The Framework of Quantum Mechanics

We now use the mathematical formalism covered in the last lecture to describe the theory of quantum mechanics. In the first section we outline four axioms that lie at the foundation of quantum mechanics. From these four axioms we derive in the second section an effective theory of quantum mechanics that applies to open quantum systems.

Contents

1 The Axioms of Quantum Mechanics .......................... 1
  1.1 The State Space Axiom ........................................... 1
  1.2 The Unitary Evolution Axiom ................................... 2
  1.3 The Measurement Axiom ......................................... 2
  1.4 The Composite System Axiom ................................... 4

2 Describing Quantum Systems using Density Matrices .............. 5
  2.1 Epistemic versus Ontological Uncertainty of a System’s State .... 5
  2.2 Properties of Density Matrices and Purifications ................ 8

3 Generalized Measurements ........................................ 12
  3.1 Physically Implementable Processes ................................ 12
  3.2 Completely Positive Maps and Quantum Instruments .............. 13
  3.3 Special Types of Generalized Measurements ..................... 16
    3.3.1 Quantum Channels ......................................... 16
    3.3.2 Measurement Instruments and POVMs ....................... 16
    3.3.3 Fine-grained Instruments ................................... 17
    3.3.4 Projective Measurements ................................... 17

4 Exercises .................................................. 18

1 The Axioms of Quantum Mechanics

The basic framework of quantum mechanics consists of four axioms, and the subject of quantum mechanics involves studying what consequences follow from these axioms. This axiomatic approach is justified by repeated experimental verification of the theoretical predictions it makes.

1.1 The State Space Axiom

Every quantum system is represented by a Hilbert space $\mathcal{H}$ called state space. Every ray in the Hilbert space corresponds to a physical state of the system.
Rays in a vector space are simply one-dimensional subspaces. The state space axiom therefore says that states of a quantum systems are identified with an entire one-dimensional subspace \( \{ \alpha | \psi \rangle : \alpha \in \mathbb{C} \} \), where \( | \psi \rangle \) is some unit vector. When performing calculations the convention is to represent physical states by unit vectors, and so henceforth we will assume that all kets are normalized unless otherwise stated. It is understood that multiplying this ket by an overall scalar does not change the physical state it represents. In particular, multiplying \( | \psi \rangle \) by an overall phase \( e^{i \phi} \) does not change any of the experimental outcomes predicted in quantum mechanics. This will be explained in greater detail when we discuss the Measurement Axiom.

There is another type of phase that does distinguish one state from another and which can lead to different experimental predictions. For a state \( | \psi \rangle \) decomposed in a linear combination

\[
| \psi \rangle = \alpha |0\rangle + \beta |1\rangle,
\]

a relative phase is a factor \( e^{i \phi} \) that is multiplied to just one of the kets but not both. For example, we would say that the vector

\[
| \psi' \rangle = \alpha |0\rangle + \beta e^{i \phi} |1\rangle
\]

differs from the vector \( | \psi \rangle \) by an relative phase, and the two represent different physical states of the system.

The State Space Axiom does not tell you the dimension of the Hilbert space associated with a given physical system. Nor does the State Space Axiom tell you which physical states of the system correspond to which rays in the Hilbert space. How both of these assignments are made depends on experimental observation. To understand this better, we need a way to describe observation and measurement in quantum mechanics. The Measurement Axiom will stipulate how this is done, but first we address of how quantum systems evolve in time.

### 1.2 The Unitary Evolution Axiom

Every isolated quantum system evolves unitarily in time.

In more detail, let \( \mathcal{H} \) be the state space for an isolated quantum system, and consider any two moments in time, \( t_0 \) and \( t_1 \). In principle it is never possible to perfectly isolate one system from the rest of the universe; this is one of the fundamental challenges in building quantum computers and related technologies. However, by carefully controlling the laboratory environment, a quantum system can be shielded from any external interaction to a sufficient degree such that treating it as an isolated system is a good approximation. The axiom says that there exists a unitary operator \( U(t_0, t_1) \) acting on \( \mathcal{H} \) such that if the system is in state \( | \psi \rangle \) at time \( t_0 \), it will be in state \( U(t_0, t_1) | \psi \rangle \) at time \( t_1 \). Since the evolution is unitary, \( U(t_0, t_1) | \psi \rangle \) is still a unit vector. In addition, the entire process can be reversed by applying the unitary \( U(t_0, t_1)^\dagger \).

### 1.3 The Measurement Axiom

Let \( \mathcal{H} \) be the state space of some quantum system. A measurement on system \( \mathcal{H} \) is represented by a hermitian operator \( O \in \mathcal{L}(\mathcal{H}) \) called an observable, and conversely, every hermitian operator \( O \in \mathcal{L}(\mathcal{H}) \) corresponds to a physical measurement on \( \mathcal{H} \).

For an observable \( O \) with spectral decomposition \( O = \sum_{k=1}^{n} \lambda_k P_{\lambda_k} \), its eigenvalues \( \lambda_k \) are the different values that can be measured when performing the measurement

\[
\sum_{k=1}^{n} \lambda_k P_{\lambda_k} \sum_{i=1}^{m} P_{\lambda_i} = \sum_{k=1}^{n} \lambda_k P_{\lambda_k}.
\]
described by $O$. If $|\psi\rangle \in \mathcal{H}$ is the pre-measurement state of the system, then value $\lambda_k$ will be measured with probability

$$p(\lambda_k) = \langle \psi | P_{\lambda_k} | \psi \rangle. \tag{1}$$

When $\lambda_k$ is measured, the post-measurement state of the system is given by

$$\frac{1}{\sqrt{p(\lambda_k)}} P_{\lambda_k} | \psi \rangle. \tag{2}$$

There is quite a bit being asserted in the Measurement Axiom. The essential idea is that measurement in quantum mechanics involves a *stochastic* process. That is, one can only assign a probability distribution $\{p(\lambda_k)\}_{k=1}^n$ to the $n$ different outcomes of a quantum measurement. In general, it cannot be known in advance what the outcome of a quantum measurement will be with full certainty. The Measurement Axiom also shows why overall phases have no physical meaning in quantum mechanics. If we have two states $|\psi\rangle$ and $|\psi'\rangle = e^{i\theta} |\psi\rangle$, then

$$p(\lambda_k) = \langle \psi | P_{\lambda_k} | \psi \rangle = \langle \psi' | P_{\lambda_k} | \psi' \rangle.$$ 

In other words, all measurement outcomes are predicted to occur with the same probability whether we identify a physical state by $|\psi\rangle$ or by $e^{i\theta} |\psi\rangle$.

Equation (2) describes the state of the system after the measurement. Conditioned on outcome $\lambda_k$, the system undergoes the transformation

$$|\psi\rangle \rightarrow \frac{1}{\sqrt{p(\lambda_k)}} P_{\lambda_k} | \psi \rangle. \tag{3}$$

The factor of $\frac{1}{\sqrt{p(\lambda_k)}}$ on the post-measurement state is to ensure that it is normalized. This transformation is sometime referred to as “collapse of the wave function.” In this course we are not dealing with wave functions since all systems are finite-dimensional, so a better fitting description of Eq. (3) would be “collapse of the state vector.” More precisely, Eq. (3) says that the state of the system is projected into the $\lambda_k$ eigenspace of observable $X$ whenever $\lambda_k$ is measured.

States represented by eigenvectors of an observable are called eigenstates, and they have a special status in the measurement process. Let $|\lambda_k\rangle$ be any eigenvector in the eigenspace $V_{\lambda_k}$ of $X$. Then $P_{\lambda_l} |\lambda_k\rangle = \delta_{lk} |\lambda_k\rangle$, which means that the if $l \neq k$, there is zero probability of obtaining outcome $\lambda_l$ when the system is prepared in state $|\lambda_k\rangle$. The only possible outcome is $\lambda_k$, and the state $|\lambda_k\rangle$ remains unchanged in the measurement process. For this reason, eigenstates of an observable are referred to as *stationary states*.

Moreover, if an arbitrary state $|\psi\rangle$ is measured and outcome $\lambda_k$ is obtained, then quantum mechanics predicts that $\lambda_k$ will necessarily be obtained again if the same measurement is immediately performed a second time. The reason is that after the first measurement the system has collapsed into a stationary state $\frac{1}{\sqrt{p(\lambda_k)}} P_{\lambda_k} | \psi \rangle$. Typically, the system will begin undergoing some nontrivial unitary evolution as soon as the first measurement is complete. This will generally cause the system to leave the stationary state, and thus the outcome of the second measurement will regain uncertainty as time progresses between the two measurements.

