

3. Density Matrix Formalism

“You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.”

– von Neumann to Shannon.

3.1 Density Operator

3.1.1 Pure states

A quantum state that is written as a state vector $|\psi\rangle$ is what we will call a *pure* state. We can also define a density operator¹ $\hat{\rho}$ corresponding to such a state by

$$\hat{\rho} \equiv |\psi\rangle\langle\psi|, \quad (3.1)$$

which is simply the projection operator onto the state $|\psi\rangle$. This also leads to

$$\hat{\rho}^2 = (|\psi\rangle\langle\psi|)(|\psi\rangle\langle\psi|) = \hat{\rho}. \quad (3.2)$$

The density matrix is the *matrix representation* of the density operator, in a particular basis. These two terms are used interchangeably in the literature, although they contain slightly different meaning. If the quantum state is a superposition $|\psi\rangle = \sum_i c_i |\phi_i\rangle$, then

$$\hat{\rho} \equiv |\psi\rangle\langle\psi| = \sum_{i,j} c_i c_j^* |\phi_i\rangle\langle\phi_j|. \quad (3.3)$$

The terms for which $i = j$ correspond to $c_i c_i^* = |c_i|^2$, the probability of finding state $|\phi_i\rangle$ in $|\psi\rangle$. These would be diagonal entries on the density matrix written in the basis $\{|\phi_i\rangle\}$. The terms for which $i \neq j$ are called *interference* or *coherence* terms, which are *off-diagonal* entries on the density matrix. Interference is typically understood to be associated with the “quantumness” of a state arising from being a superposition of states, e.g. $|\psi\rangle = \sum_i c_i |\phi_i\rangle$. However, one must be

¹The density operator was introduced by von Neumann and by Lev Landau independently in 1927, but with different motivations. The motivation that inspired Landau was the impossibility of describing a subsystem of a composite quantum system by a state vector. On the other hand, von Neumann introduced the density matrix in order to develop both quantum statistical mechanics and a theory of quantum measurements.

careful to understand that a matrix is written in terms of a particular basis, and there will be a basis in which the density matrix is diagonal. Thus, the association between off-diagonal entries and “quantumness” should not lead us to the wrong conclusion that the absence of off-diagonal terms of a density matrix means that the system is classical, or has lost coherence. Expressed in a different basis, off-diagonal terms may re-appear.

3.1.2 Trace

We now introduce the *trace* operation, which is denoted by “Tr”. It is an operation on some operator, say \hat{A} . Given some orthonormal basis $\{|\phi_i\rangle\}$ of the Hilbert space of the system,

Trace

$$\text{Tr}(\hat{A}) \equiv \sum_i \langle \phi_i | \hat{A} | \phi_i \rangle. \quad (3.4)$$

The trace operation is linear,

$$\text{Tr}(\hat{A} + \hat{B}) = \text{Tr}(\hat{A}) + \text{Tr}(\hat{B}). \quad (3.5)$$

It is also independent of the particular choice of orthonormal basis $\{|\phi_i\rangle\}$. Thus we can use any arbitrary set of orthonormal basis vectors to compute the trace.

The cyclic property of trace is also useful,

$$\text{Tr}(\hat{A}\hat{B}\hat{C}) = \text{Tr}(\hat{B}\hat{C}\hat{A}) = \text{Tr}(\hat{C}\hat{A}\hat{B}). \quad (3.6)$$

Note that $\text{Tr}(\hat{A}\hat{B}\hat{C}) \neq \text{Tr}(\hat{C}\hat{B}\hat{A})$.

The reason for introducing the trace operation is as follows. Consider the trace operation on the operator $\hat{B} = \hat{\rho}\hat{A}$, where \hat{A} is Hermitian. Taking the trace over the eigenvectors $\{|a_i\rangle\}$ of \hat{A} with corresponding eigenvalues a_i , the trace yields

$$\begin{aligned} \text{Tr}(\hat{\rho}\hat{A}) &= \sum_i \langle a_i | \hat{\rho}\hat{A} | a_i \rangle = \sum_i \langle a_i | \psi \rangle \langle \psi | \hat{A} | a_i \rangle \\ &= \sum_i a_i \langle a_i | \psi \rangle \langle \psi | a_i \rangle = \sum_i a_i |\langle a_i | \psi \rangle|^2. \end{aligned} \quad (3.7)$$

