a physically allowed state.

This principle is the most important principle of Quantum Mechanics since it allows a state to be in a superposition of two states, *giving the state freedom to be in both states at once with varying probabilities,* which was the experimental result. Further, we require a complex linear vector space which is capable of computing the probability in some form of product. Therefore, these requirements mathematically and experimentally lead us to Hilbert space.

1.2 Inner Product Spaces and Hilbert Space

A Hilbert space is a Complete Inner Product space. To understand this statement, we need to understand what an Inner Product Space is and what it means for a space to be complete.

1. What is an Inner Product Space? -

An **Inner Product** is a map $\langle \cdot , \cdot \rangle : V \times V \to C$ satisfying:

- Conjugate Symmetry:$\langle x, y\rangle = \overline{\langle y, x\rangle}$
- Linearity: $\langle a\,x + y, bz\rangle = a * b\langle x, z\rangle + b\langle y, z\rangle$
- Positive Definiteness: $\langle x, x\rangle \geq 0$, and
- $\langle x, x\rangle = 0$ if and only if $x = 0$

where an Inner product space is a Normed linear space with its norm defined as*:*

$$\|v\| = \sqrt{\langle v, v\rangle}$$

Further, a Normed linear space is a Metric space with its metric function defined as:

$$d(v, w) = \|v - w\|$$

where a metric function has to follow the following properties:-

- Positive: $d(x, y) \geq 0$
- Non-degenerate: $d(x, y) = 0$ if and only if $x = y$,
- Symmetric: $d(x, y) = d(y, x)$

- Triangle Inequality: $d(x, z) \leq d(x, y) + d(y, z)$

2. What does it mean for a Space to be Complete? -

Let (a_n) be a sequence in a metric space (X, d). The sequence $(a_n)$is said to be Cauchy. If for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(a_n, a_m) < \epsilon$ whenever $n, m > N$.

If every Cauchy sequence in a metric space X converges to an element of X, then X is said to be complete.

Cauchy-Schwarz inequality

Let $(\mathscr{H}, \langle \cdot , \cdot \rangle)$ be a real or complex inner-product space. For all vectors $|\psi\rangle, |\phi\rangle \in \mathscr{H}$

$$|\langle \psi, v\phi \rangle| \leq \|\psi\| \, \|\phi\|$$

Where the equality holds if and only if u and v are **linearly dependent** on each other.

In Hilbert space, the inequality takes the form:

$$|\langle \psi, \phi \rangle|^2 \leq \langle \psi | \psi \rangle \langle \phi | \phi \rangle$$

This result is crucial to justify Born's Probability theorem we will discuss later.

Geometrically, the theorem allows us to define angles as follows:

$$|\langle \psi, \phi \rangle|^2 = cos^2\theta \rightarrow \cos\theta = \frac{\mathrm{Re}\langle \psi | \phi \rangle}{\|\psi\| \, \|\phi\|}$$

Here, we observe that for $\theta = 0$, the states are identical and for $\theta = \pi/2$, we get orthogonal states. Manipulating the theorem further, we can reconstruct the triangle inequality from it:

$$\|\psi + \phi\| \leq \|\psi\| + \|\phi\|$$

This justifies that Hilbert space has **proper Euclidean geometry.**

Hilbert space

A Hilbert space $\mathscr{H}$ is a complex vector space equipped with an inner product $\langle \cdot , \cdot \rangle$ such that:

1. $(\mathscr{H}, \langle \cdot , \cdot \rangle)$ is an Inner product space

2. $\mathscr{H}$ is complete with respect to the norm

$$\|x\| = \sqrt{\langle x, x \rangle}$$

If $\mathscr{H}$ also contains a countable orthonormal basis, it is said to be separable.

1.3 Dual Space and the Riesz Representation Theorem

The topological Dual Space, denoted by, $\mathscr{H}^*$ is a vector space of all **bounded linear functionals** on $\mathscr{H}$. The central theorem connecting Hilbert spaces and their duals is the Riesz Representation Theorem.
Let $\mathscr{H}$ be a Hilbert space over the field $\mathbb{C}$. Denoted by $\mathscr{H}^*$, the space of all continous linear functionals : $\mathscr{H} \to \mathbb{C}$

Riesz Representation Theorem

For every continous linear functional

$$f \in \mathscr{H}^*$$

There exists a **unique** vector $y \in \mathscr{H}$ such that $\forall x \in \mathscr{H}$

$$f(x) = \langle x, y \rangle$$

Moreover;

1. Uniqueness: If $f(x) = \langle x, y \rangle = \langle x, y' \rangle \ \forall x$, then $y = y'$

2. Isometric Isomorphism: The mapping:

$$\Phi : H \to H^*, \quad y \mapsto f_y, \quad f_y(x) = \langle x \,|\, y \rangle,$$

is an anti-linear isometric isomorphism for complex Hilbert Spaces.

Using Dirac Notation, we can **physically distinguish** vectors in Hilbert space from their duals.

The Ket $|\,\psi\rangle$: A vector in $\mathscr{H}$

The Bra $\langle\psi\,|$: A vector of the Dual Space $\mathscr{H}^*$.

Therefore, a ket is a vector $|\,\psi\rangle \in \mathscr{H}$ **,** and its corresponding bra vector $\langle\psi\,|$ are linear functional:

$$\langle\psi\,| : \mathscr{H} \to \mathbb{C}, \quad |\,\phi\rangle \mapsto \langle\psi\,|\,\phi\rangle$$

Extending the theorem's statement, we can also conclude that the mapping:

$$\mathscr{H} \to \sim \mathscr{H}^*, \quad y \mapsto f_y, \quad f_y(x) = \langle x, y \rangle$$

states that every ket vector $|\,y\rangle \in \mathscr{H}$ corresponds to a unique bra $\langle y\,| \in \mathscr{H}^*$ such that

$$\langle y \mid x \rangle = \langle x \,|\, y \rangle$$

Hence, bras and kets are symmetrical; *it's as if one is viewing the same object through a primal(vectors) and a dual(linear functional) perspective*. Note that due to this symmetry, we are not allowed to use a Banach space to represent Quantum states since a Banach space doesn't contain a complete inner product space or symmetry.
Also, one must note the **antilinearity of bra**. Under the Dirac convention, the exchange

$$|\psi\rangle \rightarrow \langle\psi|$$

is antilinear.

For scalars $\alpha, \beta \in \mathbb{C}$ and vectors $|\phi\rangle, |\psi\rangle \in \mathscr{H} \Rightarrow \langle\phi|, \langle\psi| \in \mathscr{H}^*$

$$\langle\alpha\phi + \beta\psi| = \alpha^* \langle\phi| + \beta^* \langle\psi|$$

In finite finite-dimensional Hilbert space, if $|x\rangle$ is bounded, so is $f(x)$. This is not the case for infinite-dimensional Hilbert spaces. An arbitrarily small change in x can cause a large change in $f(x)$. Meaning, there are vectors x_n with $\|x_n\| \leq 1$ but $|f(x)| \rightarrow \infty$.

Since the Riesz Representation Theorem is applicable only for **bounded linear functionals**, we must only use bras corresponding to bounded linear functionals. If not, the Dual space would be unbounded, extremely large compared to the Primal Space, and isomorphism would break.

In Quantum Mechanics, **kets represent the quantum state,** whereas **bras represent linear functionals**, or the amplitude of the state. Through this, we understand the physical significance of boundedness since a slight change in state could give an abrupt, large change for non-bounded functionals. What we desire is that an infinitesimal change in the state should give an infinitesimal change in the measurement, making the physical quantities **stable under small perturbations**. We will show how this comes to be in the later sections.

Orthogonal Expansion in state bases

Let $|x_1\rangle, |x_2\rangle, \ldots$ be a basis in a Hilbert space $\mathscr{H}$. Then, for any vector $|x\rangle \in \mathscr{H}$, the expansion:

$$|x\rangle = \sum_n \langle x_n, x\rangle |x_n\rangle$$

holds.

Note that Quantum Mechanical Operators are unbounded operators, but they map $\mathscr{H} \rightarrow \mathscr{H}$, not $\mathscr{H} \rightarrow \mathbb{C}$. We will discuss operators and their theory in rigour in the later sections.

1.4 Closed Subspaces, Orthogonality, and Projections

A subset $V \subset \mathscr{H}$ is a **closed linear subspace** only if it upholds the two properties:

1. Linearity: It is closed under linear combination

2. Completeness: It contains all its limit points. If a sequence of vectors $|v_n\rangle \in V$ converges to $|v\rangle$, then $|v\rangle$ must also be in V.

In Quantum mechanics, we almost always care about **closed subspaces** because projection operators require them.

Orthogonal Projection Theorem

Let $\mathscr{H}$ be a Hilbert Space and let $\mathscr{M}$ be a **closed linear subspace** of $\mathscr{H}$. For every vector $|x\rangle \in \mathscr{H}$, there exists a **unique** vector $|m\rangle \in \mathscr{M}$ and $|n\rangle \in \mathscr{M}^\perp$ such that:

$$|\, x\rangle = |\, m\rangle + |\, n\rangle$$

where $\mathscr{M}^{\perp}$ is the orthogonal complement of $\mathscr{M}$, i.e. the set of all vectors orthogonal to every vector in $\mathscr{M}$.
The key consequence of this theorem is that we can decompose the Hilbert space as a direct sum of two subspaces, namely the subspace and its **orthogonal complement:**

$$\mathscr{H} = \mathscr{M} \oplus \mathscr{M}^{\perp}$$

Hence, we can exhaustively split states as they are either inside a subspace or outside of it.

Here, we introduce(defined later) the operator $P : \mathscr{H} \to \mathscr{H}$ called the Projection operator. The mapping $P : x \to x_0$ is linear. Moreover $P = P^2$, the physical implication of $P = P^2$ is that *if you project the same state on something twice, it will give you the same result as it would be if you did it once*. The reason for only accepting closed linear spaces is that the **closedness** of the subspace is required to guarantee the existence of the projection.

Furthermore, **the projection operator is a self-adjoint operator;** as a consequence of the projection operator has to split the Hilbert space into two distinct parts, namely, a subspace where the answer lives and another where it does not. This enforces the condition of orthogonal projections at all times, requiring orthogonal eigenspaces and eigenvalues 1 and 0.

$$\langle Px\,|\,y\rangle = \langle x\,|\,Py\rangle$$

We now define the **Range** of the projection operator:

$$V = \text{Range}(P) = \{\,|\psi\rangle \in \mathscr{H} \mid |\psi\rangle = P\,|\phi\rangle \ \text{ for some } |\phi\rangle \in \mathscr{H}\}$$

Correspondingly, its orthogonal complement is defined to be:

$$V^{\perp} = \{\,|\chi\rangle \in H \mid \langle\chi\,|\,v\rangle = 0 \text{ for all } |v\rangle \in V\}$$

Hence, it is correct to say if P is the projector onto V, then $Q = I - P$ is the projector onto $V^{\perp}$.

From the theory of linear algebra, the subspace $V^{\perp}$ is the null space of the operator $\hat{P}$, and the space decomposes as:

$$\mathscr{H} = \text{Range}(P) \oplus \text{Ker}(P)\,.$$

Through this, we can write any vector **uniquely** as a sum of two orthogonal projection operators:

$$|\psi\rangle = P\,|\psi\rangle + (I - P)\,|\psi\rangle$$

Further, we define the norm of the above equation, where P lies in V, and $I - P$ lies in $V^{\perp}$:

$$\|\psi\|^2 = \|P\psi\|^2 + \|(I-P)\psi\|^2 \Rightarrow \|P\psi\| \leq \|\psi\| \implies \|P\| \leq 1$$

Hence, the *boundedness of the projection operator* is proven, meaning **unbounded projectors cannot exist**.

For normalised wavefunctions:

$$\|\psi\| = 1 \implies \|P\psi\| = \|\psi\| \implies \|P\| = 1$$

At this stage, projection operators are purely geometric objects, where *Px is the nearest point to M and $x - Px$ is the shortest vector from x to any point in M*. Their interpretation as measurements will be developed later. We will develop a more general form of measurement, one which is unbounded and not self-adjoint in Measurement theory.

1.5 Quantum States

To connect the quantum theory with the mathematical formalism, we start by defining a quantum state using Hilbert space. Classically, we define physical states of a system via Phase space, where each '*point'* defines the momentum and position of the particle and state corresponding to a linear functional in Phase Space.

The reason we cannot use Phase space and had to extend towards Hilbert space, and why we can no longer describe States as linear functionals in Phase Space, rather as 'vectors (or more generally, rays)', has a physical intuition and a deeper algebraic reasoning.

The quantum state of an isolated physical system is completely specified by a unit ray in a complex, separable Hilbert space $\mathscr{H}$. Every non-zero vector $|\psi\rangle$ in this space corresponds to a possible physical state.

Physically, we can interpret one of the incompatibilities via **Heisenberg's uncertainty principle** for conjugate canonical variables, which states that there should in any scenario exist a **minimum** amount of error in measurement between position and momentum.

$$\Delta q \Delta p \geq \hbar/2$$

Hence, rather than a linear functional, our state becomes **delocalised in Phase Space**. To explain why the vectors are the correct mathematical description of States is explained via the superposition principle of Quantum systems, where we need a mathematical structure allowing linear combination.

Also, since the *amplitudes must interfere,* we require an inner product structure. As discussed earlier, the experimental justification of why Quantum Mechanics requires a complex linear vector space is given by the double slit experiment. However, there has been no justification for why vectors are the appropriate tool required to represent states.

Mathematically, the answer lies within the algebraic concepts of Commutative Algebra:-

Gelfand Representation Theorem
If you have a commutative C^*-algebra of observables, the "states" of that system are mathematically isomorphic to **points** in a topological space (Phase Space).

Gelfand-Naimark-Segal Theorem
Given an C^*-algebra and a state, the GNS theorem constructs a Hilbert space representation. Every abstract C^*-algebra of non-commuting observables can be represented as **operators** acting on a Hilbert Space.

Hence, by the Gelfand Representation, classical states are **positive linear functionals** on the Phase Space. In a Quantum system, the operation of measuring position and the operation of moving the system interfere with each other; mathematically, the commutator of the respective operators is non-zero, i.e. non-commutative.
The Physical reason is that the $\hat{x}$ and $\hat{p}$ are **Fourier Duals**. Hence, they are the same object viewed through different domains, respectively, position and time. Hence, by the Gelfand-Naimark-Segal Theorem, we require the Hilbert space representation.

To be specific, states are not vectors; they represent a **class of vectors or rather, 'Rays'**. The physical interpretation of any vector is via its measurement. As discussed earlier, the bras represent the amplitude, and the kets represent states. If we modify our state vector $|\psi\rangle$ as:

$$|\psi'\rangle = e^{i\alpha}|\psi\rangle$$

Using Born's Probability rule to measure the probability, the **global phase vanishes**:

$$\langle\phi|\psi'\rangle = \langle\phi|e^{i\alpha}\psi\rangle = e^{i\alpha}\langle\phi|\psi\rangle$$

$$|\langle\phi|\psi'\rangle|^2 = |\langle\phi|\psi\rangle|^2$$

Thus, all vectors differing by a global phase are physically equivalent.