We can now answer the questions of (i) how to decide the dimension of state space associated with some physical system, and (ii) how to assign the physical states of the system to rays in state space. For the first question, one could prepare multiple copies of the same physical system and carefully perform the same type of experimental observation on each of the copies. Due to the
stochastic nature of quantum measurement, different outcomes will be observed among the different copies. The experimenter counts the total number of distinct outcomes, and this represents one estimate on the dimension of the quantum system. Indeed, by the measurement axiom, whatever observation the experimenter performs, it must be described by a hermitian operator. The eigenspaces of the operator provide a direct sum decomposition of state space with the number of eigenspaces corresponding to the number of different outcomes observed in the experiments. If all the eigenspaces are one-dimensional, then this number is precisely the dimension of state space. However, if there are degenerate eigenvalues, then this number will be strictly less than the dimension. In the latter case, the experimenter would need to perform “finer-grained” measurements described by observables not having degeneracy on this eigenspace. It is not a priori given what these finer-grained measurements would be if, in fact, there is degeneracy. Instead, one usually first has a theoretical model hypothesizing the Hilbert space structure of the system, and then this model is experimentally tested.

For the second question, experimental observation also places restrictions on the correspondence between physical states and kets in state space. We have discussed above how eigenvectors of an observable must be assigned to the stationary states of the measurement in order to be consistent with the Measurement Axiom. By the spectral decomposition of a given observable, it follows that in a $d$-dimensional Hilbert space, there exists $d$ stationary states for the observable whose corresponding kets form an orthonormal basis for state space. More generally, a physical state can be represented by a ket $|\psi\rangle$ if measurement outcome $\lambda_k$ is observed at frequency $p(\lambda_k) = \langle \psi | P_{\lambda_k} | \psi \rangle$ whenever the system is prepared in that state and the observable $X = \sum_{k=1}^{n} \lambda_k P_{\lambda_k}$ is performed.

For example, the Stern-Gerlach experiment revealed that the directional spin of an electron has two possible values. Thus, an electron spin system is described by a two-dimensional Hilbert space; i.e. $\mathcal{H} \cong \mathbb{C}^2$. We can represent the $\hat{z}$-direction “spin-up” eigenstate by ket $|0\rangle$ and the “spin-down” eigenstate by $|1\rangle$. Any other spin state of the electron is represented by a superposition of these states

$$|\psi\rangle = a|0\rangle + b|1\rangle,$$

where $|a|^2 + |b|^2 = 1$. Here, $|a|^2$ and $|b|^2$ represent the frequency of observing “spin-up” and “spin-down,” respectively, when the system is prepared in state $|\psi\rangle$ and spin is measured in the $z$-direction. From our discussion on the Stern-Gerlach experiment, an $x$-direction “spin-up” eigenstate measures “spin-up” and “spin-down” in the $z$-direction with equal probability. Thus if we let $|+\rangle$ denote the $x$-direction “spin-up” eigenstate, we must have $|+\rangle = a|0\rangle + b|1\rangle$ with $|a|^2 = |b|^2 = 1/2$. Up to an overall phase, $|+\rangle$ has the form $|+\rangle = \sqrt{1/2}(|0\rangle + e^{i\theta}|b\rangle)$. The $x$-direction “spin-down” eigenstate is orthogonal to $|+\rangle$, and thus has the form $|-\rangle = \sqrt{1/2}(|0\rangle - e^{i\theta}|b\rangle)$. Customarily, the relative phase $e^{i\theta}$ for the $x$-direction eigenstates is chose as 1.

A final remark on the Measurement Axiom involves taking the expectation value of a quantum measurement. Suppose that $O = \sum_{k=1}^{n} \lambda_k P_{\lambda_k}$ is some quantum observable. By the Measurement Axiom, for pre-measurement state $|\psi\rangle$ the outcomes $\lambda_k$ are distributed according to the distribution $p(\lambda_k) = \langle \psi | P_{\lambda_k} | \psi \rangle$. This means that the expected value $\langle O \rangle$ of the measurement is given by

$$\langle O \rangle = \sum_{k=1}^{n} \lambda_k \langle \psi | P_{\lambda_k} | \psi \rangle = \langle \psi | \sum_{k=1}^{n} \lambda_k P_{\lambda_k} | \psi \rangle = \langle \psi | O | \psi \rangle.$$

\hspace{1cm} (4)

### 1.4 The Composite System Axiom

If $A$ and $B$ are two quantum systems with state spaces $\mathcal{H}^A$ and $\mathcal{H}^B$ respectively, the state space of their combined physical system is the tensor product space $\mathcal{H}^{AB} := \mathcal{H}^A \otimes \mathcal{H}^B$. For instance, a three-level system could be formed by tensoring two two-level systems.
\(\mathcal{H}^A \otimes \mathcal{H}^B\). For more systems, \(\mathcal{H}^{A_1}, \mathcal{H}^{A_2}, \ldots, \mathcal{H}^{A_n}\), the joint state space is the \(n\)-fold tensor product space \(\mathcal{H}^{A_1} \otimes \mathcal{H}^{A_2} \otimes \cdots \otimes \mathcal{H}^{A_n}\). Two quantum systems are often called a bipartite system while more than two systems are typically referred to as a multipartite system.

We already have good practice working with tensor product spaces in the previous lecture. Recall we emphasized that \(\mathcal{H}^A \otimes \mathcal{H}^B\) is much larger than the set of tensor product states. As we will see, the difference between these two sets has dramatic physical consequences. A bipartite state \(|\psi\rangle^{AB}\) is entangled if it is not a tensor product state; i.e. \(|\psi\rangle^{AB} \neq |\alpha\rangle^A |\beta\rangle^B\). A non-entangled state is also called a product or separable state. In multipartite systems, a state \(|\psi\rangle^{A_1A_2\ldots A_n}\) is entangled if it is not an \(n\)-fold tensor product of states; i.e. \(|\psi\rangle^{A_1A_2\ldots A_n} \neq |\alpha_1\rangle^{A_1}|\alpha_2\rangle^{A_2} \cdots |\alpha_n\rangle^{A_n}\). However, note that under this definition, states of the form \(|\psi\rangle^{AB}|\psi\rangle^C\) are considered entangled whenever \(|\psi\rangle^{AB}\) is entangled, despite the fact that system \(C\) is in a product state with \(A\) and \(B\). In multipartite systems, more interesting entanglement structures emerge than what is found in the bipartite case. For the time being we will only focus on bipartite systems and return to the multipartite scenario later in the course.

Let us consider how to determine whether a given \(|\psi\rangle \in \mathcal{H}^{AB}\) is entangled. The following criterion for entanglement follows almost trivially from the isomorphism \(\mathcal{H}^{AB} \cong L(\mathcal{H}^B, \mathcal{H}^A)\) established in the last lecture.

**Proposition 1.** A bipartite state \(|\psi\rangle\) is a product state iff its operator representation \(M_{|\psi\rangle}\) is rank one.

**Proof.** If \(|\psi\rangle = |\alpha\rangle |\beta\rangle\), then \(M_{|\psi\rangle} = |\alpha\rangle \langle \beta|\) which is a rank one operator. Conversely, if \(M_{|\psi\rangle} = |\alpha\rangle \langle \beta|\) is rank one, then \(|\psi\rangle = M_{|\psi\rangle} \otimes I |\psi\rangle = |\alpha\rangle |\beta\rangle\).

From elementary matrix theory, a matrix has rank \(r\) iff all of its \((r+1)\)-minors vanish, and there exists at least one nonvanishing \(r\)-minor. In \(2 \times 2\) matrices, this reduces to the condition that a nonzero matrix has vanishing determinant. Thus, as an example consider an arbitrary two-qubit state \(|\psi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle\). The operator \(M_{|\psi\rangle}\) has matrix representation \(M_{|\psi\rangle} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\), whose rank is one iff its determinant is zero. That is, the necessary and sufficient condition for \(|\psi\rangle\) to be separable is that \(ad - bc = 0\).

## 2 Describing Quantum Systems using Density Matrices

### 2.1 Epistemic versus Ontological Uncertainty of a System’s State

The State Space Axiom stipulates that every normalized vector in Hilbert space (more precisely every ray) corresponds to a physically realizable state of the system. However, in many situations it may not be possible to assign a definite state vector to the system. Traditionally these situations are classified as being either epistemological or ontological.

In the case of epistemological uncertainty, a state vector \(|\psi\rangle\) might be the appropriate description of a system’s state at a given moment in time, but sheer ignorance of this fact prevents the experimenter from assigning a definite state vector to the system. Instead the experimenter may have gained only enough information about the system to describe it as being an ensemble of possibilities. More formally, a discrete ensemble of states is a probability space \(\mathcal{E} = (\Omega, p)\), where \(\Omega = \{\psi_x\}_{x \in \mathcal{X}}\) is a sample space consisting of possible states for the system and \(p\) is a distribution on \(\Omega\). The ensemble \(\mathcal{E}\) then reflects the experimenter’s statistical description of the system. The experimenter would agree that the system is in some definite state belonging to \(\Omega\), but because of insufficient knowledge, he/she can only assign a probability \(p_x\) to the system being in state \(|\psi_x\rangle\).
Usually we will denote an ensemble of states by $E = \{ |\psi_x\rangle, p_x \}_{x \in \mathcal{X}}$. Without loss of generality, we can assume that $\mathcal{X} \subset \mathbb{R}$, and we let $X$ be the random variable ranging over $\mathcal{X}$ with distribution $p_X = p$. The values of random variable $X$ are thus interpreted as classical labels for the different quantum states belonging to the ensemble.