The term $|\langle a_i | \psi \rangle|^2$ is the probability of the outcome a_i in a measurement represented by \hat{A} . Thus $\text{Tr}(\hat{\rho}\hat{A})$ represents the average over all possible outcomes a_i of this measurement, weighted by the corresponding Born probabilities, which is the definition of the expectation value $\langle \hat{A} \rangle$ of the observable \hat{A} . Therefore, we have the *trace rule* for the expectation value:

Expectation value

$$\langle \hat{A} \rangle = \text{Tr}(\hat{\rho}\hat{A}). \quad (3.8)$$

The diagonal entries of the density matrix represent probabilities, which must sum to 1. If we took $\hat{A} = \hat{I}$, we can also see that

$$\text{Tr}(\hat{\rho}) = 1. \quad (3.9)$$

This relation simply reflects the fact that pure states are normalized, i.e. $|\langle \psi | \psi \rangle|^2 = 1$. We emphasize that the concept and interpretation of the trace fundamentally relies on the Born rule, i.e. what $|\langle a_i | \psi \rangle|^2$ meant in Eq. (3.7). The importance of this point will become clear later.

Note that although we derived these results from the density operator for a pure state, we will generalize it later to mixed states and see that Eq. (3.8) and Eq. (3.9) are true in general.

3.1.3 Mixed States

We have been dealing with pure states up till now. For pure states, the state vector $|\psi\rangle$ and the corresponding pure state density matrix $\hat{\rho}$ are completely equivalent. However, pure states are not

the only type of states that an experimenter may encounter in the lab. For example, a two-level system may be prepared in either of the two pure states, $|0\rangle$ or $|1\rangle$, with some probability p_0 and p_1 , but the observer does not know which. These pure states need not even be orthogonal, or be limited to two states. As such, these states are called statistical mixtures of pure states, or just *mixed states*, which are descriptions that express insufficient information about the state of the system. The mixed state, i.e., the resulting set of possible pure states $|\psi_i\rangle$ with associated probabilities p_i , represents a *classical ensemble*, meaning the origin of the probabilities is purely classical. These probabilities simply express the subjective ignorance of the observer about the quantum state of the system, while physically the system has indeed been prepared in a pure state (albeit not known to the observer).

We can introduce the mixed state density matrix that encapsulates the classical probabilities p_i of the pure states $|\psi_i\rangle$ as

Density operator (general definition)

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle \langle \psi_i|. \quad (3.10)$$

Probabilities have the usual properties $p_i \geq 0$ and $\sum_i p_i = 1$. The pure state density matrix, Eq. (3.1), is a special case of Eq. (3.10) with one pure state. In this definition, there is no condition for the pure states in the mixture to be orthogonal, e.g. we can prepare a mixed state of 50% $|0\rangle$ and 50% $|+\rangle$.

The density operator is Hermitian,

$$\hat{\rho}^\dagger = \hat{\rho}, \quad (3.11)$$

and non-negative (or positive semi-definite), i.e.

$$\langle i | \hat{\rho} | i \rangle \geq 0 \quad \forall |i\rangle \text{ in the Hilbert space.} \quad (3.12)$$

In other words, all eigenvalues of $\hat{\rho}$ are non-negative, as would be expected since the eigenvalues of a density matrix are the probabilities defining the mixture of pure quantum states.

We emphasize that the trace rule in Eq. (3.9) and expectation value of an observable in Eq. (3.8) is still valid with the general definition of density operators in Eq. (3.10). These will be derived in the next few examples.

■ **Example 3.1** Show that $\text{Tr}(\hat{\rho}) = 1$ for the mixed state density matrix.

■ **Example 3.2** Show that $\langle \hat{A} \rangle = \text{Tr}(\hat{\rho} \hat{A})$ for the mixed state density matrix.

3.1.4 Basis ambiguity of mixed state density matrices

Finally, we note that given a mixed state density matrix, unless one knows that the system has been prepared in a particular pure state from a set of states $\{|\psi_i\rangle\}$, but does not know which, one cannot reconstruct *uniquely* a classical probability distribution of pure states. In other words, without more specific information (e.g. about state preparation), the density matrix alone provides only information about the probabilities of *different* sets of pure states, but not about which particular set of states has been prepared. The following example illustrates this point.