Equivalence Relation

Defining an equivalence relation on $\mathscr{H}\backslash\{0\}$

$$|\psi\rangle \sim |\phi\rangle \text{ iff } |\phi\rangle = c|\psi\rangle, c \in \mathbb{C}\backslash\{0\}$$

Hence, our Quantum state is an equivalence class defined as:

$$[\psi] = \{e^{i\alpha}|\psi\rangle \mid \alpha \in [0,2\pi)\}$$

Therefore, we define a Projective Hilbert space $\mathbb{P}(\mathscr{H})$ as a set of all **rays.** Hence, **every quantum state lives in a projective Hilbert space** becuase the state's global phase is **unobservable**. This also changed the geometry of the space, where the distance is:

$$d_{FS}(\psi,\phi) = arccos|\langle\psi,\phi\rangle|$$

where a projective Hilbert space carries the **Fubini–Study metric.**

Quantum State

A physical state is a positive, normalised linear functional:

$$\omega : \mathscr{A} \to \mathbb{C}$$

on the C-algebra of observables.*

Pure states are extremal points of the convex set of states. For quantum systems $\mathscr{A}$ is non-commutative. Therefore, by GNS construction, every pure state corresponds to a ray

$$[\psi] = \{e^{i\alpha} \mid \psi\rangle \mid \alpha \in [0,2\pi)\}$$

In Hilbert space. The collection of all such rays forms the projective Hilbert space $\mathbb{P}(\mathscr{H})$.
Note that until now, we have only defined and discussed pure states. Mixed states represented by Density matrices are not rays and do not live in $\mathbb{P}(\mathscr{H})$. The formulation of mixed states will be done in Chapter 4.

More generally, a pure or mixed quantum state is a positive, normalised linear functional:

$$\omega(A) = \langle \psi, A\psi \rangle$$

on the noncommutative C*-algebra of observables. Further, by GNS, every such functional corresponds to a Hilbert space.

2. LINEAR OPERATORS IN HILBERT SPACE

This chapter develops the operator framework underlying quantum mechanics. We introduce linear operators as the only transformations compatible with superposition, then distinguish bounded from unbounded operators through domain considerations and results such as the Hellinger–Toeplitz theorem. The role of adjoints is examined to expand over and identify the difference between symmetric and self-adjoint operators, which form the mathematical basis of physical observables. Unitary operators are then presented as structure-preserving transformations governing quantum evolution, with Stone's theorem linking continuous unitary dynamics to self-adjoint generators. Together, these ideas provide the core operator-theoretic foundations for quantum measurement, dynamics, and symmetry.

2.1 Linearity of Operators

In the previous section, we stated that a Hilbert space is a Linear Complex vector space (or a Complete Inner product space).
To measure, evolve, or rotate, i.e. any action on the Hilbert space, is done by operators that preserve the structure of the space (linear superposition) acting on the abstract states. The operators which preserve the structure are called Linear Operators. They are defined as follows:

Linear Operators

Linear Operators are linear mappings that transform one vector in a Hilbert space into another vector in the same space.

$$\hat{A} : \mathscr{D}(\hat{A}) = \mathscr{H} \to \mathscr{H}$$

Further, an operator is linear if and only if:

1. Preservation of vector addition:

$$A(\,|\,\psi\rangle + |\,\phi\rangle) = A\,|\,\psi\rangle + A\,|\,\phi\rangle$$

2. Preservation of scalar multiplication:

$$A\,|\,\alpha\psi\rangle = \alpha A\,|\,\psi\rangle \text{ where } \alpha \in \mathbb{C}$$

Why is an operator needed in the first place? To answer that, we simply ask ourselves how we can represent a system change from state $|\,\psi(0)\rangle$ to state $|\,\psi(t)\rangle$. Mathematically, to preserve the structure, any operation acting has to be linear in nature, since if two systems A and B evolve in time to A' and B'. The superposition of A' and B' should also be a part of the evolved Hilbert space.

Hence, without an Operator acting on the space, our space would be static, and there would be no motion, just a collection or a set of complex vectors. To connect the Vectors to the physical phenomena, we use operators, making them crucial to the theory of Quantum Mechanics. Also, quantum superposition is always preserved by transformations caused by linear operators, which can be restated as:

All physically meaningful state evolutions must be carried out by linear operations on the Hilbert space.

Note that different kinds of operators—bounded, self-adjoint, unitary, projection—describe different physical processes. Linearity is the universal requirement, but additional structure is added depending on the role of the operator.

2.2 Bounded and Unbounded Operators

The domain $\mathscr{D}(A)$ of a linear operator A need not be equal to the entire Hilbert space $\mathscr{H}$. This is crucial becuase the Hellinger–Toeplitz theorem states that any operator which is symmetric and defined on the entire Hilbert space $\mathscr{H}$ must be bounded.

The Hellinger-Toeplitz Theorem

If a linear operator $\hat{A}$ is defined on the ***entire*** **Hilbert space and is symmetric**:

$$\langle \phi \mid A\psi \rangle = \langle A\phi \mid \psi \rangle$$

Then $\hat{A}$ **must** be bounded.

There exist many crucial symmetric operators which are not bounded, for example, the position operator $\hat{x}$, the momentum operator $\hat{p}$, the Hamiltonian $\hat{H}$, etc. Hence, we classify operators as such:

A] Bounded Operator

A linear operator A with the domain $\mathscr{D}(A) \subset \mathscr{H}$ is bounded if there exists a number $M \geq 0$, such that $\forall \mid \psi \rangle \in \mathscr{D}(A)$

$$\|A\psi\| \leq M\|\psi\|$$

where the norm of A is defined as:

$$\|A\| = inf\{M \geq 0 : \|A\psi\| \leq M\|\psi\|, \forall \psi \in \mathscr{H}\}$$

Hence, by definition $\|A\| \leq \infty$, iff A is bounded.

In Hilbert space, Boundedness is synonymous with continuity.

If an operator is bounded, it maps converging sequences to converging sequences.

Therefore, Bounded Operators represent 'finite operators'. Also, a bounded operator can be defined on the **entire** Hilbert Space $\mathscr{H}$, giving us the freedom to apply a bounded operator to any state vector without worrying about the domain restrictions; there is no need to define a dense domain.

Further, the set $\mathfrak{B}(\mathscr{H})$ of all bounded linear operators on $\mathscr{H}$, equipped with a norm and adjoint, is a Banach algebra.

B] Unbounded Operator

For a linear operator A with the domain $\mathscr{D}(A) \subset \mathscr{H}$, if there exists no $M \geq 0$ such that $\forall \mid \psi \rangle \in \mathscr{D}(A)$,

$$\|A\psi\| \leq M\|\psi\|$$

Then the operator is said to be **unbounded or discontinuous**. An unbounded operator cannot be defined on the entire Hilbert space; rather, it is defined on a dense subset of it.

Dense subset in $\mathscr{H}$

$\forall x \in \mathscr{H}$ and $\forall \epsilon > 0,$ there exists $d \in \mathscr{D}$ such that

$$\|x - d\| < \epsilon$$

Therefore, every element of $\mathscr{H}$ is a limit of a sequence of elements from $\mathscr{D}$.

Since observables must be represented by **self-adjoint** operators, and adjoints can only be defined for densely defined operators, this requires that both A and A^* to be densely defined. Hence, this imposes on the non-bounded operators to have densely defined domains. Tho this may be the physical interpretation, a mathematically sound explanation requires graph theory to explain the reason for a dense domain of the non-bounded operators.

Closed Graph Theorem

If a linear operator $T : \mathscr{H} \to \mathscr{H}$ is defined on all of $\mathscr{H}$ and has a closed graph, then T must be bounded.
Therefore, any unbounded operator cannot be defined on all of $\mathscr{H}$ while remaining closed. Since physical operators must be closed (or closable), they must be defined on a proper, usually dense, subset of H.

In quantum mechanics, unbounded operators representing observables must be closed or at least closable. Without closedness, limits of physical states would produce inconsistent results. A linear operator defined on all of $\mathscr{H}$ cannot be both closed and unbounded (Closed Graph Theorem), so the need for a proper dense domain is mathematically unavoidable.

2.3 Adjoint Operators

In Hilbert space, the adjoint operator $A^\dagger$ shifts the action of the linear operator from one side of the inner product to the other. Formally, the adjoint operator $A^\dagger$ is defined as the operator that satisfies:

$$\langle \psi, A\phi \rangle = \langle A^\dagger \psi, \phi \rangle \; \forall \; \phi \in \mathscr{D}(\hat{A}) \text{ and } \forall \; \psi \in \mathscr{D}(\hat{A^\dagger})$$

The domain $\mathscr{D}(\hat{A^\dagger})$ is the set of all vectors ψ such that there exists a vector $\chi_\psi \in \mathscr{H}$

$$\langle \psi, A\phi \rangle = \langle \chi_\psi, \phi \rangle \; \forall \; \phi \in \mathscr{D}(\hat{A})$$

Here, $A^\dagger \psi = \chi_\psi$

Symmetric (Hermitian) Operator

An operator is symmetric (Hermitian) if:

$$\langle \psi, A\phi \rangle = \langle A\psi, \phi \rangle; \forall \psi, \phi \in \mathscr{D}(A)$$

Symmetry operators are domain-dependent, i.e. the same operator may be symmetric on one domain but not on another. Also, symmetric operators may fail to have real eigenvalues.

Self-adjoint Operator

A linear operator is called self-adjoint if it satisfies two conditions:-

1. Hermitian:

$$\langle \psi, A\phi \rangle = \langle A\psi, \phi \rangle; \forall \psi, \phi \in \mathscr{D}(A)$$

2. Domains of the operator and the adjoint are equal:

$$\hat{A} = \hat{A}^{\dagger} \text{ and } \mathscr{D}(\hat{A}) = \mathscr{D}(\hat{A^{\dagger}})$$

Self-adjoint operators are domain independent, i.e. a global operator property, and by Stone's theorem, only self-adjoint operators are guaranteed to generate unitary transformations; this is discussed ahead.

The domains of the operator and adjoint perfectly coinciding is always true for finite finite-dimensional Hilbert space. Hence, all symmetric operators are self-adjoint since the condition of domain restriction is already fulfilled, but when we take into consideration the infinite-dimensional Hilbert space, $\mathscr{D}(\hat{A^{\dagger}})$ **is usually bigger than** $\mathscr{D}(\hat{A})$. Therefore, self-adjointness is a stricter condition than hermiticity.

One of the crucial properties of a self-adjoint operator is that its **eigenvalues are always real and its eigenvectors orthogonal.** This allows it to represent all the forms of measurements carried out in Hilbert space. **Hence, physical observables are always self-adjoint operators.**

To represent any physical observable, we require it to produce real eigenvalues on measurement. Due to this constraint, we only have one type of operator capable of fitting the requirement, self-adjoint operators, since only self-adjoint operators guarantee that all eigenvalues are real. Further elaboration of why self-adjoint operators are the ideal operators for physical observables will be understood via Spectral theory.

2.4 Unitary Operators

An operator $U : \mathscr{H} \to \mathscr{H}$ is Unitary if it satisfies the following two properties:

1. Isometry: It preserves the Inner product(thus the norm as well):

$$\langle U\psi, U\phi \rangle = \langle \psi, \phi \rangle \implies \| U\psi \| = \| \psi \|; \forall \psi \in H$$

2. Operator Equality: $U^{\dagger} = U^{-1}$

3. Surgective Mapping: Range(U) = $\mathscr{H}$

As self-adjoint operators have special properties regarding their eigenvalues, so do Unitary operators. If $U|\psi\rangle = \lambda|\psi\rangle$, then we can say $|\lambda| = 1$, which can be represented as $\lambda = e^{i\theta}$. Hence, **the full spectrum lies on the unit circle.**

$$\sigma(U) \subseteq \{z \in C :| z | = 1\}$$

Note that an operator satisfying only $U^\dagger U = I$ is called isometric, whereas a unitary also requires

$$UU^\dagger = I$$

Therefore, an isometry is unitary if and only if it is surjective.

Stone's Theorem

There exists a unique **Self-Adjoint operator** H for a **strongly continuous one-parameter Unitary group** $\{U(t)\}_{t\in\mathbb{R}}$ **on** $\mathscr{H}$, such that:

$$U(t) = e^{-itH}, t \in \mathbb{R}$$

Conversely, for every self-adjoint operator H, the operators:

$$U(t) = e^{-itH}$$

form a strongly continuous one-parameter unitary group.

The mathematical condition for being strongly continous is:

$$\lim_{t\to 0} \|U(t)\psi - \psi\| = 0 \quad \forall \psi \in H.$$

This also justifies why **Hermiticity for an operator is not a sufficient condition, and self-adjoint is the required and sufficient condition.**

3. Spectral Theory

Chapter 4 explains spectral theory as the infinite-dimensional version of diagonalising an operator. It starts from $L^2(X,\mu)$ as the natural Hilbert space for normalisable quantum states and shows why standard matrix diagonalisation breaks down for unbounded operators and continuous spectra. To fix this, the chapter introduces the spectrum $\sigma(A)$ and resolvent $R_\lambda(A) = (A - \lambda I)^{-1}$, defining spectral points as those where this inverse fails to exist or is unbounded. The spectrum is then split into point, continuous, and residual parts: the point spectrum gives genuine eigenvalues and bound states; the continuous spectrum captures scattering states via Weyl sequences; the residual spectrum is the leftover, where the range of $A - \lambda I$ is not dense.

3.1 L^2 Space and Diagonalisation

The space $L^2(X,\mu)$ is defined as:

$$L^2(X,\mu) = \left\{ f : X \to \mathbb{C} \;\middle|\; f \text{ is measurable and } \int_X |f(x)|^2 \, d\mu(x) < \infty \right\}$$

where the norm corresponds to:

$$\|f\|_2 = \left(\int_X |f(x)|^2 \, d\mu(x) \right)^{1/2}$$

Hence, with an equipped inner product, $L^2(X,\mu)$ becomes a Hilbert space as follows:

$$\langle f, g \rangle = \int_X f(x)^* g(x) \, d\mu(x)$$

In Quantum mechanics, the conditions correspond to:

$$L^2(\mathbb{R}^n) = \left\{ \psi : \mathbb{R}^n \to \mathbb{C} \;\middle|\; \int_{\mathbb{R}^n} |\psi(x)|^2 \, d^n x < \infty \right\}$$

where $\|\psi(x)\|^2$ corresponds to the probability density by Born's Interpretation. Hence, this states that the total probability is finite, which allows us to convert it to 1 after normalisation.

We can restate that the set of all normalisable states corresponds to $L^2(\mathbb{R}^n)$. Therefore, all physically realisable (normalizable) quantum states live in $L^2(\mathbb{R}^n)$, but operators like position or momentum can have generalised eigenvectors that **do not live in** $L^2(\mathbb{R}^n)$. The issue arises in Hilbert space when one tries to diagonalise operators which do not live in $L^2(\mathbb{R})$, i.e. not normalisable or when one tries to diagonalise unbounded operators, since they do not contain a full set of eigenvectors belonging to the Hilbert space.

To tackle the problem, we first need to understand what diagonalisation is and why it fails in Infinite-dimensional Hilbert spaces.

Diagonalisation in Finite Dimensions
A matrix $A \in \mathbb{C}^{nxn}$ is diagonalizable if there exists a basis of eigenvectors $\{ \mid v_i\rangle \}$ such that:

$$A \mid v_i\rangle = \lambda_i \mid v_i\rangle \ \forall i = 1,.....,n$$

and on this basis, the matrix becomes diagonal:

$$A = V\Lambda V^{-1}, \text{where } \Lambda = diag(\lambda_1, \ldots, \lambda_n)$$

Diagonalisation is a process by which we find a basis in which the operators act like eigenvectors. Note that when operators contain continuous spectra, one cannot 'sum' over the eigenvalues. The method of diagonalisation for finite dimensions cannot be used here, requiring us to develop a more generalised method to diagonalise the basis, which can allow us to decompose the state into 'eigencomponents' for all self-adjoint operators(we only work with self-adjoint since they are the only set of operators that guarentee a complete real spectra), be it with continous spectra or unbounded operators. This generalisation is provided by Spectral theory.