If the states in ensemble $E$ belong to Hilbert space $\mathcal{H}$, then its ensemble average is the operator on $L(\mathcal{H})$ defined by

$$\rho = \sum_{x \in \mathcal{X}} p_x |\psi_x\rangle\langle \psi_x|.$$ (5)

The ensemble average is also called the density operator or density of states generated by the particular ensemble. The density operator is the key tool for computing probabilities of measurement outcomes when the system is described by an ensemble of states. Suppose that $O = \sum_k \lambda_k P_k$ is some observable on a system described by ensemble $\{ |\psi_x\rangle, p_x \}_{x \in \mathcal{X}}$. When preforming the measurement, there are now two probabilities to consider: (i) $p(\lambda_k)$, the probability of measurement outcome $\lambda_k$, and (ii) $p_x$, the probability that the system is in pre-measurement state $|\psi_x\rangle$. The joint distribution $p(\lambda_k, x)$ is the probability that outcome $\lambda_k$ is obtained and the pre-measurement state is $|\psi_x\rangle$. This can be computed using the Measurement Axiom, which says that given state $|\psi_x\rangle$, the conditional probability of outcome $\lambda_k$ is

$$p(\lambda_k|\psi_x) = \langle \psi_x | P_k | \psi_x \rangle,$$ (6)

and the post-measurement state will be

$$\frac{P_k |\psi_x\rangle}{\sqrt{p(\lambda_k|\psi_x)}}.$$ (7)

Hence the overall probability of obtaining outcome $\lambda_k$ is given by

$$p(\lambda_k) = \sum_x p(\lambda_k|x) p_x = \sum_x p_x \langle \psi_x | P_k | \psi_x \rangle$$

$$= \sum_x p_x \text{Tr} [ |\psi_x\rangle\langle \psi_x| P_k ] = \text{Tr} \left[ \sum_x p_x |\psi_x\rangle\langle \psi_x| P_k \right] = \text{Tr} [ \rho P_k ].$$ (8)

In an actual experimental setting, the experimenter observes the outcome $\lambda_k$, and this information can then be used to better infer the state of the system. According to Bayes’ rule, the probability that the pre-measurement state was $|\psi_x\rangle$ given outcome $\lambda_k$ is

$$p(x|\lambda_k) = \frac{p(\lambda_k|x) p_x}{p(\lambda_k)}.$$ (9)

Therefore, given outcome $\lambda_k$, the experimenter’s post-measurement description of the system is given by the post-measurement ensemble

$$E_{\lambda_k} = \left\{ \frac{P_k |\psi_x\rangle}{\sqrt{p(\lambda_k|x)}}, p(x|\lambda_k) \right\}_x.$$ (10)

As an analog to Eq. (3), conditioned on outcome $\lambda_k$ which occurs with probability $p(\lambda_k)$, the density matrix transforms as

$$\rho = \sum_{x \in \mathcal{X}} p_x |\psi_x\rangle\langle \psi_x| \rightarrow \sum_{x \in \mathcal{X}} p(x|\lambda_k) \frac{P_k |\psi_x\rangle\langle \psi_x| P_k}{p(\lambda_k|x)} = \frac{1}{p(\lambda_k)} P_{\lambda_k} \rho P_{\lambda_k}. $$ (11)
In most interpretations of quantum mechanics, there is a second type of uncertainty pertaining to the state of a system. This is often called ontological uncertainty, and it is a consequence of entanglement. If systems $A$ and $B$ are in some entangled state $|\psi^{AB}\rangle$, then by the definition of $|\psi\rangle$ being entangled, state vectors $|\alpha^A\rangle$ and $|\beta^B\rangle$ cannot be assigned to either subsystem. This is sometimes described as being an ontological fact. Rather than reflecting someone’s subjective statistical description, it is an objective property of entangled systems that subsystems cannot be represented by definite state vectors. The philosophical implications of such a statement are hotly debated. Does this mean that entangled subsystems truly “don’t exist” in individual physical states, or does the mathematical formalism simply fall short of capturing all that really does exist? We will not pursue this question any further since even in phrasing it there are philosophical landmines we have started to trigger. Instead, we will find an effective way to describe entangled subsystems so that accurate measurement predictions can be made. As in the epistemological case, this will involve the density matrix.

Suppose that $O^A = \sum_k \lambda^A_k P^A_k$ and $O^B = \sum_k \lambda^B_k P^B_k$ are two quantum observables for Alice and Bob’s systems respectively. If their pre-measurement state is $|\psi^{AB}\rangle$, then the Measurement Axiom says that the joint distribution of measurement outcomes is given by

$$p(\lambda^A_j, \lambda^B_k) = \langle \psi | P^A_j \otimes P^B_k | \psi \rangle.$$  \hspace{1cm} (12)

We are now interested in the reduced distribution of outcomes for, say, system $A$. To obtain this, we sum over the $\lambda^B_k$ outcomes,

$$p(\lambda^A_j) = \sum_k p(\lambda^A_j, \lambda^B_k) = \sum_k \langle \psi | P^A_j \otimes P^B_k | \psi \rangle = \langle \psi | P^A_j \otimes \mathbb{I} | \psi \rangle = \text{Tr} \left[ P^A_j \otimes \mathbb{I} | \psi \rangle \langle \psi | \right] = \text{Tr} \left[ P^A_j \rho^A \right],$$  \hspace{1cm} (13)

where $\rho^A = \text{Tr}_B |\psi\rangle\langle\psi^AB|$ is called the reduced density operator or reduced density matrix of $|\psi\rangle\langle\psi^A|^{AB}$ for system $A$. In deriving this formula, we used the fact that $\sum_k P^B_k = \mathbb{I}$. Thus, we obtain the correct marginal distribution for measurement outcomes on system $A$ if we use the reduced density operator and the rule $p(\lambda^A_j) = \text{Tr} \left[ P^A_j \rho^A \right]$.

We have previously defined the density operator as the ensemble average for some ensemble of states $\mathcal{E}$. The reduced density operator is indeed a density operator in this sense. To see explicitly see this, note that $|\psi\rangle\langle\psi|^{AB}$ is a positive operator, and so its partial trace is also positive (see Exercise 10 of the previous lecture). Furthermore, $1 = \text{Tr} (\rho^A) = \langle \psi | \psi \rangle$, which means that the eigenvalues of $\rho^A$ sum up to unity (including multiplicities). Thus an eigenbasis of $\rho^A$ along with the associated eigenvalues provides an ensemble $\mathcal{E}$ whose ensemble average is precisely $\rho^A$. We can also see this directly from a Schmidt decomposition of $|\psi\rangle = \sum_{i=1}^{r} \sigma_i |\alpha_i\rangle |\beta_i\rangle$. The normalization condition $1 = \langle \psi | \psi \rangle$ implies that $\sum_{i=1}^{r} \sigma_i^2$, and the orthonormality of the $|\beta_i\rangle$ yields $\text{Tr}_A |\psi\rangle\langle\psi| = \sum_{i=1}^{r} \sigma_i^2 |\alpha_i\rangle \langle \alpha_i |$. Therefore, we can interpret $\rho^A$ as the ensemble average of $\mathcal{E} = \{ |\alpha_i\rangle, \sigma_i^2 \}_{i=1}^{r}$. However in this case, we should not interpret $\mathcal{E}$ as a mere statistical since no definite state vector can be assigned to an entangled subsystem, regardless of how much knowledge the experimenter has gained about the system. Nevertheless, from a mathematical perspective, every reduced density matrix is equivalent to an ensemble average.

Equation (13) also holds when systems $A$ and $B$ are themselves described by a joint density matrix $\rho^{AB} = \sum_x p_x |\psi_x\rangle\langle\psi_x|^{AB}$. Whether this density matrix description reflects epistemic un-
certainty of the joint state of system $AB$, or whether this density matrix reflects ontological un-
certainty and $\rho^{AB}$ is the reduced density operator of some tripartite entangled state $|\psi\rangle^{ABC}$, or
whether this is a mixture of both, Eqns. (8) and (13) give the same rule for computing probabilities
for measurement outcomes:

$$p(\lambda_j^A, \lambda_k^B) = \text{Tr}[P_j^A \otimes P_k^B \rho^{AB}].$$

The marginal distribution is then given by

$$p(\lambda_j^A) = \sum_k \text{Tr}[P_j^A \otimes P_k^B \rho^{AB}] = \text{Tr}[P_j^A \rho^A],$$

(14)

where $\rho^A = \text{Tr}_B \rho^{AB}$. Hence, the reduced density operator of $\rho^{AB}$ is the correct mathematical object
for computing the marginal distribution of measurement outcomes.

In summary, whether one’s ignorance of a system’s state is epistemological or ontological,
Eqns. (8) and (13) show that the process for computing measurement outcome probabilities is
the same. One describes the quantum system using a density operator. For a bipartite density
operator $\rho^{AB}$ on state space $\mathcal{H}^{AB}$, the density operator for subsystem $A$ is obtained by tracing out
the other subsystem, $\rho^A = \text{Tr}_B \rho^{AB}$. Then the process of quantum measurement is described by
the following rule.