■ **Example 3.3** Given a mixed state density matrix $\hat{\rho} = \frac{1}{2} |0\rangle \langle 0| + \frac{1}{2} |1\rangle \langle 1|$, show that it can also be written as $\hat{\rho} = \frac{1}{2} |+\rangle \langle +| + \frac{1}{2} |-\rangle \langle -|$.

3.2 Purity and Quantum Entropy

The degree of “mixedness” (or conversely, purity) of a state can be found by two different quantities – purity and von Neumann entropy.

Purity

$$\xi \equiv \text{Tr}(\hat{\rho}^2). \quad (3.13)$$

A completely mixed or maximally mixed state is such that $p_i = 1/d$, $\forall i$, where d is the dimensionality of the Hilbert space upon which the state is defined. Why is such a state “maximally mixed”? One can think of the statistical mixture as reflecting our ignorance (or information) about the state given to us. A pure state density matrix, say $\hat{\rho} = |\psi_1\rangle\langle\psi_1|$ contains no “ignorance” about the state. In contrast, a mixture of 95% $|\psi_1\rangle$ and 5% some other state(s), reduces our information about the state and increases our ignorance. The degree of information loss or ignorance is not too large; we can still be confident that most of the time we are getting $|\psi_1\rangle$. However, when the mixture becomes an equal-weight ensemble of all possible states, where the density operator is equal to the identity matrix multiplied by a normalization factor, then the degree of information loss and our ignorance is maximum. In this case, there is no information about which pure state has been prepared, and hence corresponds to maximum ignorance and information loss.

■ **Example 3.4** Show that $\hat{\rho}^2 \neq \hat{\rho}$ for the mixed state density matrix, and deduce that $\text{Tr}(\hat{\rho}^2) \leq 1$.

PH3406 (2024) T.S.KOH

The von Neumann entropy² is a generalization of entropy in classical statistical mechanics to the case of quantum mechanical density operators. It quantifies the information content for a quantum system in a specified state, and is given by

von Neumann Entropy

$$S(\hat{\rho}) := -\text{Tr}(\hat{\rho} \log_2 \hat{\rho}) = -\sum_i \lambda_i \log_2 \lambda_i, \quad (3.14)$$

where λ_i are the eigenvalues of the density matrix $\hat{\rho}$ and $0 \log_2(0) := 0$.

By convention, the case $\lambda_i = 0$ is handled by defining $0 \log_2(0) := 0$, such that states absent from a mixture do not enter into the value of the entropy. Pure states have zero entropy; mixed states have entropy greater than zero. In quantum information the logarithms are usually taken to be base 2, giving a maximum entropy of 1 for a qubit in a maximally mixed state.

The logarithm of a density matrix may be calculated as follows: (1) write out the Taylor series expansion of $\log_2(x)$ about $x = 1$. (2) Substitute x with $\hat{\rho}$. Then, one gets $\log_2(\hat{\rho}) = k \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(\hat{\rho}-1)^n}{n}$, where k is a constant.³ However, this is rather cumbersome. The more straightforward way is to recognize that there exists a basis where the density matrix is diagonal. Taking the log of the diagonalized density matrix is just taking the log of each diagonal element separately, element by element, which is the second equality in Eq. (3.14).

■ **Example 3.5** Calculate the von Neumann entropy for a pure state $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and for a completely mixed state $\hat{\rho} = \frac{1}{2}\mathbb{I}$.

²In information theory, Shannon entropy is defined in a similar but not identical way. For diagonal matrices, the two entropy definitions are the same. According to Claude Shannon, “*My greatest concern was what to call it. I thought of calling it ‘information’, but the word was overly used, so I decided to call it ‘uncertainty’. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, ‘You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.’*

³Recall: $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$ and $\frac{d \log_2(x)}{dx} = \frac{1}{\ln(2)} \frac{d \ln(x)}{dx} = \frac{1}{x \ln(2)}$. Finding the pattern in the derivatives, we can arrive at $\log_2(x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(x-1)^n}{n \ln(2)}$. Note that different texts use a different base for the log in the definition. They simply differ by a constant from change of base.