3.2 Spectrum and Resolvent

Now that the physical requirement of the theorem is put forth, we will formulate the mathematical problem:

Given a self-adjoint operator A on a Hilbert space $\mathscr{H}$, find a decomposition of the Hilbert space into its 'spectral subspaces' mathematically corresponding to subsets of the spectrum $\sigma(A)$, such that A acts like an eigenvector on each subspace, and the functions of A and measurement projections of A are defined.

The linear algebra problem leading to the generalisation is:

For a linear operator A on Hilbert space $\mathscr{H}$:

$$A \mid \psi\rangle = \lambda \mid \psi\rangle$$

in a finite-dimensional space, the bracket $(A - \lambda I)$ can only be zero if λ is an eigenvalue of A, making it non-invertible. When we talk about infinite dimensions, an operator can fail to invert even if λ is not an eigenvalue since $|\psi\rangle$ may not exist in $\mathscr{H}$. Therefore, we cannot just talk about eigenvalues.

We rewrite the equation as:

$$(A - \lambda I)\,|\psi\rangle = 0$$

If λ is an eigenvalue, $A - \lambda I$ is not invertible on $\mathscr{H}$. If λ is not in the spectrum, $A - \lambda I$ is invertible if its inverse is bounded and its resolvent is:

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

Hence, we can say, spectral points are exactly the λ where the resolvent fails to exist or is unbounded. We must define the **Spectrum** $\sigma(\hat{A})$ as the set of *all* λ where our operator is non-invertible.

On the other hand, **the resolvent set** $\rho(\hat{A})$ is the set containing all the values for which the equation can be uniquely solved. Rigorously, a complex number λ is in the resolvent set if the operator $(\hat{A} - \lambda \hat{I})$is a **bijection** with a **bounded inverse.**

Therefore, **the spectrum set** $\sigma(\hat{A})$ is the complement of the resolvent set, where the inverse $\hat{R}_\lambda$ either does not exist or is unbounded

$$\sigma(A) = \left\{ \lambda \in \mathbb{C} \;\middle|\; A - \lambda I \text{ is not boundedly invertible} \right\}$$

3.3 Classification of Spectra

The spectral set can be exhaustively divided into three categories:

1. Point Spectrum

These are the eigenvalues of the operator; they are not injective, stating there exists a non-zero function such that it maps to zero:

$$(A - \lambda I)\,|\psi\rangle = 0 \implies A\,|\psi\rangle = \lambda\,|\psi\rangle$$

These specify the bound states where the particle is trapped or localised, and the state

$$|\psi\rangle \in L^2(\mathbb{R}^n) \Rightarrow \lim_{|x|\to\infty} \psi(x) = 0$$

The values are points because only point values of E allow the function to hit zero for the limit.

2. Continous Spectrum

These are the values that have a one-to-one mapping with respect to $(A - \lambda I)$, but the inverse mapping is unbounded. Meaning there is no single vector in this set that satisfies $A\,|\psi\rangle = \lambda\,|\psi\rangle$, but there exists a sequence of unit vectors $|\psi_n\rangle$ that get arbitrarily close:

$$\lim_{n\to\infty} \|(A - \lambda I)\,|\psi_n\rangle\| = 0$$

These are called ‘Weyl Sequences’.

The wavefunction does not vanish at infinity. Instead, it behaves like a plane wave, corresponding to an infinite area under the curve, i.e non-normalisable, these states represent scattering states. This implies that the state lives outside the Hilbert space, making it an invalid eigenstate. This collection of energies forms the continous spectrum.

2. Residual Spectrum

The last one of the three is the residual spectrum, corresponding to the leftover of the residual set when one takes out the discrete and continous spectrum. Here, the range of $(A - \lambda I)$ is not dense.

For $\lambda \in \sigma_r(A)$, therefore, the adjoint $A^\dagger$ has:

$$ker(A^\dagger - \lambda^* I) \neq \{0\}$$

For all self-adjoint operators on Hilbert space, the residual spectrum is empty. Therefore, if $A = A^\dagger$ is self-adjoint, then:

$$\sigma_r(A) = \varnothing$$

Therefore, non-empty residual sets appear in non-Hermitian systems.

3.4 Spectral Measures and Projection-Valued Measures

Before going into the rigour of Spectral theory, we will understand what measurement is briefly; the core of measurement theory in Quantum Mechanics requires its own separate chapter.

Borel Set
A Borel set is any set you can build starting from open intervals using countable unions, countable intersections and complements.

$$\mathscr{B}(\mathbb{R}) = \sigma(\text{all open intervals})$$

Therefore, Borel sets are the allowed subsets of $\mathbb{R}$ one can 'measure' in making them the largest family of subsets of $\mathbb{R}$ where probability can be consistently defined corresponding to physically meaningful measurable regions of possible outcomes of an observable.

Consider an observable A with an eigenvalue λ and its corresponding eigenspace E_λ. Suppose, on measurement, it returns the value λ, then the post-measurement state must lie entirely in E_λ , i.e. obtained by projecting the original state onto E_λ.

$$P_\lambda = |\,\psi_\lambda\rangle\langle\psi_\lambda\,|$$

Further, when one projects the post-measurement state onto E_λ, it should give the same result, stating the property of the projection operators:

$$P = P^2$$

More generally than not, we want to compute the probability of an observable's measure lying in a 'range' or some region, call that region B. Therefore, we need a projection that projects onto the subspace spanned by all the eigenvectors with eigenvalues in B.

Take $\mathscr{H}_B = span\{\psi_\lambda : \lambda \in B\}$ and define the projection operator onto the space B as $E(B)$. In a discrete spectrum:

$$E(B) = \sum_{\lambda \in B} P_\lambda$$

The sum is well-defined only when the spectrum in B is at most countable. Though for a continous spectrum, we cannot sum over the eigenvalues. Here, we need to define the allowed subsets of the space where the projection operator can measure. For this reason, we introduced Borel sets as they give us the family of sets for which measurement can be applied, i.e. projection can be done.

Therefore, we define such a measurement as a projection-valued measure. For continuous spectra, the spectral family becomes an operator-valued measure instead of a sum.

Projection-valued Measure (PVM)
A projection-valued measure is a map:

$$E : \mathscr{B}(\mathbb{R}) \to \mathscr{P}(\mathscr{H})$$

That assigns projections to Borel sets satisfying additivity, normalisation and orthogonality.

Properties of PVM:

1. Normalisation: $E(\mathbb{R}) = 1$ and $E(\varnothing) = 0$

2. Countable Additivity: If $\{B_i\}$ are disjoint Borel Sets:

$$E\left(\bigcup_i B_i\right) = \sum_i E(B_i)$$

3. Orthogonality: $E(B_1)E(B_2) = 0$ if $B_1 \cap B_2 = \varnothing$

4. Projection: $E(B) = E(B)^2 = E(B)^\dagger$

Thus, a PVM assigns to every measurable set of outcomes a projection operator whose expectation value gives the probability of that outcome. This is the mathematical bridge between projections and measurement. The physical name of the set $E(B)$ is called the Spectral family; it is the cumulative probability distribution for the observable.

Spectral Family
A spectral family or resolution of the identity is a one-parameter family of projection operators.

$$\{E_\lambda\}_{\lambda\in\mathbb{R}}$$

satisfying:

1. Monotonicity: $\lambda_1 < \lambda_2 \Rightarrow E_{\lambda_1} \leq E_{\lambda_2}$

2. Right-continuity: $E_\lambda = \lim_{\mu\to\lambda} E_\mu$

3. Limits: $\lim_{\lambda\to-\infty} E_\lambda = 0, \quad \lim_{\lambda\to+\infty} E_\lambda = I$

Given a PVM: $E : \mathscr{B}(\mathbb{R}) \to \mathscr{P}(\mathscr{H})$, we can define the spectral family $\{E_\lambda\}_{\lambda\in\mathbb{R}}$ as follows:

$$E_\lambda := E((-\infty, \lambda]), \quad \lambda \in \mathbb{R}\,.$$

Conversely, given a spectral family $\{E_\lambda\}_{\lambda\in\mathbb{R}}$, we can recover the PVM as follows:

For any Borel set $\Delta \subseteq \mathbb{R}$, $E(\Delta)$is the projection, such that:

$$\langle\psi \,|\, E(\Delta) \,|\, \psi\rangle = \int_\Delta d\langle\psi \,|\, E_\lambda \,|\, \psi\rangle, \quad \forall\psi \in \mathscr{H}\,.$$

Therefore, a PVM gives a spectral family by taking cumulative projections, whereas a spectral family gives a PVM by taking the operator-valued Stieltjes measure.

3.5 Spectral Theorem for Self-Adjoint Operators

Having developed all the physical and mathematical tools, we now define the form of measurement directly used in Spectral theory.

Stieltjes measure
There exists a unique Borel measure μ_F on $\mathbb{R}$ such that for any interval, we define:

$$\mu_F\big((a,b]\big) = F(b) - F(a)\,.$$

Thus, a Stieltjes measure is the measure generated by a monotone function.

Lebesgue–Stieltjes Integral
For a Borel measurable g, the Lebesgue–Stieltjes integral is:

$$\int_{\mathbb{R}} g(x)\, d\mu_F(x)$$

It is a Lebesgue integral with respect to a Stieltjes measure.

The reason the Riemann integral was not applicable here is becuase we do not integrate with respect to a Lebesgue measure dx, rather we integrate with respect to a spectral measure E which can be discrete, continuous, or singular depending on the operator.

Hence, the Riemann integral is not compatible, and we require an integral with measure adhering to discrete, continuous, or singular intervals. The Lebesgue–Stieltjes measure allows the measure to 'jump', be 'flat' or have singular spectra. Another requirement from the integral is to deal with unbounded operators, on $\mathbb{R}$ which the Riemann integral cannot.

Spectral theory for bounded Self-adjoint operators
Let $A \in B(\mathscr{H})$ be self-adjoint. Then there exists a unique PVM E on the Borel sets of $\sigma(A) \subset \mathbb{R}$ such that:

$$A = \int_{\sigma(A)} \lambda \, dE(\lambda),$$

where the integral is an operator integral for all $\psi \in \mathscr{H}$

$$A\psi = \int_{\sigma(A)} \lambda \, dE(\lambda)\psi$$

Spectral theory for unbounded Self-adjoint operators
Let A be self-adjoint and densely defined. There exists a unique PVM E on $\mathbb{R}$ such that for each Borel measurable $f : \mathbb{R} \to \mathbb{C}$, one can define:

$$f(A) = \int_{\mathbb{R}} f(\lambda)\, dE(\lambda)$$

with the domain:

$$\mathscr{D}(f(A)) = \left\{ \psi \in \mathscr{H} : \int_{\mathbb{R}} |f(\lambda)|^2 \, d\langle \psi \,|\, E(\Delta) \,|\, \psi \rangle < \infty \right\}$$

Hence, we have:

$$A = \int_{\mathbb{R}} \lambda \, dE(\lambda)$$

with the domain:

$$\mathscr{D}(A) = \left\{ \psi : \int_{\mathbb{R}} \lambda^2 \, d\langle \psi \,|\, E(\Delta) \,|\, \psi \rangle < \infty \right\}$$

Going back to the initial linear algebra problem, we now show:

$$(A - zI)x = y$$

If $z \in \rho(A)$, the linear equation has a unique solution:

$$x = R(z)y = \int_{\sigma(A)} (\lambda - z)^{-1} \, dE(\lambda) \, y$$

Hence, we have developed a mathematical method to rewrite our operator A as a multiplication by its eigenvalues on a L^2 space regardless of the operator being bounded, unbounded, continous or discrete.

On measurement of A, we get a unique PVM $E(\,\cdot\,)$ on the Borel subset of $\mathbb{R}$. Therefore, this theorem provides us a generalised manner of 'diagonalising' the operator in a basis of eigenvectors.

Generalised Diagonalisation
There is a unitary mapping:

$$U : \mathscr{H} \to \int_{\Lambda}^{\oplus} H_\lambda \, d\mu(\lambda)$$

such that UAU^{-1} acts as a multiplication of λ on $\mathscr{H}_\lambda$.

$$(UAU^{-1}f)(\lambda) = \lambda f(\lambda) \,.$$

This form of continuous diagonalisation is also called spectral decomposition.

4. Statistical States and Measurement Theory

This chapter formulates quantum mechanics as a statistical theory of states and measurements. Quantum states are defined as positive, normalised linear functionals on the algebra of observables, leading to the density matrix representation via trace-class operators. Born's rule is derived using spectral measures, while general measurement processes are described using projection-valued measures, POVMs, and quantum instruments. Measurement dynamics are modelled through system–apparatus coupling, Kraus operators, and completely positive trace-preserving maps, with state update rules interpreted as quantum conditioning on a noncommutative probability space. The chapter clarifies the distinction between pure and mixed states, proper and improper mixtures, and the operational meaning of preparation and measurement.

4.1 Statistical Structure of Quantum States

In reality, we need more than a pure state to measure; we need both the measuring device and the state, which would lead to a mixture of states, destroying the pure state. Hence, we cannot physically measure an isolated pure state. Isolation cannot occur since even the environment surrounding the state would lead to a mixture.

Physical measurement requires a system S, a measuring device M and an environment E. The process of measurement couples the system, making it an entangled state of system+device. Hence, we require something more than a pure state.

A mixed state is a collection:

$$\{p_i, |\psi\rangle\}$$

Where p_i represents classical uncertainty about the underlying quantum states.

Pure State
Let $\mathscr{H}$ be a complex separable Hilbert space. A pure state is an equivalence class of unit vectors $\psi \in \mathscr{H}$ with $\|\psi\| = 1$ having a global invariant phase.

Equivalently, it is the rank-one orthogonal projection:

$$P_\psi = |\psi\rangle\langle\psi| \in \mathscr{B}(\mathscr{H})$$

In C^*-algebra formulation, a pure state is an extremal point in the convex set of states on the algebra of observables, that is:

$$\phi \text{ is pure} \iff \phi = \lambda\phi_1 + (1-\lambda)\phi_2,\ 0 \le \lambda \le 1, \Rightarrow \phi = \phi_1 = \phi_2$$

A pure state contains the maximal possible information; it cannot be written as a mixture of others and in Hilbert space, a pure state is of the form:

$$\phi_\psi(A) = \langle\psi|A|\psi\rangle$$

Convex Combination
Let V be a vector space over $\mathbb{C}$. A convex combination of elements $x_1, x_2, \ldots, x_n \in V$ is:

$$\sum_{i=1}^{n} \lambda_i x_i$$

where $\lambda_1 \geq 0, \forall i$ and $\sum_{i=1}^{n} \lambda_i = 1$.

Convex Set
A subset $C \subseteq V$ is convex if $\forall\, x, y \in C$ and $t \in [0,1]$, we have:

$$(1-t)x + ty \in C$$

That is, the entire line segment between any two points in C lies in C.

An extremal point (pure state)cannot be written as a convex combination of other states, whereas a non-extremal(mixed state) point can always be written as one. Note that the set of all Quantum States is a convex set.