If a quantum system is described by density operator $\rho$ and $O = \sum_k \lambda_k P_{\lambda_k}$ is some
quantum observable, then the probability of obtaining measurement outcome $\lambda_k$ is

$$p(\lambda_k) = \text{Tr}[\rho P_{\lambda_k}].$$

(15)

Conditioned on outcome $\lambda_k$, the post-measurement description of the system is density
operator

$$\frac{1}{p(\lambda_k)} P_{\lambda_k} \rho P_{\lambda_k}.$$

(16)

This rule can be seen as a generalization of the Measurement Axiom to density operators. Indeed,
if $\rho = |\psi\rangle \langle \psi|$ is a rank-one projector, then this rule reduces precisely to the Measurement Axiom
presented above. Density operators are used so frequently in quantum mechanics that it is cus-
tomary to refer to them as “states” of the same system, even though they are generally not true
state vectors in state space. To distinguish the two cases, $\rho$ is called a pure state if it is a rank-one
projector, and it is called a mixed state if its rank is greater than one.

### 2.2 Properties of Density Matrices and Purifications

Density operators have a number of important properties that we now discuss. By definition, a
density operator $\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$ is the ensemble average of some ensemble $\mathcal{E} = \{|\psi_i\rangle, p_i\}_i$. From
its form, we note that

$$\text{Tr}(\rho) = \sum_i p_i \text{Tr}(|\psi_i\rangle \langle \psi_i|) = \sum_i p_i = 1.$$

Furthermore, for any $|\varphi\rangle$ we have

$$\langle \varphi | \rho | \varphi \rangle = \sum_i p_i \langle \varphi | \psi_i \rangle \langle \psi_i | \varphi \rangle = \sum_i p_i |\langle \varphi | \psi_i \rangle|^2 \geq 0.$$

Thus we have established the first basic property of a density matrix.
Proposition 2. Every density operator is a trace-one positive operator.

We next turn to the notion of a purifying system. If $\rho^A$ is a mixed state on $\mathcal{H}^A$, then a purification of $\rho^A$ is a bipartite pure state $|\Psi\rangle^{AB} \in \mathcal{H}^{AB}$ for some second system $B$ such that $\text{Tr}_B |\Psi\rangle^{AB} = \rho^A$. In the state $|\Psi\rangle^{AB}$, system $B$ is systems referred to as the purifying system or $\rho^A$. There are many different purifications and purifying systems for a given mixed state. For example, consider the totally mixed state on a qubit system. In general, the totally mixed state on a $d$-dimensional system is the normalized identity: $\rho = \mathbb{I}/d$. For a qubit this takes the form

$$\rho = \frac{1}{2} (|0\rangle\langle 0| + |1\rangle\langle 1|).$$

Two different purifications of this state are

$$|\Psi\rangle^{AB} = \frac{1}{\sqrt{2}}(|0\rangle |0\rangle + |1\rangle |1\rangle)$$

$$|\Psi'\rangle^{AB} = \sqrt{\frac{2}{3}}(|\varphi_0\rangle |0\rangle + |\varphi_1\rangle |1\rangle + |\varphi_2\rangle |2\rangle),$$

where

$$|\varphi_0\rangle = |0\rangle, \quad |\varphi_1\rangle = \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle, \quad |\varphi_2\rangle = \frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle.$$ (19)

It is easy to verify that $\text{Tr}_B |\Psi\rangle^{AB} = \rho$. Notice that the purifying system $B$ is two-dimensional while $B'$ is three-dimensional. Physically this may reflect a scenario where, say, $B$ is some spin-1/2 system while $B'$ is some spin-1 system. However, mathematically we can think of $\mathcal{H}^B$ as just being some subspace of $\mathcal{H}^{B'}$. In general then, when comparing two purifications $|\Psi\rangle$ and $|\Psi'\rangle$ of the same reduced density matrix $\rho$, we can mathematically treat the purifying systems as being the same system; the difference is that the support spaces of the purifying systems will not necessarily be the same in the two states. This begs the question of whether there is any mathematical relationship between two states $|\Psi\rangle^{AB}$ and $|\Psi'\rangle^{AB}$ that purify the same density operator $\rho^A$. The next lemma identifies the relationship.

Lemma 1. Two states $|\Psi\rangle^{AB}$ and $|\Psi'\rangle^{AB}$ in $\mathcal{H}^{AB}$ purify the same mixed state $\rho^A$ iff $|\Psi\rangle = \mathbb{I} \otimes U |\Psi'\rangle$ for some unitary $U$ acting on $\mathcal{H}^B$.

Proof. If $|\Psi\rangle = \mathbb{I} \otimes U |\Psi'\rangle$, then obviously $\text{Tr}_B |\Psi\rangle \otimes |\Psi'\rangle = \text{Tr}_B |\Psi'\rangle \otimes |\Psi'\rangle$. The converse follows from Corollary 1 of the previous lecture. Indeed, if both $|\Psi\rangle$ and $|\Psi'\rangle$ have the same reduced state on system $A$, then they have Schmidt decompositions of the form

$$|\Psi\rangle = \sum_{i=1}^{r} \sigma_i |\alpha_i\rangle |\beta_i\rangle$$

$$|\Psi'\rangle = \sum_{i=1}^{r} \sigma_i |\alpha_i\rangle |\beta'_i\rangle.$$ 

Hence $|\Psi\rangle = \mathbb{I} \otimes U |\Psi'\rangle$, with $U$ being a unitary such that $U |\beta'_i\rangle = |\beta_i\rangle$ for all $i = 1, \cdots, r$. \qed

From the example of Eqns. (17) and (18), we see that $\mathcal{E} = \{ |\psi\rangle, \frac{1}{2} \}_{i=0}^{1}$ and $\mathcal{E}' = \{ |\varphi\rangle, \frac{3}{2} \}_{i=0}^{2}$ are two different ensembles with the same ensemble average $\mathbb{I}/2$. The first ensemble can be "padded with zeros" so that it has the same size as the second. In general, "padding with zeros" is an
expression refers to adding zero elements to some set so that it size is of desired length. In this case, we add an arbitrary state \(|\psi_2\rangle\) to \(\mathcal{E}\) that occurs with probability \(p_2 = 0\). Similar to the question asked in Lemma 1, we can now ask whether there is a relationship between any two ensembles that generate the same density matrix. The following lemma provides the answer.

**Lemma 2.** Two pure state ensembles \(\mathcal{E} = \{ |\psi_i\rangle, p_i \}_{i=1}^{l'}\) and \(\mathcal{E}' = \{ |\psi'_i\rangle, p'_i \}_{i=1}^{l'}\) have the same ensemble average, i.e., \(\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| = \sum_i p'_i |\psi'_i\rangle \langle \psi'_i|\), iff there exists an \(l_0 \times l_0\) unitary matrix \(U\) with elements \(u_{ij}\) such that

\[
\sqrt{p'_i} |\psi'_i\rangle = \sum_{j=1}^{l_0} u_{ij} \sqrt{p_j} |\psi_j\rangle \quad \text{for } i = 1, \ldots, l_0
\]  

where \(l_0 = \max\{l, l'\}\), and the ensemble with the smallest number of states is “padded” with a sufficient number of zero-probable states so that it has \(l_0\) elements.

**Proof.** Suppose that Eq. (20) holds. Then

\[
\sum_{i=1}^{l_0} p'_i |\psi'_i\rangle \langle \psi'_i| = \sum_{j=1}^{l_0} u_{ij} \sqrt{p_j} \sqrt{p'_j} |\psi_j\rangle \langle \psi'_j| = \sum_{i=1}^{l_0} p_i |\psi_i\rangle \langle \psi_i|,
\]

where we use the fact that the \(u_{ij}\) constitute a unitary matrix. Conversely, suppose that \(\rho = \sum_{i=1}^{l_0} p'_i |\psi'_i\rangle \langle \psi'_i| = \sum_{i=1}^{l_0} p_i |\psi_i\rangle \langle \psi_i|\). Then the bipartite states

\[
|\Psi\rangle^{AB} = \sum_{i=1}^{l_0} \sqrt{p_i} |\psi_i\rangle^A |\psi_i\rangle^B
\]

\[
|\Psi'\rangle^{AB} = \sum_{i=1}^{l_0} \sqrt{p'_i} |\psi'_i\rangle^A |\psi'_i\rangle^B
\]

provide two purifications of \(\rho\) with an \(l_0\)-dimensional purifying system \(B\). Hence by Lemma 1 there exists a unitary \(U\) acting on \(\mathcal{H}^B\) such that \(|\Psi\rangle = I \otimes U |\Psi'\rangle\). The action of \(U\) on basis vector \(|\psi_i\rangle^B\) has the form \(U |\psi_i\rangle^B = \sum_{j=1}^{l_0} u_{ij} |\psi_i\rangle^B\). Thus, we have

\[
\sum_{i=1}^{l_0} \sqrt{p'_i} |\psi'_i\rangle^A |\psi'_i\rangle^B = \sum_{j=1}^{l_0} \sqrt{p_j} |\psi_j\rangle^A U |\psi_j\rangle^B
\]

\[
= \sum_{i=1}^{l_0} \sum_{j=1}^{l_0} u_{ij} \sqrt{p_j} |\psi_i\rangle^A |\psi'_i\rangle^B.
\]