To develop our theory beyond states, we need to define the mathematical starting point. What do we expect from a state? From the theory of statistics, we know the Mean to be linearly distributed. Therefore, the expectation value of any observable in an exceptable state must contain linearity.

For an expectation value mapping of $\omega : B(\mathscr{H}) \to \mathbb{C}$

$$\omega(aA + bB) = a\omega(A) + b\omega(B)$$

Furthermore, every measurement outcome shall possess non-negative variance.

$$\omega(A^\dagger A) \geq 0$$

and also, the probabilities should sum up to one, leading us to the condition:

$$\omega(I) = 1$$

Therefore, the starting point gives us the requirement that states are positive, normalised linear functionals.

Trace-class Operator
Let $\mathscr{H}$ be a Hilbert space. An operator $A : \mathscr{H} \to \mathscr{H}$ is trace-class if:

$$\|A\|_1 := \mathrm{Tr}\left(\sqrt{A^\dagger A}\right) < \infty$$

Equivalently, for a set of orthogonal bases $\{e_i\} \in \mathscr{H}$,

$$\|A\|_1 = \sum_n \langle e_n | A^\dagger A | e_n \rangle < \infty$$

Schatten-class duality Theorem
Every positive, normal, linear functional on $B(\mathscr{H})$ is of the form:

$$\omega(A) = Tr(\rho A)$$

For a unique positive trace-class operator $\rho \geq 0$ with $Tr(\rho) = 1$.

This operator is called the density matrix. The representation of pure states through the density matrix is done via rank-1 projectors

$$\rho_{\psi} = |\,\psi\rangle\langle\psi\,|$$

The reason for pure states to be only rank-1 and anything above being a mixed state comes from spectral theory where one can diagonalise any projector:

$$\rho = \sum_i p_i\,|\psi_i\rangle\langle\psi_i|$$

allowing it to be represented as a convex combination, making it a non-extremal point, hence a mixed state.

The properties of density matrices are as follows:

1. Hermitian: $\rho = \rho^{\dagger}$
2. Positive semidefinite: $\langle\phi \mid \rho \mid \phi\rangle \geq 0$
3. Trace is always equal to 1: $Tr(\rho) = 1$

Suppose we pick eigenbasis $|i\rangle$ of any observable. The diagonal terms of the density matrix ρ_{ii} represent the probability of finding the system in the state $|i\rangle$. Whereas the off-diagonal elements describe the quantum coherence of states which correspond to the quantum interference between two systems $|i\rangle$ and $|j\rangle$. Hence, to distinguish between pure states and mixed states, we compute $Tr(\rho^2)$:

$$\begin{aligned} \rho \text{ is a pure state } &\iff Tr(\rho^2) = 1 \\ \rho \text{ is a mixed state } &\iff Tr(\rho^2) < 1 \end{aligned}$$

4.2 Expectation Values and the Born Rule

The expectation value of an observable A is given by:

$$\langle A\rangle_{\psi} = \langle\psi \mid A \mid \psi\rangle$$

where $|\psi\rangle \in \mathscr{H}$ is a pure state. For a mixed state described by the density matrix ρ, representing a statistical ensemble of states $\{p_i\,, |\psi\rangle\}$, the expectation value corresponds to:

$$\langle A\rangle_{\rho} = Tr(\rho A) = \sum_i p_i\langle\psi_i \mid A \mid \psi_i\rangle$$

Similarly, take A to be self-adjoint, by spectral theory:

$$A = \int_{\mathbb{R}} \lambda \, dE_A(\lambda)$$

where $E_A(\lambda)$ is a PVM. To find the probability of A giving a value from a Borel set Δ on measurement is:

$$P_\rho(A \in \Delta) := \mathrm{Tr}\big(\rho \, E_A(\Delta)\big)$$

Born's rule
The probability of obtaining an outcome from the Borel set Δ if you measure A is a state ρ is:

$$P_\rho(A \in \Delta) := \mathrm{Tr}\big(\rho \, E_A(\Delta)\big)$$

For mixed states, to get the probability of the entire ensemble, you extend the pure state result linearly as:

$$\sum_i p_i \, |\psi_i\rangle\langle\psi_i|$$

4.3 Preparation of states

In the previous section, we understood that the probability of obtaining a result λ on the measurement of the system E is:

$$P(\lambda) = Tr(\rho E_\lambda)$$

Here, we observe that the probability of obtaining the result λ is completely dependent on ρ. Hence, it becomes extremely important to know what ρ is. Preparation is the physical process that produces the state ρ, which is required to predict measurement outcomes.

Gleason's theorem
Let μ be a function that assigns probability to projectors P. There exists a unique density operator ρ such that:

$$\mu(P) = Tr(\rho P)$$

This shows that all valid probability assignments for quantum measurements are described by density matrices.

One of the key requirements of preparation is that it has to be a repeatable process to produce a statistical output ρ. Hence, without knowing how the state is prepared, we cannot predict anything since probability is incomputable. Therefore, if we do not prepare any state, the system will not be in any specific state.

A quantum state is not a property of an isolated particle; rather, it is a mathematical description of the information about the state. Classically, preparing a state corresponds to stating the values of position and momentum.

Preparation
A preratation is mathematically modelled as a quantum channel:

$$E : \rho_{in} \mapsto \rho_{out}$$

where a quantum channel is a completely positive trace-preserving map. The CPTP map contains all physically allowable transformations of a quantum state.

Some examples of preparation are as follows:

1. Projective preparation: System is measured via PVM, and the outcome is preserved as:

$$\rho \mapsto \frac{P_i \rho P_i}{\mathrm{Tr}(\rho P_i)}$$

2. Randomised preparation: Choosing a state $|\psi_i\rangle$ with its respective probability p_i

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

3. Thermal preparation: Coupling to a heat bath at a given temperature T:

$$\rho = \frac{e^{-\beta H}}{Z}$$

4. Unitary preparation: Evolution under $\rho \mapsto U\rho U^\dagger$

Finally, we distinguish between classical ignorance and quantum entanglement in the following manner:

1. Proper mixing: Describes a classical ensemble of quantum states, where it's a superposition of states, namely randomised preparation.

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

2. Improper mixing: When systems are entangled, we trace out the environment as follows:

$$\rho_S = \mathrm{Tr}_E\left(|\Psi_{SE}\rangle\langle\Psi_{SE}|\right)$$

A rigorous study on the environment and entanglement will be done later in the article.

4.4 Generalised Measurements (POVMs)

Earlier, we discussed that our Projective-valued measures (PVM) were analogous to spectral measurement. The reason why PVM were not a sufficient form of measurement and why something more general was required was due to the strict assumptions of PVM:

1. Measurement is sharp
2. Measurement is repeatable
3. Measurement operators are orthogonal projections

But these were not the mathematical origins of a more generalised measurement. When one tries to perform a projective measurement on a joint system of the system and a measuring device, the effective measurement on the system is called a POVM, positive operator-valued measure.

Formally:

$$E_i = \mathrm{Tr}_A\left[(I_S \otimes \rho_A)\, P_i^{SA}\right]$$

where P_i^{SA} corresponds to the projectors on the system+measuring apparatus. Hence, POVMs are not required to be orthogonal or idempotent($E = E^2$) making it a more general form of measurement:

$$PVM \subset POVM$$

Positive Operator-Values Measure
Let $\mathscr{H}$ be a separable Hilbert space, and $(\mathbb{R}, \mathscr{B})$ be a measurable space. A POVM is a map:

$$E : \mathscr{B} \to B(\mathscr{H})$$

Satisfying:

1. $E(\Delta) \geq 0,\ \forall \Delta \in \mathscr{B}$
2. $E(\varnothing) = 0, E(\Omega) = I_{\mathscr{H}}$
3. For disjoint sets $\{B_i\}$

$$E\left(\bigcup_i B_i\right) = \sum_i E(B_i)$$

A PVM is a POVM where each $E(\Delta)$ is an orthogonal projection:

$$E(\Delta)^2 = E(\Delta) = E(\Delta)^\dagger$$

4.5 Measurement Dynamics and Kraus Operators

Say your system S is living in a state $|\psi\rangle_S$ in the Hilbert space $\mathscr{H}_S$ is measured by coupling it to a measuring device A in the Hilbert space $\mathscr{H}_A$; further, say the device is kept in a standard state $|k\rangle_A$.

As discussed before, the system evolves in time via a unitary operator, giving us the final state of the coupled system as:

$$|\Psi_{\text{final}}\rangle = U\big(|\psi\rangle_S \otimes |k\rangle_A\big)$$

Let us perform the standard projective measurement (PVM) as we formulated it only for device A.

Take $P_m = I_S \otimes |m\rangle_A\langle m|$, where $\{\, |m\rangle_A \,\}$ is the orthogonal basis for the measuring device's Hilbert Space.

Applying Born's rule, we get:

$$p(m) = \langle \Psi_{\text{final}} | P_m | \Psi_{\text{final}} \rangle$$

Kraus Representation Theory
Let $\mathscr{H}$ be a finite-dimensional Hilbert space, and let $B(\mathscr{H})$ be the space of bounded operators in $\mathscr{H}$.

A linear map E (quantum channel, hence CPTP mapping) is completely positive if and only if there exists a finite set of operators

$$M_i : \mathscr{H} \to \mathscr{H}$$

such that for all density operators ρ:

$$E(\rho) = \sum_i M_i \rho M_i^\dagger$$

Here, the operator M_i is called the Kraus Operator, where each M_i corresponds to an **outcome** of a generalised measurement (POVM). Note that:

$$\sum_i M_i^\dagger M_i = I$$

Furthermore, if $\{K_i\}$ and $\{L_j\}$ describe the same channel, there exists a unitary U such that:

$$L_j = \sum_i U_{ji} K_i$$

Hence, Kraus operators are not unique.

Now, we define a set of operators known as Kraus operators acting only on the system S:

$$U\big(|\psi\rangle_S \otimes |0\rangle_A \big) = \sum_m \big(M_m |\psi\rangle_S \big) \otimes |m\rangle_A$$

Using the above formulated operators, we calculate the probability again as:

$$p(m) = \left(\sum_k \langle \psi |_S M_k^\dagger \otimes \langle k |_A \right) \big(I_S \otimes |m\rangle_A \langle m| \big) \left(\sum_j M_j |\psi\rangle_S \otimes |j\rangle_A \right)$$

Using the orthogonality relations of the basis, the expression simplifies to a neat expression:

$$p(m) = \langle \psi |_S M_m^\dagger M_m |\psi\rangle_S$$

Setting $E_m = M_m^\dagger M_m$, we reestablish Born's probability rule.

Hence, $\{ E_m \}$ are known as POVM, positive operator valued measurements

$$p(m) = \langle \psi |_S E_m |\psi\rangle_S$$

Analysing the expression, we clearly see two properties:

1. Positive Semidefinite - $E_m \geq 0$
2. Completeness - Since $\sum_m p(m) = 1$, we can say $\sum_m E_m = \mathbb{I}$

If we consider PVMs to be 'sharp' measurements, POVMs are considered to be 'unsharp' measurements. In our general PVM, we slice the Hilbert space into two sections, giving us an exact answer. Rather, in POVMS, Hilbert space is overlapped with a probability distribution over outcomes that does not correspond to mutually orthogonal subspaces, meaning if your state falls into cloud A, it might also be in cloud B with respective probabilities.

Naimark Dilation Theorem
Every POVM on a Hilbert space is the compression of a Projection-Valued Measure (PVM) on a larger Hilbert space.

Formally, there exists an Auxiliary Hilbert space $\mathcal{K}$ powered with an isometry $V : \mathcal{H} \to \mathcal{K}$, and a PVM $P : \mathcal{B} \to B(\mathcal{H})$ such that for all measurable sets $\Delta \in \mathcal{B}$:

$$E(\Delta) = V^\dagger P(\Delta) V$$

This is often referred to as the dilation of POVM to a sharp PVM.

The measurement outcomes are not mutually exclusive properties of the state, but mutually exclusive outcomes of the apparatus. The reason POVMs are more general than PVMs is that, in PVMs, measuring an observable by projecting it onto an eigenstate collapses the initial state, often destroying the initial information. In POVM, since you are extracting partial information, the initial state is not destroyed, which disturbs it less. Note that not all POVMs have this property; some POVMs disturb the state more than PVMs.

4.6 Post-Measurement State

POVMs provide us with the probability, whereas the state update after measurement requires a set of Kraus operators.

Quantum Instrument
A Quantum instrument is a set of maps:

$$\{\mathcal{I}_i\}$$

such that, each $\mathcal{I}_i$ is CPTP, whereas the sum corresponds to a quantum channel E:

$$\sum_i \mathcal{I}_i = E$$

Each outcome mapping occurs with a probability $p(i) = \mathrm{Tr}[I_i(\rho)]$, leaving the measured state in an 'updated state' as:

$$\rho_i = \frac{\mathcal{I}_i(\rho)}{p(i)}$$

The subtle difference between a Quantum channel and a Quantum instrument is that a Quantum channel does not define a measurement uniquely, i.e. there are infinitely many measurements whose sum will equal E. Hence, a channel with a specified outcome corresponds to a Quantum Instrument. Using the Kraus operator, we say:

$$E(\rho) = \sum_{\alpha} M_{\alpha} \rho M_{\alpha}^{\dagger} \Rightarrow \text{Quantum Channel}$$

$$\mathcal{I}_i(\rho) = \sum_{\alpha} K_{i\alpha} \rho K_{i\alpha}^{\dagger} \Rightarrow \text{Quantum Instrument}$$

The sum index 'i' gives a unique identity to each measurement.

Measurement
Let $\mathcal{H}_S$ be the system's Hilbert space. A measurement is defined by a family of maps:

$$\{\mathcal{I}_i : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)\}$$

where $\mathcal{I}$ represents a quantum instrument.

Take the setup with system S and an apparatus A. The initial state is $\rho \otimes |0\rangle\langle 0|$. Applying a unitary operator $U : \mathcal{H}_S \otimes \mathcal{H}_A \to \mathcal{H}_S \otimes \mathcal{H}_A$ on the composite system:

$$\Omega = U(\rho \otimes |0\rangle\langle 0|)U^{\dagger}$$

Suppose the apparatus has a preferred orthonormalisation basis $\{|i_A\rangle\}$. Measuring the state S in these bases corresponds to the PVM:

$$\{\mathcal{I}_S \otimes |i\rangle\langle i|\}$$

The conditional probability of obtaining the measurement in the state $|i_A\rangle$ is:

$$\tilde{\rho}_i = \mathrm{Tr}_A \left[(\mathcal{I}_S \otimes |i\rangle\langle i|)\Omega(\mathcal{I}_S \otimes |i\rangle\langle i|)\right]$$

Further, normalising the probability, we obtain:

$$\rho_i = \frac{\tilde{\rho}_i}{p(i)}$$

where $p(i) = Tr(\tilde{\rho}_i)$. Using quantum instruments, we can write the conditional probability of obtaining the measurement in $|i_A\rangle$ as:

$$\rho_i = \frac{\mathcal{I}_i(\rho)}{p(i)}$$

where $p(i) = Tr[\mathcal{I}_i(\rho)]$.

Therefore, the mapping $\rho \to \rho_i$ is called quantum conditioning. Note that Quantum measurement is Bayesian conditioning on a noncommutative probability space. The quantum state after measurement is conditioned on the measurement record, and therefore depends on the experimental history.