Applying a partial contraction of \(|\psi_i\rangle^B\) on both sides generates the condition of Eq. (20). \(\square\)

We finally connect Lemmas 1 and 2 in terms of a physical task performed by two parties, called Alice and Bob. Suppose that Alice and Bob hold a bipartite pure state \(|\Psi\rangle^{AB}\) on joint system \(\mathcal{H}^{AB}\). In this case Bob holds a purification of Alice’s reduced state \(\rho^A\), and Lemma 1 characterizes all pure-state ensembles \(\mathcal{E}\) whose ensemble average is \(\rho^A\). As noted above, because of the entanglement, these ensembles do not reflect statistical descriptions for Alice or anyone else. However, for a given ensemble \(\mathcal{E} = \{ |\psi_i\rangle, p_i \}_{i}\) that generates \(\rho^A\), is it possible for Bob to perform a measurement on his system that breaks the entanglement and leaves Alice’s system in state \(|\psi_i\rangle\) with probability \(p_i\)? Phrased differently, if Alice and Bob share many identical copies of \(|\Psi\rangle^{AB}\), is it possible for
Bob to perform the same measurement on each copy such that with very high probability roughly a fraction $p_i$ of Alice’s systems are in state $|\psi_i\rangle$, for all $i$?

To see how this might work, consider a Schmidt decomposition $|\Psi\rangle^{AB} = \sum_{k=1}^{r} \sigma_k |\alpha_k\rangle^A |\beta_k\rangle^B$ where $\{|\beta_k\rangle\}_{k=1}^{d_B}$ forms an orthonormal basis for $\mathcal{H}^B$. The operator $O = \sum_{k=1}^{d_B} |\beta_k\rangle \langle \beta_k|$ is hermitian and therefore is a quantum observable for Bob’s system. When he performs this measurement and obtains outcome $k$, the joint post-measurement state is given by

$$\frac{1}{p_k} (I \otimes |k\rangle \langle k|) |\Psi\rangle^{AB} = |\alpha_k\rangle^A |\beta_k\rangle^B,$$

where $p_k = \sigma_k^2$. Now once Bob learns the outcome of his measurement, he knows precisely that Alice’s system is in state $|\alpha_k\rangle$. However, if Alice only learns that Bob performed the measurement but she does not learn the outcome, then her description of the system is statistical and given by the ensemble $E = \{ |\alpha_k\rangle, p_k \}_{k=1}^{r}$. Note that $E$ has an ensemble average given by the reduced density matrix $\rho^A = \text{Tr}_B |\Psi\rangle \langle \Psi|^{AB}$, and we say that through this process Bob can prepare the **ensemble** $E$ for Alice. The question is whether Bob can also prepare any other ensemble for Alice that generates $\rho^A$. We now show how any such preparation can always be accomplished.

Let $E’ = \{ |\Psi_i\rangle, p_i’ \}_{i=1}^{r’}$ be any other ensemble whose ensemble average is also $\rho^A$. The state $|\Psi’\rangle^{AB} = \sum_{i=1}^{r’} \sqrt{p_i’} |\Psi_i\rangle^A |\beta_i\rangle^B$ is a purification of $\rho^A$ where $B’$ is an $r’$-dimensional system, and since the rank of $\rho^A$ is $r$, we must have $r \leq r’$. In the discussion leading up to Lemma 2, we saw how mathematically $\mathcal{H}^B$ could be regarded as a subspace of $\mathcal{H}^{B’}$. However, in physical implementations, system $B$ will be of fixed dimension and so it cannot be transformed to obtain any general state in the larger space $\mathcal{H}^{B’}$. What Bob must do is introduce a second system, called an **ancilla system**, which effectively increases the dimension of his system. Let $\tilde{B}$ be any system of dimension $d’$ such that $rd \geq r’$. Then the dimension of the joint system $\mathcal{H}^{BB\tilde{B}}$ (which is at least $rd$) will be no less than the dimension of $\mathcal{H}^{B’}$, and so now $\mathcal{H}^{B’}$ is a subspace of $\mathcal{H}^{BB\tilde{B}}$. Since Bob has full control of both systems $B$ and $\tilde{B}$, any state in $\mathcal{H}^{B’}$ is now accessible to him. In other words, $|\Psi’\rangle^{AB}$ is also realizable as a mathematically equivalent state $|\Psi\rangle^{AB\tilde{B}}$ on systems $AB\tilde{B}$. Therefore, to prepare ensemble $E’$ for Alice, Bob first introduces ancilla system $\tilde{B}$ in some initial state $|0\rangle^{\tilde{B}}$ so that the joint state on $\mathcal{H}^{AB\tilde{B}}$ is $|\Psi\rangle^{AB\tilde{B}}$. Obviously $\text{Tr}_{B\tilde{B}}(|\Psi\rangle \langle \Psi|^{AB\tilde{B}} \otimes |0\rangle \langle 0|^{\tilde{B}})$ is a purification of $\rho^A$, and so by Lemma 2 there exists a unitary $U^{BB\tilde{B}}$ acting on $\mathcal{H}^{BB\tilde{B}}$ such that

$$I \otimes U^{BB\tilde{B}} (|\Psi\rangle^{AB\tilde{B}} |0\rangle^{\tilde{B}}) = |\Psi’\rangle^{AB\tilde{B}} = \sum_{i=1}^{r’} \sqrt{p_i’} |\Psi_i\rangle^A |\beta_i\rangle^{B\tilde{B}},$$

where the $|\beta_i\rangle^{B\tilde{B}}$ are computational basis vectors for system $B\tilde{B}$. Bob performs this unitary and then measures system $B\tilde{B}$ using observable $O = \sum_k |k\rangle \langle k|^{BB\tilde{B}}$. Alice’s post-measurement description of her system is given by the ensemble $E’ = \{ |\Psi_i\rangle, p_i’ \}_{i=1}^{r’}$, as desired. We have thus proven the following theorem.

**Theorem 1.** Suppose that Alice and Bob share pure state $|\Psi\rangle^{AB}$ on system $\mathcal{H}^{AB}$. If $E = \{ |\psi_i\rangle, p_i \}_{i=1}^{r}$ is any ensemble with ensemble average $\text{Tr}_B |\Psi\rangle \langle \Psi|^{B} = \sum_i p_i |\psi_i\rangle \langle \psi_i|$, then Bob can prepare this ensemble for Alice by interacting his system with an ancilla system and performing a measurement on both.
3 Generalized Measurements

3.1 Physically Implementable Processes

In Thm. 1 Bob is able to prepare for Alice any ensemble that generates her density matrix by a process that involves introducing an extra ancilla system. This idea turns out to be very powerful, and it allows an experimenter to realize certain system transformations that are more general than what is described by the Unitary Evolution and Measurement Axioms. For a quantum system $S$ with state space $\mathcal{H}^S$, recall that $L(\mathcal{H}^S)$ consists of all linear operator acting on $S$, and every density operator for $S$ belongs to $L(\mathcal{H}^S)$. The goal here is to identify all the physically implementable maps $E : L(\mathcal{H}^S) \rightarrow L(\mathcal{H}^S)$ that can be performed on density operators of system $S$ by combining the Unitary Evolution and Measurement Axioms with an ancilla system $\tilde{S}$. Linear maps of this form are often called superoperators since they are linear maps acting on linear operators themselves. If the system is originally in state $\rho$, then the map $E$ will transform it into $E(\rho)$, perhaps with just some probability $p$. In principle we could even consider superoperators of the form $E : L(\mathcal{H}^S) \rightarrow L(\mathcal{H}^S)$, where the output could be a density operator acting on an entirely different system $S'$, and in many cases we will interested in working with such channels. For simplicity we will not consider such maps in the following discussion, but they can be analogously characterized using the framework described below.

For a process to be physically implementable means that there exists a physical procedure for generating the state transformations described by the process that is consistent with the axioms of quantum mechanics. According to the Unitary Evolution and Measurement Axioms, the most general procedure reduces to the following steps: (i) introducing an ancilla system $\tilde{S}$ in some pure state $|0\rangle\langle0|^{\tilde{S}}$, (ii) applying a joint unitary $U^{SS\tilde{S}}$ to systems $S\tilde{S}$, and (iii) then performing an observable $O^{S\tilde{S}}$ with eigenspace projectors $\{P_k^{S}\}_k$ on systems $\tilde{S}$. Focusing just on system $S$, these three steps generate for each outcome $k$ the map $E_k : L(H^S) \rightarrow L(H^S)$ given by

\[
E_k(\rho^S) = \text{Tr}_\tilde{S}[(I^S \otimes P_k^{\tilde{S}})U^{SS\tilde{S}}(\rho^S \otimes |0\rangle\langle0|^{\tilde{S}})U^{1SS\tilde{S}}(I^S \otimes P_k^{\tilde{S}})]
\]

\[
= \text{Tr}_\tilde{S}[(I^S \otimes P_k^{\tilde{S}})U^{SS\tilde{S}}(\rho^S \otimes |0\rangle\langle0|^{\tilde{S}})U^{1SS\tilde{S}}],
\]

and $p_k = \text{Tr}[P_k^{S}U^{SS\tilde{S}}(\rho^S \otimes |0\rangle\langle0|^{\tilde{S}})U^{1SS\tilde{S}}]$. The second equality in Eq. (24) comes from the cyclicity of the trace and the fact that $P_k^{S}$ is a projector.