Lüders collapse rule
Let $\{P_i\}$ be a PVM(a sharp, repeatable measurement) on $\mathscr{H}$. If the outcome i is obtained, the post-measurement state is:

$$\rho \mapsto \frac{P_i \rho P_i}{\mathrm{Tr}(\rho P_i)}$$

If the outcome is non-selective, i.e. a quantum channel(dephasing the measurement bias), the state becomes:

$$\rho \to \sum_i P_i \rho P_i$$

Note that Lüders' collapse rule represents the least disturbed updated state compatible with a PVM.

5. Quantum Kinetics and Representation Theory

This chapter develops the theory of composite quantum systems and introduces entanglement as a purely kinematical consequence of the tensor product structure of Hilbert space. Beginning with the construction of joint state spaces and reduced descriptions via partial trace, it is shown that subsystems of a globally pure state need not possess pure states of their own. The Schmidt decomposition is established as the canonical form for bipartite pure states, providing a complete and basis-independent characterisation of entanglement. Entanglement is rigorously defined as the failure of separability, and its quantitative measure is identified with the von Neumann entropy of the reduced state. The chapter further analyses the physical meaning of entanglement through measurement correlations, nonlocal state update, and the distinction between quantum and classical correlations. Taken together, these results demonstrate that entanglement is not a dynamical effect or interaction-induced phenomenon, but an intrinsic structural feature of quantum theory that underlies the departure from classical separability.

5.1 Physical Space in Quantum Mechanics

The configuration space of a physical system is the set of all possible configurations the system can occupy. Hence, for N particles, the configuration space's size is $\mathbb{R}^{3N}$. Configuration space cannot and does not define any dynamics of the system; it is purely a static representation of the system and hence can be mathematically defined as a set.

On the other hand, a Physical space is the structure defined by a group of spatial translation actions on a physical system. This represents the dynamics of a system and requires a binary operator; hence, mathematically, a group is the ideal tool to describe a Physical space.

Group
A group is a pair $(G, \circ)$, where G is a set and '$\circ$' is a binary operation on G such that:

1. Closure: $\forall g_1, g_2 \in G, g_1 \circ g_2 \in G$

2. Associativity: $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$

3. Multiplicative Identity: There exists an element $e \in G$ such that

$$g \circ e = e \circ g = g$$

4. Multiplicative Inverse: For every $g \in G$, there exists $g^{-1} \in G$ such that

$$g \circ g^{-1} = g^{-1} \circ g = e$$

A group is the mathematical encoding of reversible transformations. It can perform the following tasks:

- An element of G corresponds to an action one can perform on a system
- Group operations correspond to doing one action after another
- The identity operator corresponds to doing nothing
- Inverse action corresponds to undoing an action

Some examples in physics are:

1. Spatial Translations

$$G = (\mathbb{R}^3, +)$$

Element = displacement a
Operation = vector addition
Identity = 0
Inverse = $-a$

2. Time Translations

$$G = (\mathbb{R}, +)$$

Element = displacement a
Operation = scalar addition
Identity = 0
Inverse = $-a$

Physical Space
Physical space is a strongly continous group of spatial translations $G = (\mathbb{R}^3, +)$ acting on a quantum state space via a unitary representation.

$$a \to U(a) \in \mathcal{U}(\mathcal{H}) \text{ such that } U(a_1 + a_2) = U(a_1)U(a_2)$$

and probability is invariant, i.e. conserved under this transformation.

Spatial Localisation
A quantum system is said to be localised in a spatial region $\Delta \subset \mathbb{R}^3$ relative to a given measurement context if the probability of obtaining a position outcome in Δ is close to unity.

Mathematically;

$$\text{Prob}_\rho(\Delta) = \text{Tr}(\rho E(\Delta)) \approx 1$$

Covariance
A position observable is defined, not assumed, by the means of translational covariance.

$$U(a)^\dagger E(\Delta) U(a) = E(\Delta + a)$$

Note that now we are comfortable in defining what a transition probability is:

$$p_\rho(\Delta; a) := \text{Tr}(\rho E(\Delta + a))$$

We can now discuss on the two properties of the Physical Space:

1. Homogeneity: A physical space is homogeneous if no spatial location is physically distinguished from any other. Since a spatial location in Quantum mechanics is described by a family of probabilities, the invariant quantity to describe the homogeneity of space corresponds to:

$$\mathrm{Tr}(\rho E(\Delta)) = \mathrm{Tr}(U(a)\rho U(a)^{\dagger} E(\Delta + a))$$

Therefore, for homogeneity, the transition probability is always conserved. Physically, this states that detecting a system in a region depends only on the relative displacement of the system. In Group theory, this property is represented by the translational group $G = (\mathbb{R}^3, +)$.

2. Isotropy: A physical space is isotropic if no spatial direction is physically distinguished from any other. In Quantum mechanics, this operator is implemented by a unitary transformation:

$$R \to U(R)$$

In Group theory, this operator is defined by the rotation group $SO(3)$. Physically, this states that the statistical measurement of any experiment only depends on its relative orientation and not its absolute orientation. From the perspective of Measurement theory, we describe the process as:

$$U(a, R)^{\dagger} E(\Delta) U(a, R) = E(R^{-1}(\Delta - a))$$

Combining both the 'symmetries' of the Physical space, we define the Euclidean group:

$$E(3) = \mathbb{R}^3 \rtimes SO(3)$$

Wigner's Theorem
Let $\mathscr{H}$ be a complex Hilbert space, and let:

$$W : \mathbb{P}(\mathscr{H}) \to \mathbb{P}(\mathscr{H})$$

be a bijective map on the projective Hilbert space such that for all pure states, $|\psi\rangle, |\phi\rangle$

$$|\langle \psi | \phi \rangle|^2 = |\langle W\psi | W\phi \rangle|^2$$

Then there exists a unitary or an antiunitary operator U on $\mathscr{H}$ such that:

$$W([\psi]) = [U\psi]$$

for the equivalence class $[\psi]$

Hence, Wigner's theorem proves that physically equivalent descriptions must give the same probabilities. Therefore, Physical space acting on the Hilbert space $\mathscr{H}$ through a strong continous unitary representation:

$$(a, R) \to U(a, R)$$

preserves all transition probabilities.

The reason only unitary operators are capable of preserving transition probability is becuase of their inherent structure of preserving the norm. By Born's rule and measurement theory, we defined the probability by the modulus square of the inner product. Then:

$$|\langle U\psi | U\phi \rangle|^2 = |\langle \psi | U^{\dagger} U | \phi \rangle|^2 = |\langle \psi | \phi \rangle|^2$$

Even though Wigner's theorem allows antiunitary operators, physically they dont represent transitions for a self-adjoint operator in Hilbert space because antiunitary operators cannot form a strongly continous one-parameter group. Hence, by Stone's theorem, we discard them due to the absence of an infinitesimal generator.

5.2. Wavefunction and the Position Operator

One might think that since Hilbert space represents states of a Quantum system, it corresponds to the configuration space of the system, though that is far from the truth. The distinction is understood by the quantity wavefunction.

Wavefunction
Let $|\psi\rangle \in \mathscr{H}$ be a normalised state. The wavefunction of $|\psi\rangle$, an element of $L^2(\mathbb{R}^d)$ in the position representation is the function:

$$\psi(\mathbf{x}) = \langle \mathbf{x} | \psi \rangle$$

Where $|\mathbf{x}\rangle$ corresponds to the generalised eigenvectors of the configuration space X such that

$$\|\psi\|^2 = \int_{\mathbb{R}^d} |\psi(x)|^2 \, dx = 1$$

Hence, for N particles:

$$\psi(x_1, \ldots, x_N) = \langle x_1, \ldots, x_N | \psi \rangle$$

Note that $|\mathbf{x}\rangle \notin \mathscr{H}$, rather, they live in the dual of a rigged Hilbert space:

$$\Phi \subset \mathscr{H} \subset \Phi^\times$$

Satisfying the following property:

$$X|x\rangle = x|x\rangle, \quad \langle x|x'\rangle = \delta(x - x')$$

Where X is the position operator.

By Wigner's theorem, for spatial translations, there exists a unitary representation $U(a)$ such that:

$$U(a)E(\Delta)U(a)^{-1} = E(\Delta + a)$$

We now arrive at the position operator:

Position Operator
Given a PVM $E(\,\cdot\,)$, the spectral theorem defines a unique self-adjoint operator:

$$X = \int_{\mathbb{R}^3} \mathbf{x} \, dE(\mathbf{x})$$

Where X corresponds to the position operator.

By Stone's theorem, for every self-adjoint operator, in our case A, the operator

$$U(t) = e^{\frac{-i}{\hbar}aA}$$

form a strongly continuous one-parameter unitary group. Therefore, for the spatial translational group $(\mathbb{R}^3, +)$:

$$U_j(a) = e^{-\frac{i}{\hbar}aP_j}$$

Where P_j is a self-adjoint operator and the generator of translation in Quantum Mechanics.

Similarly, we use Stone's theorem again for the rotation group $SO(3)$:

$$U_{\hat{n}}(\theta) = e^{-\frac{i}{\hbar}\theta(\hat{n}\cdot\mathbf{J})}$$

where $\mathbf{J}$ corresponds to the generator of rotations in Quantum Mechanics.

Having defined the Position operator, we write its eigenvalue equation:

$$X\,|\mathbf{x}\rangle = \mathbf{x}\,|\mathbf{x}\rangle$$

Then any state can be expanded into its continous basis by spectral theory as follows:

$$|\psi\rangle = \int \langle x\,|\psi\rangle\,|x\rangle\,d^3x \Rightarrow |\psi\rangle = \int \psi(\mathbf{x})\,|x\rangle\,d^3x$$

For the discrete case, if $\{|n\rangle\}$ are a set of discrete orthonormal bases, any state is a sum:

$$|\psi\rangle = \sum_n c_n|n\rangle, \quad c_n = \langle n\,|\psi\rangle$$

Hence, the interpretation of the wavefunction is as the probability amplitude and by Born's rule:

$$\Pr(X \in \Omega) = \int_\Omega |\psi(x)|^2\,d^3x$$

5.3 Canonical Commutation Relations

Weyl's Theorem and Relation
Any strongly continuous unitary representation of the Heisenberg group satisfying the Weyl relations gives rise to operators $\mathbf{X}$ and $\mathbf{P}$, obeying CCR, and vice versa (on a dense domain).

For two unitary operators:

$$U(a) = e^{-ia\cdot P/\hbar}, \quad V(b) = e^{ib\cdot X/\hbar}$$

Weyl's relation states:

$$U(a)V(b) = e^{\frac{i}{\hbar}a.b}V(b)U(a)$$

Is defined on all of $\mathscr{H}$.

Translation in position and translation in momentum do not commute; they differ by a phase. Since they only change by a relative phase, the probabilities do not change; this crucial restriction causes a unitary transformation to be the only allowed way for non-commutivity to show up.

In the previous section, we had defined a translational operation on an operator $E(a)$ that corresponds to:

$$U(a)^{\dagger}E(\Delta)U(a) = E(\Delta + a)$$

Taking $E(a)$ to be the position operator $\mathbf{X}$, we get:

$$U(a)^{\dagger}\mathbf{X}U(a) = \mathbf{X} + aI$$

Using the exponential form of the unitary operator from Stone's theorem, we expand:

$$U(a) = I - \frac{i}{\hbar}a_j P_j + O(a^2) \text{ and } U(a)^{\dagger} = I + \frac{i}{\hbar}a_j P_j + O(a^2)$$

Inputting these expressions into $U(a)^{\dagger}\mathbf{X}U(a) = \mathbf{X} + aI$, we get:

$$U(a)^{\dagger}X_i U(a) = X_i + \frac{i}{\hbar}a_j[P_j, X_i] + O(a^2)$$

From comparing the exact and the approximated equations and matching the coefficients, we conclude:

$$\frac{i}{\hbar}a_j[P_j, X_i] = a_j\delta_{ij}I \Rightarrow [P_j, X_i] = -i\hbar\delta_{ij}I$$

Note that both $\mathbf{X}$ and $\mathbf{P}$ are unbounded operators and the canonical commutation relation does not hold on the entire Hilbert space, rather it holds on a dense invariant domain called the Schwartz space $\mathscr{S}$. Therefore, we rigorously define:

$$[X, P]\psi = i\hbar\psi \quad \forall\psi \in \mathscr{S}$$

Schwartz Space
The Schwartz space $\mathscr{S}(\mathbb{R}^n)$ on $\mathbb{R}^n$ is the space of all smooth functions $f(x)$ such that every derivative decays faster than any power of $|x|$. Formally:

$$f \in S(\mathbb{R}^n) \iff \sup_{x\in\mathbb{R}^n} |x^{\alpha}\partial^{\beta}f(x)| < \infty \quad \forall\alpha, \beta$$

The need for Schwartz Space $\mathscr{S}(\mathbb{R}^n)$ is because our operators $\mathbf{X}$ and $\mathbf{P}$ are unbounded and hence their commutator has a domain $\psi \in \mathscr{D}(XP) \cap \mathscr{D}(PX)$, which does not correspond to the entire Hilbert space. Hence, we need to define the space for which our commutation relation is valid. The Schwartz Space is dense in $L^2(\mathbb{R}^n)$ a Fourier transform or time evolution. This makes it the perfect candidate for the physically largest and most meaningful domain to define the commutator relation.

If we input the exponential form of the operators into Weyl's relation and differentiate both sides first with a and next with b, evaluating the differentials at $a = 0$ and $b = 0$ respectively, we will get our Canonical Commutation Relation.

One of the most important applications of the CCR relation is to derive Heisenberg's Uncertainty Principle. Take two self-adjoint operators A and B, and define $\tilde{A}$ and $\tilde{B}$ as:

$$\tilde{A} = A - \langle A \rangle \text{ and } \tilde{B} = B - \langle B \rangle$$

The uncertainty of any observable is defined as:

$$(\Delta A)^2 = \langle \psi \,|\, A^2 \,|\, \psi \rangle - \langle \psi \,|\, A \,|\, \psi \rangle^2 = \langle \psi \,|\, (\tilde{A})^2 \,|\, \psi \rangle$$

By Cauchy-Schwarz Inequality:

$$\|\tilde{A}\psi\| \|\tilde{B}\psi\| \geq |\langle \tilde{A}\psi, \tilde{B}\psi \rangle|$$

splitting the inner product into its commutator and anti-commutator:

$$\langle \tilde{A}\psi \,|\, \tilde{B}\psi \rangle = \frac{1}{2}\langle [\tilde{A}, \tilde{B}] \rangle + \frac{1}{2}\langle \{\tilde{A}, \tilde{B}\} \rangle$$

Hence, from the above expressions:

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

For $A = \mathbf{X}$ and $B = \mathbf{P}$, using the canonical commutation relations, we get:

$$\Delta \mathbf{X} \Delta \mathbf{P} \geq \frac{\hbar}{2}$$

In the context of measure theory, two POVMs $M = \{E_x\}$ and $N = \{F_y\}$ are jointly measurable iff there exists a POVM $G = \{G_{x,y}\}$ such that:

$$E_x = \sum_y G_{x,y}, \quad F_y = \sum_x G_{x,y}$$

Hence, if two measurements are not jointly measurable, they exhibit uncertainty.

In Classical mechanics, Phase space is a commutative geometry, meaning the position and momentum functions commute, meaning the phase space is static. In Quantum mechanics, $\mathbf{X}$ and $\mathbf{P}$ operators do not commute, leading to a non-commutative geometry. Through our knowledge of $\mathbf{P}$ being a generator of $\mathbf{X}$, we can intuitively understand that the space itself responds to motion in a Quantum manner. Hence, each point in the 'Phase Space' of quantum mechanics represents an Action, whether it be shifting the position or boosting the momentum.

Therefore, as a corollary, we state that two observables can be jointly sharp in a state iff they commute. This uncertainty is instrument-independent and does not depict an uncertainty in measurement. Rather, it is a fundamental Quantum property of observables and the non-commuting geometry of Quantum Phase space.

5.4 Representation Theory

There are many ways to represent Quantum Mechanics mathematically. Some are listed below:

1. Abstract Hilbert-space representation: In this form of representation, we take the states as an Equivalence class and provide POVMs as a form of measurement. This is the 'purest' form of representation of Quantum Mechanics, since it is representation-independent.

2. Group-theoretic representation: In this form of representation, we describe the kinematics of operators in forms of different types of groups and via Stone's theorem, we define Unitary Operators through Self-adjoint operators defining observables.

3. CCR representation: In this form of representation, we use Weyl's relation to justify commutation relations and define generators from Stone's theorem.

4. Schrödinger (position-space) representation: In this form of representation, the position and momentum operators are defined and from Spectral theory, we formulate the wavefunction.

5. Momentum (momentum-space) representation: Fourier Dual of Schrödinger (position-space) representation.

6. Energy (spectral) representation: In this form of representation, we use the Spectral theorem on H and expand states in terms of their eigenbasis.

The Schrödinger representation is the unitary representation of Weyl's relation on

$$\mathscr{H} = L^2(\mathbb{R}^n, dx)$$

Now, we choose to describe an arbitrary state by its position, where the orthonormal set of basis vectors is $\{|x_i\rangle\}$ such that:

$$\mathbf{X}|x_i\rangle = x_i|x_i\rangle$$

If $U(a)$ translates the system by a distance a:

$$U(a)|x\rangle = |x + a\rangle$$

Defining the wavefunction as $\psi(x) := \langle x|\psi\rangle$, we get:

$$\langle x|U(a)|\psi\rangle = \psi(x - a) \Rightarrow (U(a)\psi)(x) = \psi(x - a)$$

Differentiating the last form with respect to a we get:

$$\frac{d}{da}\psi(x-a)\bigg|_{a=0} = -\frac{d}{dx}\psi(x)$$

Further, Stone's theorem provides us with the generator $\mathbf{P}$ for the translational motion, where:

$$U(a) = e^{-iaP/\hbar} \Rightarrow P = i\hbar\frac{d}{da}U(a)\bigg|_{a=0}$$

Inputting the derivative of $U(a)$ into the expression of $\mathbf{P}$ we obtain the expression of the momentum operator as:

$$\mathbf{P}\psi(x) = i\hbar\left(-\frac{d}{dx}\psi(x)\right) \Rightarrow \mathbf{P} = i\hbar\left(-\frac{d}{dx}\right)$$

Therefore, the general form of the momentum operator is $P = -i\hbar\partial_x$

Stone–von Neumann theorem for finite degrees of freedom
Let $\mathscr{H}$ be a separable Hilbert space. Consider two strongly continuous one-parameter unitary groups satisfying Weyl relations.

Every irreducible, strongly continous, unitary representation of the Weyl relations on a separable Hilbert space is unitarily equivalent to the Schrödinger representation on

$$L^2(\mathbb{R}^n)$$

This statement implies that there is only one quantum mechanics for a particle in $\mathbb{R}^n$

Note that the theorem fails for infinite degrees of freedom, i.e. fields.

Unitary Equivalence
Two representations $(\mathscr{H}_1, \mathbf{X}_1, \mathbf{P}_1)$ and $(\mathscr{H}_2, \mathbf{X}_2, \mathbf{P}_2)$ are physically equivalent if there exists a unitary operator $W : \mathscr{H}_1 \to \mathscr{H}_2$ such that:

$$X_2 = WX_1W^\dagger, \quad P_2 = WP_1W^\dagger$$

Hence, all representations of Quantum Mechanics are equivalent in nature.

The Heisenberg matrix representation starts with assuming an abstract Hilbert space $\mathscr{H}$ and two self-adjoint operators Q and P. Take an orthonormal set of basis $\{|n\rangle\}$, where:

$$Q_{mn} = \langle m|Q|n\rangle, \quad P_{mn} = \langle m|P|n\rangle$$

correspond to an infinite matrix satisfying CCR.

By the Stone-von Neumann theorem, there exists a unitary mapping W such that

$$WQW^{-1} = X, \quad WPW^{-1} = -i\hbar\frac{d}{dx}$$

For the wavefunction $\psi(x)$ and an arbitrary state $|\psi\rangle \in \mathscr{H}$, we define:

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

Further, we take:

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

Using $|x\rangle$ as generalised eigenvectors of Q arising from its spectral decomposition, the eigenvalue equation is $\langle x|Q = x\langle x|$.

Using this, we get:

$$(WQ\,|\psi\rangle)(x) = x\psi(x)$$

Hence, we just showed

$$WQW^{-1} = X$$

That clause is supplied by **irreducibility**.

A subtlety to note is that in Schrödinger's representation, the operators remain the same, and the States go under a unitary transformation. Whereas, in Heisenberg's representation, the operators go under a unitary transformation, and the states remain the same. We can understand this via the converging point both interpretations agree on, the Expectation value.

$$\langle\psi(t) \mid A \mid \psi(t)\rangle$$

Since the expectation value of an observable is a physical result, no matter what Quantum representation one chooses, the answer and form should be identical under a unitary transformation.

In Schrödinger's representation (or active transformation), we define the unitary transformations via states evolving as:

$$|\psi_S(t)\rangle = U(t)\,|\psi_S(0)\rangle$$

Therefore, the expectation value becomes:

$$\langle A\rangle = \langle\psi_S(t) \mid A \mid \psi_S(t)\rangle$$

Notice that the operator remains unchanged.

In Heisenberg's Picture (or passive transformation), the states do not go under any transformation:

$$|\psi_H\rangle = |\psi_S(t)\rangle$$

Rather, the operators evolve:

$$A_H(t) = U^\dagger(t)A_S U(t)$$

Therefore, the expectation value corresponds to:

$$\langle A\rangle = \langle\psi_H|A_H(t)\,|\psi_H\rangle$$

Note that even though the operators evolve via unitary transformations, their spectral content is unchanged, i.e. their eigenvalues and probabilities stay intact. The equivalence of representation is simply the statement that unitary transformations preserve inner products.

5.5 Symmetries and Heisenberg's Equation of Motion

In Heisenberg's perspective, we define the motion of operators through Heisenberg's Equation of motion of an Operator as:

$$\frac{d}{dt}A_H(t) = \frac{i}{\hbar}[H, A_H(t)] + \left(\frac{\partial A}{\partial t}\right)_H$$

We know,

$$A_H(t) = U^\dagger(t) A_S U(t)$$

By Stone's theorem, we define the generator of time translation to be a self-adjoint operator called the Hamiltonian H.

$$U(t) = e^{-iHt/\hbar}$$

We rewrite $A_H(t)$ as:

$$A_H(t) = e^{\frac{i}{\hbar}Ht} A_S e^{-\frac{i}{\hbar}Ht}$$

We differentiate the following expression with respect to t assuming no explicit time dependence for the operator, we arrive at:

$$\frac{d}{dt} A_H(t) = \frac{i}{\hbar}[H, A_H(t)]$$

Therefore, for the explicit time dependence of the operator A_H, we get Heisenberg's Equation of Motion.

A general symmetry in quantum mechanics is any action or transformation that leaves the expectation value invariant after it. The operation that leaves the expectation value invariant is well known to us by now to be the unitary transformations. Therefore, a transformation $U(\alpha)$ is a symmetry transformation if:

$$A(\alpha) = U^\dagger(\alpha) A U(\alpha)$$

We further expand this concept via Stone's theorem, which allows us to define the generators G of such symmetry transformations:

$$U(\alpha) = e^{-i\alpha G/\hbar}$$

where α corresponds to the group parameter. Inputting the exponential into the expression of $A(\alpha)$ and differentiating with respect to α, we arrive at:

$$\frac{dA}{d\alpha} = \frac{i}{\hbar}[G, A]$$

Comparing with Heisenberg's Equation of Motion for non-explicit dependence on time, we identify the Hamiltonian H to be the symmetry generator for group parameter t or more precisely, time translation symmetry.

A Quantum Particle in space has the following symmetries:

1. Time Translational Symmetry:

This symmetry puts the constraint of all instants of time being physically equivalent, meaning all physical laws act the same for t and $t + dt$. Mathematically, there exists a map

$$t \mapsto U(t)$$

such that:

A] Group property: $U(t_1)U(t_2) = U(t_1 + t_2), \quad U(0) = I$

B] Strong continuity

C] Unitary

By the Constraint of time translation symmetry, the following equation must hold for the time translation generator H:

$$U^{\dagger}(t)HU(t) = H, \forall t \iff \frac{\partial H}{\partial t} = 0$$

Therefore, time translational symmetry corresponds to conservation of the Hamiltonian(interpreted as Energy) of the system.

2. Space Translational Symmetry

As we have discussed before, the Homogeneity of space in Quantum Mechanics corresponds to no spatial location being physically distinguished from any other. Therefore, all physical laws remain the same if a system is present at position x or $x + a$. This implies the following:

$$U^{\dagger}(a)HU(a) = H$$

Differentiating the following expression at $a = 0$, we get:

$$\text{using } \frac{d}{da}\left(e^{Aa}Be^{-Aa}\right)\Big|_{a=0} = [A, B]$$

$$[H, \mathbf{P}] = 0$$

Further from Heisenberg's equation of motion for non-explicit dependence on time, we can say:

$$\frac{d\mathbf{P}}{dt} = 0$$

Therefore, space translational symmetry corresponds to the conservation of total linear momentum of the system.

3. Rotational Symmetry

From the Isometric property of space in Quantum Mechanics, no spatial direction is physically distinguished from any other. Following the same procedure, we define:

$$U(\theta) = e^{-i\theta\cdot\mathbf{J}/\hbar}$$

where J is the generator of rotational symmetry.

$$\frac{dA}{d\theta} = \frac{i}{\hbar}[\mathbf{J}, A] \text{ and } \frac{d\mathbf{J}}{dt} = 0$$

Therefore, rotational symmetry corresponds to conservation of total angular momentum(orbital + intrinsic spin) of the system.

4. Boost Symmetry

A Galilean boost shifts the reference frame, describing how the observables change:

$$\mathbf{x} \to \mathbf{x}' = \mathbf{x} + vt$$

Since a change of inertial or reference frame must preserve all probabilities, the change is represented via a unitary transformation. Therefore, my Stone's theorem:

$$U(v) = e^{-\frac{i}{\hbar} v \cdot K}$$

Differentiating $\mathbf{x} \to \mathbf{x}' = \mathbf{x} + vt$, we get:

$$\frac{dx_i}{dv_j} = t\delta_{ij}$$

Comparing the general form of Abstract symmetry in Quantum Mechanics, we get the commutation relation:

$$[K_j, x_i] = i\hbar t \delta_{ij}$$

Performing the same steps for the momentum observable:

$$\mathbf{p} \to \mathbf{p} + mv$$

We arrive at the commutation relation:

$$[K_j, p_i] = i\hbar m \delta_{ij}$$

Here, mass is the obstruction to commuting boosts and translations. This is the first time 'mass' has shown up in Quantum Mechanics as a weight to the inertial frames; more specifically, mass is the measure of how quantum states respond to changes of inertial frame.

Bargmann Theorem
Projective representations of the Galilean group are classified by a general charge m

The two commutation relations constrain the generator K to be:

$$K = m\mathbf{x} - t\mathbf{p}$$

By Boost symmetry, we know that the time derivative of K is zero

Therefore, from:

$$\frac{d}{dt}K(t) = \frac{i}{\hbar}[H, K(t)] + \left(\frac{\partial K}{\partial t}\right)_H$$

We deduce $[H, x_i] = -\frac{i\hbar}{m} p_i$. Further, by translational symmetry, we state:

$$H = f(\mathbf{p})$$

Using both relations:

$$[H, x_i] = \sum_j \frac{\partial f}{\partial p_j}[p_j, x_i] = -i\hbar \frac{\partial f}{\partial p_i} \Rightarrow \frac{\partial f}{\partial p_i} = \frac{p_i}{m}$$

Therefore,

$$H = \frac{\mathbf{p}^2}{2m} + c$$

Galilean Group
The Galilean group is the set of all transformations that preserve the laws of Newtonian mechanics under changes between inertial reference frames. The Galilian Algebra consists of the following commutation relations:

$[J_i, J_j] = i\hbar\epsilon_{ijk}J_k$ $[J_i, P_j] = i\hbar\epsilon_{ijk}P_k$ $[J_i, K_j] = i\hbar\epsilon_{ijk}K_k$ $[P_i, P_j] = 0$

$[K_i, K_j] = 0$ $[H, P_i] = 0$ $[H, K_i] = -i\hbar P_i.$ $[P_i, K_j] = i\hbar m\delta_{ij}$

Noether's Theorem
If the Hamiltonian is invariant under a continuous unitary transformation, then the generator G of that transformation is a conserved operator.

$$[H, G] = 0 \Rightarrow \frac{d\langle G\rangle}{dt} = 0$$

Where G is the symmetry transformation generator from:

$$U(\alpha) = e^{-i\alpha G/\hbar}$$

5.6 Schrödinger's Equation of Motion

In Schrödinger's representation of Quantum Mechanics, we defined:

$$(X_i\psi)(x) = x_i\psi(x), \quad (P_i\psi)(x) = -i\hbar\,\partial_i\psi(x)$$

As the position and momentum operators respectively on $L^2(\mathbb{R}^3)$.