In this three-step process, it may be questioned why we assume in step (i) that system $\tilde{S}$ is initially in a pure state $|0\rangle\langle0|^{\tilde{S}}$, and also why we assume that the measurement in step (iii) acts only on system $\tilde{S}$ rather than both systems $S$ and $\tilde{S}$. Concerning the purity assumption in step (i), if the initial state in $\tilde{S}$ were mixed, say $\rho^{\tilde{S}}$, then we could consider introduce a second ancilla system $\tilde{S}'$ such that $|0\rangle^{SS\tilde{S}'}$ is a purification of $\rho^{\tilde{S}}$. Then we have

\[
\text{Tr}_{SS'}[(I^S \otimes P_k^{S} \otimes I^{\tilde{S}'}) (U^{SS\tilde{S}'\tilde{S}'} \otimes I^{\tilde{S}'}) (\rho^S \otimes |0\rangle\langle0|^S) (U^{SS\tilde{S}'\tilde{S}'} \otimes I^{\tilde{S}'})] = \text{Tr}_{\tilde{S}'}[(I^S \otimes P_k^{S} \otimes I^{\tilde{S}'})]U^{SS\tilde{S}'\tilde{S}'}(\rho^S \otimes \rho^{\tilde{S}}) U^{1SS\tilde{S}'}.
\]

Thus, by just considering a larger ancilla system $SS'$, there is no loss of generality in assuming that the initial ancilla state is pure. A similar argument shows that there is also no loss of generality in assuming that the observable in step (iii) acts only on system $\tilde{S}$. If $\{P_k^{S\tilde{S}}\}_k$ is a general collection of eigenspace projectors for observable $O^{SS\tilde{S}} = \sum_k kP_k^{SS\tilde{S}}$ on both $S$ and $\tilde{S}$, then the action of $O^{SS\tilde{S}}$ can be equivalently implemented by introducing a second ancilla system $\tilde{S}'$ in state $|0\rangle\langle0|^{SS\tilde{S}'}$, applying the isometry $V^{SS\tilde{S}'\tilde{S}} = \sum_k P_k^{SS\tilde{S}} \otimes |k\rangle\langle0|^\tilde{S}'$, and then measuring system $\tilde{S}'$ using some observable that is
diagonal in the computational basis $\{|k\rangle\rangle_k$. Since this isometry can be extended to a unitary acting on systems $SSS'$ (see Lemma 2 of the previous lecture), the entire process is again given by the three steps above, except now with a larger ancilla system $SS'$.

Let us now analyze further the maps $E_k$ given in Eq. (24). The first property to notice is that $E_k$ is a linear map since the partial trace is a linear operation. In other words, $E_k(\rho + \sigma) = aE_k(\rho) + bE_k(\sigma)$ for any $a, b \in \mathbb{C}$ and $\rho, \sigma \in L(H^S)$. Additional structure to the maps $E_k$ can be seen by writing $U^{SS}$ in the eigenbasis of the observable $O^S$. Let $P^S_k = \sum \langle e_{k,l} \rangle \langle e_{k,l} \rangle$ be a spectral decomposition of the $k^{th}$ eigenspace projector of $O^S$ so that $\{|e_{k,l}\rangle\rangle_k,l\}$ provides an orthonormal basis for $H^S$. Then we can decompose $U^{SS}$ as

$$U^{SS} = \sum_{k,l} M^S_{k,l} \otimes |e_{k,l}\rangle\langle 0| + \hat{U},$$  

where $M^S_{k,l}$ are operators acting on $H^S$ and is some operator acting on $H^{SS}$ such that $\hat{U}(\{|\psi\rangle\rangle_0\}) = 0$ for all $|\psi\rangle \in H^S$. Every operator can always be written in this way, but the condition of $U^{SS}$ being unitary further implies that $\sum_{k,l} M^S_{k,l} M^S_{k,l} = I^S$, which is called the completion condition.

Substituting Eq. (25) into Eq. (24) gives the simplified form

$$E_k(\rho^S) = \sum_{l} M_{k,l} \rho^S M^l_{k,l}.$$  

(26)

Let us pause for a moment and reiterate what Eq. (26) physically represents. With system $S$ originally in state $\rho^S$, an ancilla system $S$ and interacted with $S$ via the unitary $U^{SS}$. The ancilla system was then measured, and value $k$ was the measurement outcome. Eq. (26) then describes the post-measurement state of system $S$ given that $k$ was measured on $S$. Actually, the latter statement is not entirely correct since $E_k(\rho^S)$ will not be normalized in general. If $p_k$ is the probability of outcome $k$ when measuring $S$, then the normalized post-measurement state on system $S$ is

$$\frac{1}{p_k} E_k(\rho^S) = \frac{1}{p_k} \sum_{l} M_{k,l} \rho^S M^l_{k,l},$$  

(27)

where

$$p_k = \sum_{l} \text{Tr}[M_{k,l} \rho^S M^l_{k,l}] \leq \sum_{k,l} \text{Tr}[M_{k,l} \rho^S M^l_{k,l}] = \sum_{k,l} \text{Tr}[M^S_{k,l} M_{k,l} \rho^S] = \text{Tr}[\rho^S] = 1.$$  

(28)

### 3.2 Completely Positive Maps and Quantum Instruments

From Eq. (26) we see that every $E_k$ has the property of being a positive map. That is, if for any input positive operator $X$, the output $E_k(X)$ is also positive, a fact that be easily verified by considering

$$\langle \psi | E_k(X) | \psi \rangle = \sum_{l} \langle \psi | M_{k,l} X M^l_{k,l} | \psi \rangle$$

for any state $|\psi\rangle$. However, $E_k$ has an even stronger property known as complete positivity. A superoperator $T : L(H^S) \rightarrow L(H^S)$ is said to be completely positive (CP) if, for any auxiliary system $E$, the operator $T \otimes \mathcal{I}^E : L(H^{SE}) \rightarrow L(H^{SE})$ is a positive operator, where $\mathcal{I}^E$ is the identity superoperator acting on $L(H^E)$. These definitions also apply to superoperators with a different
output space, \( T : L(\mathcal{H}^S) \rightarrow L(\mathcal{H}^E) \). To see that \( \mathcal{E}_k \) is CP, consider again Eq. (26) and let \( Y \in L(\mathcal{H}^{SE}) \) be an arbitrary operator. Then
\[
\text{Tr} \left[ \sum_k \mathcal{E}_k(X) \right] = \sum_k \text{Tr}[M_{k,l}X] = \sum_k \text{Tr}[M_{k,l}^0] \text{Tr}[X] = \text{Tr}[X],
\]
for any \( X \in L(\mathcal{H}^S) \). In other words, the superoperator \( \sum_k \mathcal{E}_k \) preserves the trace of \( X \). In general, any operation having this property is called \textit{trace-preserving}. Assume that \( S \) is a \( d \)-dimensional system. For each \( \mathcal{E}_k \) we associate the operator \( \Omega_{\mathcal{E}_k} \in \text{L}(\mathcal{H}^S \otimes \mathcal{H}^S) \) defined by
\[
\Omega_{\mathcal{E}_k} := \mathcal{E}_k \otimes I(\phi_d^+ \langle \phi_d^+ |),
\]
where \( |\phi_d^+\rangle = \sum_{i=1}^d |ii\rangle \) is an normalized vector \( \mathcal{H}^S \otimes \mathcal{H}^S \) introduced in the previous lecture. The operator \( \Omega_{\mathcal{E}_k} \) is called the \textit{Choi matrix} of the map \( \mathcal{E}_k \). For notation simplicity, let us label the first system represented in this tensor product space as \( S_1 \) and the second as \( S_2 \). Let \( \sigma \) be any operator in \( \text{L}(\mathcal{H}^S) \) whose transpose in the computational basis is denoted by \( \sigma^T \). Since \( I \) is the identity acting on \( S_2 \) in Eq. (31), then it commutes with a map that multiplies \( I \otimes \sigma^T \) from the left. Applying such a map to both sides of Eq. (31) and tracing out subsystem \( S_2 \) yields
\[
\text{Tr}_{S_2}[I \otimes \sigma^T \Omega_{\mathcal{E}_k}] = \text{Tr}_{S_2}[\mathcal{E}_k \otimes I((I \otimes \sigma^T)|\phi_d^+ \langle \phi_d^+ |)] = \text{Tr}_{S_2}[\mathcal{E}_k \otimes I((I \otimes \sqrt{\sigma^T})|\phi_d^+ \langle \phi_d^+ |)] = \text{Tr}_{S_2}[\mathcal{E}_k \otimes I((\sqrt{\sigma} \otimes \mathbb{I})|\phi_d^+ \langle \phi_d^+ |)] = \mathcal{E}_k(\sigma).
\]