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle$$

Inputting the exponential form of $U(t)$ and differentiating the equation with respect to t, we get the abstract Schrödinger equation in terms of the state vector:

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = H\,|\psi(t)\rangle$$

To represent this in terms of the wavefunction, we take the inner product of the equation with respect to the position basis(Schrödinger representation):

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \langle x\,|\,H\,|\,\psi(t)\rangle$$

From the previous section, we defined the Hamiltonian to be:

$$H = \frac{\mathbf{p}^2}{2m} + c$$

For a single non-relativistic particle in a potential $V(\mathbf{x})$:

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

Therefore,

$$\langle x \,|\, H \,|\, \psi(t) \rangle = \langle x \,|\, H \,|\, \psi \rangle = \frac{1}{2m} \langle x \,|\, p^2 \,|\, \psi \rangle + \langle x \,|\, V(x) \,|\, \psi \rangle$$

From the definition of the momentum operator in Schrödinger's representation, we define $\mathbf{p}^2$ as follows:

$$p^2 = -\hbar^2 \frac{\partial^2}{\partial x^2}$$

Therefore, we write the famous time-dependent Schrödinger Equation for a single non-relativistic particle in a potential $V(\mathbf{x})$:

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{x}, t) = \left(-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x}) \right) \psi(\mathbf{x}, t)$$

Using the method of separable variables on the time-dependent Schrödinger Equation:

$$\psi(x, t) = \phi(x) T(t)$$

Taking the constant to equate as E, we arrive at the following solutions:

$$T(t) = e^{-iE\frac{t}{\hbar}}$$

$$H\phi(x) = E\phi(x)$$

where the Hamiltonian for a single non-relativistic particle in a potential $V(\mathbf{x})$ is:

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x})$$

Hence, from the time-independent Schrödinger Equation for a single non-relativistic particle in a potential $V(\mathbf{x})$, we interpret E as the eigenvalues and $|\phi\rangle$ as eigenstates. The time evolution is as it should be, a pure phase.

$$|\psi(t)\rangle = e^{-iEt/\hbar} |\phi\rangle \Rightarrow |\psi(x,t)|^2 = |\phi(x)|^2$$

Using the time-dependent Schrödinger Equation, we will now move forward to derive the Continuity Equation. Take the probability density defined by Born's rule:

$$\rho(x, t) = \psi^*(x, t) \psi(x, t)$$

For all probabilities to be conserved, the time derivative of probability with respect to time should be zero. The partial of probability density with time corresponds to:

$$\frac{\partial \rho}{\partial t} = \psi^* \frac{\partial \psi}{\partial t} + \psi \frac{\partial \psi^*}{\partial t}$$

We use the TDSE to substitute the time derivative of the wavefunction and its conjugate(take the conjugate of TDSE)

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}\left[\frac{i\hbar}{2m}\left(\psi^* \frac{\partial \psi}{\partial x} - \psi \frac{\partial \psi^*}{\partial x}\right)\right] = 0$$

Hence, we define the probability current as:

$$j(x,t) = \frac{i\hbar}{2m}\left(\psi^* \frac{\partial \psi}{\partial x} - \psi \frac{\partial \psi^*}{\partial x}\right)$$

Further, the continuity equation is formulated as:

$$\frac{\partial \rho}{\partial t} + \frac{\partial j}{\partial x} = 0$$

Where ρ corresponds to the probability density and j corresponds to the probability current. The operator form of probability current is defined as:

$$\mathbf{j}(\mathbf{x}) = \frac{\mathbf{p}\,|\,\mathbf{x}\rangle\langle\mathbf{x}\,| + |\,\mathbf{x}\rangle\langle\mathbf{x}\,|\,\mathbf{p}}{2m}$$

5.7 State Transformation

Let ρ be a density operator such that $\rho \geq 0$ and $Tr(\rho) = 1$ with:

$$\omega(A) = Tr(\rho A)$$

State Transformation
A state transformation is a linear map

$$\tau : T(\mathscr{H}) \to T(\mathscr{H})$$

On the trace-class operator, such that the mapping is CPTP

Measurement-Induced Transformation
As discussed in Measurement theory, the state update rule defines the state to transform to:

$$\rho \mapsto \frac{M_k \rho M_k^\dagger}{\mathrm{Tr}(M_k \rho M_k^\dagger)}$$

After measurement, where M_k corresponds to the Kraus Operator

The extremal points of the CPTP set are the Symmetry-Induced State transformation, which are reversible and have generators. Note that unitary (or antiunitary) channels form a special reversible subset of CPTP maps; generic CPTP maps are irreversible.

Symmetry-Induced State Transformation
A symmetry acts on states by a family of maps:

$$g \to T_g$$

Such that:

1. Probability preservation: $\mathrm{Tr}(\rho E) = \mathrm{Tr}(T_g(\rho)T_g^*(E))$

2. Group Preservation: $T_{g_1 g_2} = T_{g_1} \circ T_{g_2}$

For probability preservation, the continous transformation is unitary. Therefore:

$$T_g = U^\dagger(g)\rho U(g)$$

Note that there exist non-unitary transformations described by semi-groups in Quantum Mechanics governed by the Lindblad–GKLS equation.

In the Unitary transformation, we proceed further via Weyl's Relation, we state:

$$U(g_1)U(g_2) = e^{i\alpha(g_1,g_2)}U(g_1 g_2)$$

Take, for example, time translation. The State transformation for evolving in time is:

$$\rho \to e^{-iH\frac{t}{\hbar}}\rho e^{iH\frac{t}{\hbar}}$$

The same structure is assumed for other generators.

5.8 Classical Limit

It is common to connect the concepts of classical mechanics as an intuitive approach, and Quantum Mechanics is derived from that intuition. One will soon hit a wall through this procedure since Quantum Mechanics is the more fundamental concept and Classical Mechanics is the limiting case of it, not the other way around. In this section, we will introduce some extensions of Quantum Mechanical concepts and how they are interpreted in Classical Mechanics.

The classical limit is often named as $\hbar \to 0$. One should not take this to be mathematically accurate as a form of taking a limit, but rather understand it as Quantum effects become negligible on the classical scale.

$$\hbar \ll S_{classical}$$

The reason why $\hbar$ was the chosen candidate to compare classical behaviour is due to its strong presence in CCRs, where for $\hbar \to 0$, we observe that:

$$[\mathbf{X}, \mathbf{P}] = 0$$

Hence, the need for non-commutating geometry vanishes. This is not the only reason; consider the TDSE, where the classical limit of forces requires a rapid change in the wavefunction, the phase of quantum amplitudes oscillates so rapidly that only classical trajectories with stationary action survive, making observables commute and dynamics reduce to Hamiltonian mechanics.

Other effects are also capable of quantifying Quantum Mechanical behaviour, namely:

1. Semiclassical Limit $\hbar \to 0$
2. Large Quantum Number: $n \gg 1$
3. Macroscoping limit
4. Coarse-graining

Ehrenfest's Theorem
For a quantum system with Hamiltonian

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

and a state $|\psi(t)\rangle$ evolving via the Schrödinger equation. The expectation values of the position and momentum operators obey:

$$\frac{d}{dt}\langle x \rangle = \frac{1}{m}\langle p \rangle$$

$$\frac{d}{dt}\langle p \rangle = -\left\langle \frac{\partial V(x)}{\partial x} \right\rangle$$

Further, for a general operator A:

$$\frac{d}{dt}\langle A(t) \rangle = \frac{i}{\hbar}\langle [H, A(t)] \rangle + \left\langle \frac{\partial A(t)}{\partial t} \right\rangle$$

Ehrenfest's theorem is the bridge between quantum dynamics and classical mechanics, showing us that classical equations govern the behaviour of the averages and not their corresponding individual quantum processes. Classically, we know:

$$\frac{dA}{dt} = \{A, H\}$$

where $\{\}$ corresponds to Poisson's Bracket. Using Ehrenfest's theorem, we induce the relation between the commutators and Poisson's Bracket as:

$$\frac{1}{i\hbar}[\hat{A}, \hat{B}] \xrightarrow{\hbar \to 0} \{A, B\}$$

Until now, we have discussed how non-commutative geometry disappears for the classical limit, a consequence of that being the shift of commutative relations to Poisson's bracket. Also, we state why only the path with stationary action survives via TDSE. As a final bridge, we explore Wigner's Function.

Wigner's Function
For a quantum state described by a density operator ρ, the Wigner function $W(x, p)$ is defined as:

$$W(x,p) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dy\, e^{-\frac{i}{\hbar}py}\left\langle x+\frac{y}{2}\middle|\hat{\rho}\middle|x-\frac{y}{2}\right\rangle$$

It is a real-valued function on classical phase space (x, p). Hence, the Wigner function provides an exact phase-space formulation of quantum mechanics as a non-positive quasi-probability distribution on classical phase space.

The properties of the function are:

1. $W(x,p) \in \mathbb{R}$

2. Normalisation - $\int dx\, dp\, W(x,p) = 1$

3. Marginal Probabilities:

$$\int dp\, W(x,p) = |\psi(x)|^2 \qquad \int dx\; W(x,p) = |\tilde{\psi}(p)|^2$$

4. Expectation values:

$$\langle A \rangle = \int dx dp A(x,p) W(x,p)$$

Note that $W(x,p)$ is not always positive. For the regions where $W(x,p) < 0$ we interpret it as intrinsic Quantum Behaviour, such as Quantum Interference, or non-classical correlations.

6. Composite Quantum System and Entanglement

This chapter characterises composite quantum systems by utilising the tensor product structure of Hilbert space, establishing entanglement as a kinematical consequence of joint state spaces rather than a dynamical effect. It demonstrates that while a global system may exist in a pure state, its subsystems recovered via the partial trace are frequently mixed states, a phenomenon known as improper mixing that reflects the nonexistence of independent local states. The Schmidt decomposition is introduced as the canonical framework for bipartite pure states, where a Schmidt rank $r > 1$ serves as the rigorous mathematical criterion for entanglement. Through the lens of the EPR paradox and Bell's theorem, the text proves that quantum mechanics violates local realism, showing that the "best possible knowledge of a whole does not include the best possible knowledge of its parts. The chapter concludes by quantifying these non-classical correlations using the von Neumann entropy of the reduced density matrix, identifying the entropy of entanglement as the Shannon entropy of the Schmidt coefficients.

6.1 Tensor Product in Hilbert Space

Up until this point, we have considered one system at a time. Many physical processes involve multiple systems, which may or may not interact with each other, and are simultaneously changing. To study such systems, called 'Composite Systems', we need the mathematical tool called Tensor products.

Tensor Product
Given two complex vector spaces, $\mathcal{H}_1$ and $\mathcal{H}_2$, the algebraic tensor product:

$$\mathcal{H}_1 \otimes_{alg} \mathcal{H}_2$$

is a vector space generated by the bases:

$$\psi \otimes \phi$$

where $\psi \in \mathcal{H}_1$ and $\phi \in \mathcal{H}_2$, such that:

1. Bilinearity:

$$(\psi_1 + \psi_2) \otimes \phi = \psi_1 \otimes \phi + \psi_2 \otimes \phi$$

$$\psi \otimes (\phi_1 + \phi_2) = \psi \otimes \phi_1 + \psi \otimes \phi_2$$

2. Scalar multiplication:

$$(\alpha\psi) \otimes \phi = \psi \otimes (\alpha\phi)$$

The inner product for the system belonging to $\mathcal{H}_1 \otimes_{alg} \mathcal{H}_2$ behave as:

$$\langle \psi_1 \otimes \phi_1 \mid \psi_2 \otimes \phi_2 \rangle = \langle \psi_1 \mid \psi_2 \rangle \langle \phi_1 \mid \phi_2 \rangle$$

The issue with the algebraic tensor product corresponds to its incomplete nature, meaning the algebraic tensor product may not always be complete with respect to the norm. Hence, there exist Cauchy sequences of finite sums of simple tensors whose limits are not finite sums of simple tensors.

Tensor Product in Hilbert Space
The joint system whose inner product has an induced norm:

$$\|x\| = \sqrt{\langle x, x \rangle}$$

Therefore,

$$\mathcal{H}_1 \otimes \mathcal{H}_2$$

are complete with respect to the norm induced by the tensor product inner product. In quantum mechanics, the state space of a composite system is not the Cartesian product of subsystem state spaces but the tensor product. This is because the Cartesian product does not form a complete space.

Universal Property of Tensor Products
There exists a bilinear map:

$$\otimes : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{H}_1 \otimes \mathcal{H}_2$$

Such that for any Hilbert space $\mathcal{K}$ and any bounded bilinear map:

$$B : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{K}$$

There exists a unique bounded linear map:

$$\tilde{B} : \mathcal{H}_1 \otimes \mathcal{H}_2 \to \mathcal{K}$$

Satisfying $B(\psi, \phi) = \tilde{B}(\psi \otimes \phi)$.

This property **defines** the tensor product up to unique isomorphism. Before moving to composite systems, we shall get acquainted with how tensor products work with bases and operators. For two orthonormal bases $\{e_i\} \subset \mathcal{H}_1$ and $\{f_j\} \subset \mathcal{H}_2$, we define the orthonormal bases of the Hilbert tensor product as:

$$\{e_i \otimes f_j\}_{ij}$$

Hence, the dimension of the tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is:

$$\dim(\mathcal{H}_1 \otimes \mathcal{H}_2) = \dim(\mathcal{H}_1) \cdot \dim(\mathcal{H}_2)$$

Further, give two bounded operators A acting on $\mathcal{H}_1$ and B acting on $\mathcal{H}_2$. We define:

$$(A \otimes B)(\psi \otimes \phi) = A\psi \otimes B\phi$$

such that:

1. Norm is preserved: $\|A \otimes B\| = \|A\|\|B\|$

2. Distributive Dagger: $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$

Composite System
Let S_1 and S_2 be two quantum systems with $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. The composite system:

$$S = S_1 + S_2$$

is a quantum system whose Hilbert space is the tensor product of $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$$

The properties of a Composite System are:

1. Independent preparation: If S_1 is prepared in ρ_1 and S_2 is prepared in ρ_2, then the composite state must depend bilinearly on (ρ_1, ρ_2), implying:

$$(\rho_1, \rho_2) \rightarrow \rho_1 \otimes \rho_2$$

2. Local measurements: If A is an observable on S_1, measurement A while doing no measurement on S_2 is represented by:

$$A \otimes I$$

3. Joint measurement: A joint measurement on an observable A of S_1 and B of S_2 corresponds to:

$$\sum_k A_k \otimes B_k$$

4. Multiple subsystems: Extending the concept from only two subsystems to multiple N subsystems requires us to take the tensor product of those N subsystems as follows:

$$H = \bigotimes_{k=1}^{N} H_k$$

Note that a system with two distinct subsystems is called a bipartite quantum system. Before moving to the next section, we get ourselves acquainted with correlations between states.

Correlation of states
Let $\mathcal{H}_A, \mathcal{H}_B$ be Hilbert spaces and let:

$$\rho_{AB} \in T(\mathcal{H}_A \otimes \mathcal{H}_B)$$

be a bipartite quantum state. The state ρ_{AB} is uncorrelated(a product state) iff for all observables O_A, O_B the equality holds:

$$\langle O_A \otimes O_B \rangle = \langle O_A \rangle \langle O_B \rangle$$

whereas, a bipartite state ρ_{AB} is said to be correlated iff:

$$\rho_{AB} \neq \rho_A \otimes \rho_B$$

Equivalently, for all observables O_A, O_B:

$$\mathrm{Tr}\left[(O_A \otimes O_B)\,\rho_{AB}\right] \neq \mathrm{Tr}(O_A \rho_A)\ \mathrm{Tr}(O_B \rho_B)$$

where ρ_A and ρ_B correspond to partial traces of subsystem A and B respectively.

For our convenience, we define the correlation function:

$$C(O_A, O_B) := \langle O_A \otimes O_B \rangle - \langle O_A \rangle \langle O_B \rangle$$

Therefore, for $C = 0 \ \forall O_A, O_B$ the bipartite states are product states, and for $C \neq 0$ for atleast one O_A, O_B the bipartite states is correlated.

There exist two types of correlations:

1. Classical correlation: A state is classically correlated if it can be written as:

$$\rho_{AB} = \sum_i p_i \, \rho_A^{(i)} \otimes \rho_B^{(i)}, \qquad p_i \geq 0, \ \sum_i p_i = 1$$

such states are separable.

2. Quantum Correlation: Two states having quantum correlation are called entangled states, such that:

$$\rho_{AB} \neq \sum_i p_i \, \rho_A^{(i)} \otimes \rho_B^{(i)}$$

Correlations exist when the state of the composite system contains information not present in the individual subsystem states.