**Proposition 3.** Every physically implementable process is described by a quantum instrument.

Our next goal is to show that the converse of this proposition is also true. Suppose that \( \{\mathcal{E}_k\}_k \) is a collection of CP maps acting on \( L(\mathcal{H}^S) \) with \( \sum_k \mathcal{E}_k \) being trace-preserving. Assume that \( S \) is a \( d \)-dimensional system. For each \( \mathcal{E}_k \) we associate the operator \( \Omega_{\mathcal{E}_k} \in \text{L}(\mathcal{H}^S \otimes \mathcal{H}^S) \) defined by

\[
\text{Tr}_{S_2}[\mathcal{E}_k \otimes I((I \otimes \sigma^T)|\phi_d^+ \langle \phi_d^+ |)] = \text{Tr}_{S_2}[\mathcal{E}_k \otimes I((I \otimes \sqrt{\sigma^T})|\phi_d^+ \langle \phi_d^+ |)] = \text{Tr}_{S_2}[\mathcal{E}_k \otimes I((\sqrt{\sigma} \otimes \mathbb{I})|\phi_d^+ \langle \phi_d^+ |)] = \mathcal{E}_k(\sigma).
\]

14
In this derivation, the second line follows form writing $\sigma^T = \sqrt{\sigma^T} \sqrt{\sigma^T}$ and using cyclicity of the trace, the third line follows from the ricochet property and the fact that $\sqrt{\sigma^T} = \sigma$ (a fact that can verified from the spectral decomposition of $\sigma^T$), and finally the fourth line follows from the fact that $\text{Tr}_{S_2}[\phi_d^{++} \phi_d^{+}] = I_{S_1}$. Hence the action of $\mathcal{E}_k$ on any density operator $\sigma$ can be determined equivalently computed using its Choi matrix $\Omega_{\mathcal{E}_k}$.

Since $\mathcal{E}_k$ is CP and $\phi_d^{++}$ is a positive operator, then $\Omega_{\mathcal{E}_k}$ is also positive. Thus, it has an orthonormal eigenbasis $\{|\psi_{k,l}\rangle\}_{l=1}^{d^2}$ with spectral decomposition

$$\Omega_{\mathcal{E}_k} = \sum_{l=1}^{d^2} \lambda_{k,l} |\psi_{k,l}\rangle \langle \psi_{k,l}| = \sum_{l=1}^{d^2} \lambda_{k,l} (M_{k,l} \otimes I)|\psi_{k,l}\rangle \langle \psi_{k,l}| (M_{k,l}^\dagger \otimes I).$$

(33)

Now if we multiply both sides by $I \otimes \sigma^T$ and trace out the second system, we have

$$\text{Tr}_{S_2}[I \otimes \sigma^T \Omega_{\mathcal{E}_k}] = \sum_{l=1}^{d^2} \lambda_{k,l} (M_{k,l} \otimes I)|\sigma M_{k,l}^\dagger\rangle \langle M_{k,l}^\dagger | = \sum_{l=1}^{d^2} M_{k,l} \sigma M_{k,l}^\dagger,$$

(34)

where $M_{k,l} := \sqrt{\lambda_{k,l}} M_{k,l}$. Putting together Eqns. (32) and (34) yields the equality

$$\mathcal{E}_k(\sigma) = \sum_{l=1}^{d^2} M_{k,l} \sigma M_{k,l}^\dagger.$$  

(35)

The operators $\{M_{k,l}\}_{l=1}^{d^2}$ are called Kraus operators for the map $\mathcal{E}_k$, and Eq. (35) is called a Kraus operator representation of $\mathcal{E}_k$. Furthermore, from the assumption that $\sum_k \mathcal{E}_k$ is a trace-preserving map, we must have that

$$\text{Tr} \left( \sum_{k=1}^{d^2} M_{k,l}^\dagger M_{k,l} \right) X = \text{Tr}[X]$$

(36)

for every operator $X \in L(H)$. It is not difficult to show that the only this equation can only be true for every $X$ if $\sum_k \sum_{l=1}^{d^2} M_{k,l}^\dagger M_{k,l} = I_S$; i.e. if the operators $\{M_{k,l}\}_{l=1}^{d^2} S$ satisfy the completion condition. This can most easily be seen by considering the spectral decomposition of $\sum_k \sum_{l=1}^{d^2} M_{k,l}^\dagger M_{k,l}$ and then taking $X$ to be the various eigenspace projectors.

Eq. (35) has exactly the same form that we observed for physically implementable CP maps in Eq. (26). Here we have shown that any general CP map also has the same structure. In fact, any set of Kraus operators $\{M_{k,l}\}_{k,l}$ satisfying the completion relation can be converted into a physically implementable map; one just used the operators $M_{k,l}$ to define a unitary $U^{SS}$ as in Eq. (25). This establishes the converse to Prop. 3.

**Proposition 4.** Every quantum instrument represents a physically implementable process.

Propositions 3 and 4 can be united in the following theorem, which describes the notion of generalized measurement in quantum information science.
Theorem 2 (The Generalized Measurement Theorem). A generalized measurement on system $S$ consists of a quantum instrument $\{E_k\}_k$. When performing the instrument on pre-measurement state $\rho^S$, outcome $k$ is obtained with probability

$$p_k = \text{Tr}[E_k(\rho^S)].$$

The post-measurement state is given by $\frac{1}{p_k}E_k(\rho^S)$. Thus a generalized quantum measurement involves a stochastic transformation on system $S$ given by:

$$\rho^S \mapsto \frac{1}{p_k}E_k(\rho^S)$$

with probability $p_k$.

Again, in the actual implementation of a generalized measurement, the outcome $k$ described in this theorem is the measurement outcome obtained from some observable on an ancilla system $\tilde{S}$.

3.3 Special Types of Generalized Measurements

We highlight four special families of quantum instruments that are used quite regularly in the subject.

3.3.1 Quantum Channels

A quantum channel is any completely positive trace-preserving (CPTP) map acting on density matrices. Thus a quantum channel is a quantum instrument consisting of a single CP map, $\mathcal{E}$. However, for any instrument $\{E_k\}_k$, the sum $\sum_k E_k$ is always a quantum channel since it is a CPTP map. This channel is very important since it represents how the density operator description of a quantum system evolves when a generalized measurement is performed, but the outcome of this measurement is not known. For a system initially described by density operator $\rho^S$, the generalized measurement process induces a stochastic transformation

$$\rho^S \mapsto \frac{1}{p_k}E_k(\rho^S),$$

occurring with probability $p_k$. Analogous to the post-measurement ensemble generated by some quantum observable, we can interpret $\{\frac{1}{p_k}E_k(\rho^S), p_k\}_k$ as being the post-measurement ensemble of mixed states and

$$\hat{\rho}^S = \sum_k p_k \left( \frac{1}{p_k}E_k(\rho^S) \right) = \sum_k E_k(\rho^S)$$

as being the post-measurement ensemble average. The transformation $\rho^S \to \hat{\rho}^S$ is equivalently described as system $S$ being sent through the quantum channel $\mathcal{E} := \sum_k E_k$.

3.3.2 Measurement Instruments and POVMs

A special type of instrument involves CP maps $\mathcal{T} : L(\mathcal{H}^S) \to \mathbb{C}$ of the form

$$\mathcal{T}(X) = \text{Tr}[Xi],$$

for some fixed positive operator $\Pi \in L(\mathcal{H}^S)$. To see that this is CP, suppose that $Y \in \mathcal{H}^{SE}$ is a positive operator. Then

$$\mathcal{T} \otimes \mathcal{I}^E(Y) = \text{Tr}_S[Y(\Pi \otimes I)] = \text{Tr}_S[(\Pi^{1/2} \otimes I)Y(\Pi^{1/2} \otimes I)],$$

(39)
which is positive since it is the partial trace of a positive operator \((\Pi^{1/2} \otimes \mathbb{I})Y(\Pi^{1/2} \otimes \mathbb{I})\). We are interested in a family of CP maps \(\{\mathcal{E}_x\}_{x \in \mathcal{X}}\) with \(\mathcal{X}\) being some finite set and with each \(\mathcal{E}_x\) having the form of Eq. (38) and such that \(\sum_{x \in \mathcal{X}} \mathcal{E}_x\) is trace-preserving. Here we are explicitly denoting the set \(\mathcal{X}\) to which the different outcomes \(x\) belong since it plays a specific role in the definition of these instruments. If we let \(\Pi_x\) denote the positive operator defining each of the CP maps \(\mathcal{E}_x\), then the trace-preserving condition becomes

\[
\operatorname{Tr}[\rho] = \operatorname{Tr}\left[\sum_{x \in \mathcal{X}} \mathcal{E}_x(\rho)\right] = \operatorname{Tr}\left[\sum_{x \in \mathcal{X}} \Pi_x \rho\right].
\]

Once again, this equality can hold for all \(X\) if and only if \(\mathbb{I}^S = \sum_{x \in \mathcal{X}} \Pi_x\). Thus, we have a valid quantum instrument \(\{\mathcal{E}_x\}_{x \in \mathcal{X}}\) with

\[
\mathcal{E}_x(\rho) = \operatorname{Tr}[\rho \Pi_x]
\]

provided that \(\sum_{x \in \mathcal{X}} \Pi_x = \mathbb{I}^S\).