6.2 Schmidt Decomposition and Mixing of States

Suppose we consider a bipartite quantum system $\mathscr{H}$ and an arbitrary state $|\psi\rangle \in \mathscr{H}$, we can write the state as:

$$|\psi\rangle = \sum_{i,j} c_{ij} |i\rangle_A |j\rangle_B$$

Where $|i\rangle_A$ are bases spanning $\mathscr{H}_A$, $|j\rangle_B$ are bases spanning $\mathscr{H}_B$, and c_{ij} corresponds to the coefficient of expansion. This representation often becomes tedious due to its double-sum nature. To solve this issue, we define the Schmidt decomposition. The idea of Schmidt decomposition is that any pure state of a bipartite quantum system can be written as a sum of perfectly correlated orthonormal states of the two subsystems, decomposing into independent correlation channels

Schmidt decomposition method finds new bases:

$$\{ |\mathbf{u}_k\rangle_A \}, \ \{ |\mathbf{v}_k\rangle_B \}$$

Such that on basis expansion, we get a one-to-one pairing for the two systems, stating that if subsystem A is in $|\mathbf{u}_k\rangle$, then subsystem B has to be in $|\mathbf{v}_k\rangle$. This form of one-to-one pairing is obtained via Tensor products:

$$|\psi\rangle = \sum_{k=1}^{r} \sqrt{\lambda_k}\, |\mathbf{u}_k\rangle_A \otimes |\mathbf{v}_k\rangle_B$$

where λ_k is the Schmidt decomposition coefficient.

Schmidt Decomposition
Let $\mathscr{H}_A, \mathscr{H}_B$ be separable Hilbert spaces and $|\psi\rangle \in \mathscr{H}_{AB} = \mathscr{H}_A \otimes \mathscr{H}_B$ be a normalised pure state.

A Schmidt decomposition of the state $|\psi\rangle$ is a representation of the form:

$$|\psi\rangle = \sum_{k=1}^{r} \sqrt{\lambda_k}\, |\mathbf{u}_k\rangle_A \otimes |\mathbf{v}_k\rangle_B$$

where r is called the Schmidt rank of $|\psi\rangle$ such that:

$$\{|\mathbf{u}_k\rangle\}_{k=1}^{r} \subset \mathscr{H}_A, \{|\mathbf{v}_k\rangle\}_{k=1}^{r} \subset \mathscr{H}_B$$

both being an orthonormal set, and $\lambda_k > 0, \forall k$ with:

$$\sum_{k=1}^{r} \lambda_k = 1$$

Note that every pure state admits a Schmidt decomposition. To work with Mixed states, we first define $T(\mathscr{H})$, a space of trace-class operators on $\mathscr{H}$, where a state of the composite system $\mathscr{H}_A \otimes \mathscr{H}_B$ corresponds to:

$$\rho_{AB} \in T(\mathscr{H}_A \otimes \mathscr{H}_B)$$

Partial Trace
The partial trace over the subsystem B is a unique linear map:

$$\mathrm{Tr}_B : T(\mathscr{H}_A \otimes \mathscr{H}_B) \longrightarrow T(\mathscr{H}_A)$$

Such that for all bounded operators $X \in B(\mathscr{H}_A)$:

$$Tr_{\mathscr{H}_A}\big(X\, Tr_B(\rho_{AB})\big) = Tr_{\mathscr{H}_A \otimes \mathscr{H}_B}\big((X \otimes I_B)\rho_{AB}\big)$$

Hence, we have:

$$\mathrm{Tr}_B(\rho_{AB}) = \sum_k (I_A \otimes \langle b_k|)\, \rho_{AB}\, (I_A \otimes |b_k\rangle)$$

This is a Kraus decomposition with:

$$K_k = I_A \otimes \langle b_k|$$

Partial trace represents a complete loss of access to subsystem B. To expand on the topic, we note that the Partial trace is a CPTP mapping which guarantees physicality of the concept. Expanding on

this idea, we can state a theorem supporting the local measurement property mentioned before. Let E_b be any CPTP mapping acting only on subsystem B. Then:

$$\mathrm{Tr}_B\left[(I_A \otimes E_B)(\rho_{AB})\right] = \rho_A$$

Reduced State
The reduced state of a subsystem A associated with the composite state ρ_{AB} is:

$$\rho_A = Tr_B(\rho_{AB})$$

Note that the reduced state ρ_A contains all information accessible to subsystem A. Therefore, for any observable O_A acting only on A:

$$\langle O_A \rangle = \mathrm{Tr}_{AB}\left[(O_A \otimes I_B)\,\rho_{AB}\right] = \mathrm{Tr}_A\left(O_A \rho_A\right)$$

Hence, no measurement on A alone can distinguish ρ_{AB} from ρ_A. Using Schmidt decomposition, we define:

$$\rho_A = \sum_k \lambda_k \,|\, u_k\rangle\langle u_k \,|\,, \quad \rho_B = \sum_k \lambda_k \,|\, v_k\rangle\langle v_k \,|$$

Using Kraus operators, we write:

$$\rho_A(t) = \sum_\alpha K_\alpha(t)\,\rho_A(0)\,K_\alpha^\dagger(t)$$

This has an immediate impact on the non-unitary evolutions of the subsystem. Assume the global system to evolve unitarily:

$$\rho_{AB}(t) = U_{AB}(t)\,\rho_{AB}(0)\,U_{AB}^\dagger(t) \Rightarrow \rho_A(t) = \mathrm{Tr}_B\left[\,\rho_{AB}(t)\,\right]$$

This chain of equations confirms that $U_{AB} \neq U_A \otimes U_B$

$$\rho_A(t) \neq U_A(t)\,\rho_A(0)\,U_A^\dagger(t)$$

Rather, it is nonunitary due to coupling with inaccessible degrees of freedom. Hence,

$$\rho_A(t) = \sum_\alpha K_\alpha(t)\,\rho_A(0)\,K_\alpha^\dagger(t)$$

The Schmidt rank satisfies:

$$r = rank(\rho_A) = rank(\rho_B)$$

Further, we can identify whether a global state is entangled or not via the Schmidt rank. Firstly, through the rigorous manner, a global state is factorisable if:

$$|\,\psi\rangle_{AB} = |\,\phi\rangle_A \otimes |\,\chi\rangle_B$$

If and only if this is true, the subsystem A is not entangled and ρ_A is a pure state. For any global state:

$$|\psi\rangle_{AB} \neq |\phi\rangle_A \otimes |\chi\rangle_B$$

The state is entangled, further:

$$\rho_A = \mathrm{Tr}_B\big(|\psi\rangle\langle\psi|\big)$$

is necessarily mixed. Through the Schmidt rank, we can distinguish as follows:

$$r = 1 \Rightarrow \text{Separable States}$$
$$r > 1 \Rightarrow \text{Entangled State}$$

The global density matrix corresponds to:

$$\rho_{AB} = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B$$

Hence, the reduced state ρ_A is:

$$\rho_A = \sum_i \lambda_i |i\rangle\langle i| \Rightarrow \rho_A^2 = \sum_i \lambda_i^2 |i\rangle\langle i|$$

$Tr(\rho_A^2) = Tr(\rho_A)$ if and only if there exists only one value of λ_i , and that is only possible for $r = 1$. Therefore, any reduced state, i.e. a subsystem, is entangled unless it's a product state. This arises due to properties of the Tensor product, which allows superpositions in tensor-product spaces to not factor into superpositions of subsystems. Therefore, Global purity does not imply local purity.

To extend the concept of mixing, we categorise them into pure mixing and impure mixing:

1. Proper Mixing - A state is properly mixed if each system in the ensemble is actually prepared in one of the states $|\phi_i\rangle$:

$$\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$$

where p_i represents the classical uncertainty.

2. Improper Mixing - A state is improperly mixed if:

$$\rho_A = \mathrm{Tr}_B\big(|\psi\rangle\langle\psi|\big)$$

For some entangled state $|\psi\rangle$. In this case, there exists no information about which pure state subsystem A is in. Further, both mixing mathematically identical density matrices but physically inequivalent. Classical correlations(proper mixing) reflect ignorance about local states; entanglement(improper mixing) reflects the nonexistence of local states.

Through the method of purification of mixed states, we can depict that every mixed state can be seen as the reduced state of a larger pure state.

Schmidt Purification
Let ρ_S be a density operator on a Hilbert space $\mathscr{H}_S$. A purification of ρ_S is a triple $(\mathscr{H}_E, |\Psi\rangle, Tr_E)$ such that:

1. $\mathscr{H}_E$ is an auxiliary Hilbert space

2. $|\Psi\rangle \in \mathscr{H}_S \otimes \mathscr{H}_E$ is a pure state

3. The reduced state on S satisfies:

$$\rho_S = \mathrm{Tr}_E\big(|\Psi\rangle\langle\Psi|\big)$$

with the dimensional constraint of:

$$dim(\mathscr{H}_E) \geq rank(\rho_S)$$

This constraint exists becuase each nonzero eigenvalue of ρ_S requires an independent environmental degree of freedom. Hence, a smaller environment cannot support the process
Therefore, a purification of ρ_S is a pure state of a larger system whose partial trace reproduces ρ_S. Note that all purifications of a ρ_S given differ only by unitaries on the purifying system and are not unique. Hence, applying the partial trace on a pure global state would give us mixed local states, and purifying mixed local states would give us a non-unique pure global state.

Stinespring Dilation Theorem
For every CPTP map E, there exists:

1. An auxiliary Hilbert space $\mathscr{H}_E$

2. A unitary operator U on $\mathscr{H}_S \otimes \mathscr{H}_E$

3. A fixed pure state $|0\rangle_E$

such that:

$$E(\rho) = \mathrm{Tr}_E\big[\, U\big(\rho \otimes |0\rangle\langle 0|\big)U^\dagger\big]$$

Therefore, every quantum operation is a unitary evolution on a larger system, followed by ignoring part of it.

6.3 Entanglement of Pure States

In 1935, Einstein, Podolsky and Rosen formulated the EPR paradox. It was well known by virtue of the principle of causality that physical influences cannot propagate faster than light. One of EPR's claims was:

Claim: The physical state of a composite system is fully determined by the physical states of its spatially separated subsystems.

This property is called local decomposability. In our previous section, we studied how a global state can be pure, but its subsystems are mixed. This fact completely denies EPR's claim.

Let's take an example to understand. Suppose we have a two-particle state in which the position and momentum of both particles are perfectly correlated.

Now, assume two observers, Alice and Bob. If Alice measures an observable A on particle 1, she can with certainty predict Bob's measurement result for observable B. This seems like particles are space-like seperated which is not allowed by the Special theory of relativity. Even Quantum Mechanics does not allow this form of measurement since it allows one to assign both momentum and position of a particle simultaneously.

It was Schrödinger who realised the issue is not regarding uncertainty, but rather it is due to the inseparability or the entangled nature of the two particles. The global system has a definite wavefunction; the subsystems do not. Meaning, Bob's particle does not have a state by itself. He stated:

"The best possible knowledge of a whole does not include the best possible knowledge of its parts"

Therefore, we can say that the entanglement of a pure state is the failure of spatial separation to imply state separation.
Mathematically, the EPR paradox fails through Bell's theorem. Bell assumes the three properties which formualte the concept of 'Local Realism', which are as follows:

1. Realism: Measurement outcomes are determined by pre-existing properties.
2. Locality: The outcome of an observable A does not depend on the setting of B
3. Statistical Independence: Hidden variables are uncorrelated with measurement choices.

Take the two observers, Alice and Bob. Say each of them takes two measurements a, a' and b, b' their respective outcomes are restricted to ± 1. The correlation is defined as:

$$E(a,b) = \int d\lambda\, \rho(\lambda) A(a,\lambda) B(b,\lambda)$$

Further, we define the quantity S:

$$S = E(a,b) + E(a,b') + E(a',b) - E(a',b')$$

Bell-CHSH Inequality
For local realism:

$$|S| \leq 2$$

Every Bell-violating state is entangled, but not all entangled states violate the Bell Inequality. This is because the core reasoning of Entanglement and Bell's Inequaltiy are different. Entanglement says the state is non-seperable whereas Bell non-locality says the impossibility of classical explanations for observed correlations.

This inequality was violated by Quantum mechanical measurements. Therefore, the conclusion was that Quantum Mechanics disobeys at least one of the three assumptions for local realism stated to construct the inequality. Realism states that measurement outcomes reveal pre-existing, measurement-independent properties of the system; this assumption was rejected by Quantum Mechanics since quantum states encode probabilities for measurement outcomes, not actual

properties of the states. Measurement on part of an entangled system updates the global state, not just the local subsystem. Therefore, Bell stated Bell's theorem:

Bell's theorem
No theory satisfying local realism can reproduce all quantum mechanical predictions.

Entangled States
A state ρ_{AB} is entangled if it is not separable as:

$$\rho_{AB} \neq \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}$$

This corresponds to a non-classical correlation which cannot be expressed as a convex combination of local states.

Hence, EPR assumed:

$$Reality(AB) = Reality(A) + Reality(B)$$

Entanglement is the negation of the above statement. Therefore, the reality in Quantum Mechanics is non-separable.

6.4 Measurement and Entropy of Entangled Pure States

Let S be the system and M corresponding to the apparatus, the initial state is:

$$|\psi\rangle_S = \sum_i c_i |i\rangle, \quad |0\rangle_M$$

such that:

$$U_{SM}(|i\rangle_S |0\rangle_M) = |i\rangle_S |m_i\rangle_M$$

Therefore, we have the entangled pure state $|\Psi\rangle_{SM}$

$$|\Psi\rangle_{SM} = \sum_i c_i |i\rangle_S |m_i\rangle_M$$

Where the reduced state ρ_S corresponds to a mixed state:

$$\rho_S = \mathrm{Tr}_M(|\Psi\rangle\langle\Psi|) = \sum_i |c_i|^2 |i\rangle\langle i|$$

On a projective measurement, the state collapses to:

$$\rho \to \frac{P_i \rho P_i}{\mathrm{Tr}(P_i \rho)}$$

POVMs are required because we cannot measure the full entangled state. Therefore, limited access required the development of POVMs

Suppose Alice performs a POVM $\{E_k\}$on the entangled state ρ_{AB}. The conditional state of Bob is then:

$$\rho_{B|k} = \frac{\mathrm{Tr}_A[(E_k \otimes I)\rho_{AB}]}{p(k)}$$

This would seem to break locality, but the key fact to take into consideration is that Bob's unconditional state is unchanged, i.e. avering out the result restores locality. Further, after measurement, the global system-apparatus state turns into a classical-quantum state after decoherence, where quantum coherence turns into classical records.

$$\rho_{SM}^{post} = \sum_i p_i |i\rangle\langle i|_M \otimes \rho_S^{(i)}$$

This allows measurement outcomes to be separable, as well as satisfy Bell's Inequality.

To quantify entanglement or answer how much the subsystems of a bipartite quantum $|\psi\rangle_{AB} \in \mathscr{H}_A \otimes \mathscr{H}_B$ state entangle requires us to define the entropy of Entanglement where the entropy of a subsystem measures how strongly it is correlated with the rest of the universe.

Von Neumann Entropy/ Entropy of Entanglement for a pure state
The von Neumann Entropy of either reduced state:

$$E(|\psi\rangle_{AB}) = S(\rho_A) = S(\rho_B)$$

where:

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho)$$

The equality for entropy of the two reduced states holds if and only if the global state is a pure state.

Shannon Entropy for a pure state
For $|\psi\rangle_{AB} \in \mathscr{H}_A \otimes \mathscr{H}_B$, the Schmidt decomposition describes the reduced state ρ_A as:

$$\rho_A = \sum_i \lambda_i |i_A\rangle\langle i_A|$$

Hence, we define the Shannon Entropy as:

$$E(X) = -\sum_i \lambda_i \log \lambda_i$$

Therefore Entropy of entanglement is exactly the Shannon entropy of the Schmidt coefficients.
The entropy is zero or minimal for product states, and the maximum entropy is for an d dimensional entangled state is:

$$E_{max} = \log(d)$$