Any set of positive operators \(\{\Pi_x\}_{x \in \mathcal{X}}\) that sum up to the identity is called a **positive-operator valued measure** (POVM). It is called this because for every density matrix \(\rho \in \mathcal{L}(\mathcal{H}^S)\), the function \(p : \mathcal{X} \to \mathbb{R}\) given by

\[
p(x) = \operatorname{Tr}[\rho \Pi_x]
\]

defines a probability measure on \(\mathcal{X}\). POVMs are used in quantum information to characterize scenarios in which only the classical data of measurement outcomes are of interest.

For example, suppose that Alice wishes to send Bob a yes/no message using a quantum system. If her message is yes, she prepares some state \(\rho_0\) and sends it to Bob, while if her message is no, she prepares a different state \(\rho_1\) and sends it to him. On Bob’s end, he must measure the incoming quantum system and determine if it is in state \(\rho_0\) or state \(\rho_1\). Based on this result, he can decode Alice’s message as either being yes or no. In this case, Bob describes his quantum instrument by a POVM \(\{\Pi_0, \Pi_1\}\). When Bob obtains outcome \(k \in \{0, 1\}\), he guesses that Alice sent state \(\rho_k\). Thus the probability that he correctly decodes Alice’s message when she says “yes” is \(\operatorname{Tr}[\rho_0 \Pi_0]\), and the probability that he correctly decodes her message when she says “no” is \(\operatorname{Tr}[\rho_1 \Pi_1]\).

### 3.3.3 Fine-grained Instruments

An instrument \(\{\mathcal{E}_k\}_k\) is called **fine-grained** if each \(\mathcal{E}_k\) has a single Kraus operator; i.e.

\[
\mathcal{E}_k(\rho) = M_k \rho M_k^\dagger.
\]

In this case, the entire instrument is more easily characterized as simply being a set of Kraus operators \(\{M_k\}_k\) such that \(\sum_k M_k^\dagger M_k = \mathbb{I}\). When performing such an instrument on initial state \(\rho\), outcome \(k\) is obtained with probability \(p_k = \operatorname{Tr}[M_k^\dagger M_k \rho]\), and when it does, the post-measurement state is \(\frac{1}{p_k} M_k \rho M_k^\dagger\).

### 3.3.4 Projective Measurements

The final special type of instruments we consider are called **projective measurements**. This is a fine-grained instrument in which the Kraus operators form a set of orthogonal projectors. That is, the instrument is defined by a set of projectors \(\{P_k\}_k\) such that \(P_k P_j = \delta_{jk} P_k\) and \(\sum_k P_k = \mathbb{I}\). We immediately see that every normal operator, and hence every observable, generates a projective measurement by considering its set of eigenspace projectors. Thus quantum observables represent just one type of quantum instrument, and the Generalized Measurement Theorem actually includes the Measurement Axiom as a special subcase.
4 Exercises

Exercise 1

For a three dimensional quantum system, consider the two observables

\[ A = a_1 (1 - |\varphi\rangle\langle\varphi|) + a_2 |\varphi\rangle\langle\varphi|; \]
\[ B = b_1 |1\rangle\langle1| + b_2 |2\rangle\langle2| + b_3 |3\rangle\langle3|, \]

where \( |\varphi\rangle = \sqrt{1/3} (|1\rangle + |2\rangle + |3\rangle) \). Let \( |\psi\rangle = x|1\rangle + y|2\rangle + z|3\rangle \) be the initial state of the system where \( x, y, \) and \( z \) are arbitrary constants. An experimenter first performs observable \( A \) on her system. If she obtains outcome \( a_2 \), she stops. On the other hand, if outcome \( a_1 \) is obtained, she performs observable \( B \). What are the possible post-measurement states of this entire process, and what are their respective probabilities?

Exercise 2

Prove that a density operator \( \rho \) is pure iff \( \text{Tr}[\rho^2] = 1 \).

Exercise 3

Determine the values of \( a, b, c, d, e \) for which the state

\[ |\Psi\rangle = ad|00\rangle + ae|02\rangle + bd|10\rangle + be|12\rangle + cd|20\rangle + cd|22\rangle \]

is a product state.

Exercise 4

Prove the following statements.

(a) For any two product states \( |\psi_1\rangle = |\alpha_1\rangle|\beta_1\rangle \) and \( |\psi_2\rangle = |\alpha_2\rangle|\beta_2\rangle \) in \( \mathbb{C}^2 \otimes \mathbb{C}^2 \), let \( S \) be the subspace spanned by \( |\psi_1\rangle \) and \( |\psi_2\rangle \). Then either all states in \( S \) are product states, or \( |\psi_1\rangle \) and \( |\psi_2\rangle \) are the only product states in \( S \).

(b) Any two-dimensional subspace of \( \mathbb{C}^2 \otimes \mathbb{C}^2 \) contains at least one product state.

Exercise 5

Consider a two-qubit system initially in the state \( |\Psi\rangle^{AB} = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle \). Compute the post-measurement ensemble prepared for Alice when Bob performs a measurement on system \( B \) given by the observable \( \sigma_y \).

Exercise 6

Prove that any purification of pure state must be a pure product state.

Exercise 7

Consider the density operator \( \rho = \frac{1}{2} (|0\rangle\langle0| + |+\rangle\langle+|) \).

1. Compute two distinct purifications of \( \rho \).

2. Give two pure-state ensembles whose ensemble average is \( \rho \).
Exercise 8

For any integer \( N \), let \( U = e^{-i\pi \sigma_y / N} \) be a unitary operator defined on \( \mathbb{C}^2 \).

(a) Write out the matrix representation of \( U^k \) for any \( k = 0, \ldots, N - 1 \).

(b) Define the operators \( M_k = \gamma U^k|0\rangle\langle 0|U^k \) for \( k = 0, \ldots, N \) and some constant \( \gamma \). Show that for a suitable choice of \( \gamma \), the operators \( \{ M_k \}_{k=0}^{N-1} \) represent a generalized measurement on \( \mathbb{C}^2 \). What is the value of \( \gamma \)?

(Hint. It may be helpful to recall the relationship
\[
\sum_{k=0}^{N-1} e^{i2\pi k/N} = \sum_{k=0}^{N-1} \left( e^{i2\pi/N} \right)^k = e^{2\pi i} - 1 \quad \frac{1}{e^{2\pi i/N} - 1} = 0,
\]
which is a geometric series expansion for \( e^{i2\pi/N} \).)

Exercise 9

Suppose that Alice and Bob share the two-qubit entangled state \( |\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \). Consider the fine-grained generalized measurement given by Kraus operators \( \{ M_0, M_1 \} \) where
\[
M_0 = \begin{pmatrix} \sqrt{1-\epsilon} & 0 \\ 0 & 1 \end{pmatrix}, \quad M_1 = \begin{pmatrix} \sqrt{\epsilon} & 0 \\ 0 & 0 \end{pmatrix}.
\]
Suppose that Alice and Bob each perform this generalized measurement on their respective subsystem. What are the four possible post-measurement states, and with what probability does each occur?

Exercise 10

The fully dephasing map \( \Delta \) is a quantum channel that acts on a \( X \in \mathbb{L}(\mathbb{C}^d) \) according to
\[
\Delta(X) = \sum_{i=1}^{d} |i\rangle\langle i|X|i\rangle\langle i|.
\]

(a) Restrict attention to \( \mathbb{C}^3 \), and consider the superoperator
\[
\mathcal{T}(X) = 2 \left( \Delta(X) + \Delta(\pi X \pi^\dagger) \right) - X \quad \text{for } X \in \mathbb{L}(\mathbb{C}^3),
\]
where \( \pi = |1\rangle\langle 2| + |2\rangle\langle 3| + |3\rangle\langle 1| \) is a permutation. Is \( \mathcal{T} \) completely positive?

(b) Is \( \mathcal{T} \) given in Eq. (45) a positive map?

(Hint. For (a), consider \( \langle \phi_3^+ | \Lambda(\phi_3^+) \otimes I | \phi_3^+ \rangle \). For part (b), note that \( \Lambda(X) \) will always be hermitian when \( X \geq 0 \). Then use Sylvester’s criterion to check for positivity of \( \Lambda \).)

(Bibliography note. The map in Eq. (45) was first presented by Choi in Ref. [Cho75].)

References