Properties of von Neumann entropy

- $S(\hat{\rho}) = 0$ iff. $\hat{\rho}$ represents a pure state.
- $S(\hat{\rho}) = \log_2(d)$ is maximal if $\hat{\rho}$ represents a maximally mixed state, where d is the dimension of the Hilbert space.
- $S(\hat{\rho})$ is invariant under change of basis of $\hat{\rho}$, i.e. $S(\hat{\rho}) = S(\hat{U}^\dagger \hat{\rho} \hat{U})$.
- $S(\hat{\rho})$ is additive for independent systems, i.e. $S(\hat{\rho}_A + \hat{\rho}_B) = S(\hat{\rho}_A) + S(\hat{\rho}_B)$.
- $S(\hat{\rho})$ is sub-additive in general, i.e. $S(\hat{\rho}_{AB}) \leq S(\hat{\rho}_A) + S(\hat{\rho}_B)$, where $\hat{\rho}_A$ and $\hat{\rho}_B$ are reduced density matrices of the composite system AB .

3.3 Reduced Density Matrices and Partial Trace

As we saw in the previous chapter, if a system comprises two subsystems \mathcal{A} and \mathcal{B} , then a state described by $|\Psi\rangle = |\psi\rangle_A |\phi\rangle_B$ is not entangled. This is called a product state. In the same way, if the density operator of the system can be written as a tensor product of density operators of each subsystem, $\hat{\rho} = \hat{\rho}_A \otimes \hat{\rho}_B$, then the subsystems are not entangled, and there are no quantum correlations between them.

Now, let us suppose that the experimenter has no access to one of the subsystems, say \mathcal{B} , and can only make measurements on \mathcal{A} . For example, consider the following pure state of the composite system,

$$|\Psi\rangle = \frac{1}{\sqrt{2}} (|a_1\rangle |b_1\rangle + |a_2\rangle |b_2\rangle), \quad (3.15)$$

$$\text{equivalently, } \hat{\rho} = |\Psi\rangle \langle \Psi| = \frac{1}{2} \sum_{i,j=1}^2 |a_i\rangle \langle a_j| \otimes |b_i\rangle \langle b_j|. \quad (3.16)$$

Then, let $\{|\psi_k\rangle\}$ and $\{|\phi_l\rangle\}$ be orthonormal basis sets of \mathcal{A} and \mathcal{B} respectively. An observable acting on \mathcal{A} only is given by $\hat{O} \equiv \hat{O}_A \otimes \hat{I}_B$. Then the expectation value of the observable can be calculated using the trace rule, giving

$$\begin{aligned} \langle O \rangle &= \text{Tr}(\hat{\rho} \hat{O}) \\ &= \sum_{k,l} \langle \phi_l | \langle \psi_k | \hat{\rho} (\hat{O}_A \otimes \hat{I}_B) | \psi_k \rangle | \phi_l \rangle \\ &= \sum_k \langle \psi_k | \underbrace{\left(\sum_l \langle \phi_l | \hat{\rho} | \phi_l \rangle \right)}_{\text{Tr}_B(\hat{\rho}) \equiv \hat{\rho}_A} \hat{O}_A | \psi_k \rangle \\ &= \sum_k \langle \psi_k | \hat{\rho}_A \hat{O}_A | \psi_k \rangle \\ &= \text{Tr}_A(\hat{\rho}_A \hat{O}_A). \end{aligned} \quad (3.17)$$

Reduced density matrix

Therefore, the mathematical object that will encapsulate all possible information that can be obtained by the observer on \mathcal{A} is given by the reduced density matrix $\hat{\rho}_A$, defined as

$$\hat{\rho}_A \equiv \text{Tr}_B(\hat{\rho}), \quad (3.18)$$

where $\hat{\rho}$ is the full density matrix and Tr_B represents a *partial trace* over all the basis states of \mathcal{B} . The partial trace may be interpreted as taking an average over the degrees of freedom of the unobserved system \mathcal{B} .

The concept of reduced density matrices can be generalized from bipartite entangled states to any pure state $|\psi\rangle$ or its density matrix $\hat{\rho} \equiv |\psi\rangle \langle \psi|$, describing entanglement between N

subsystems. Consider an observable \hat{O} pertaining to system i only:

$$\hat{O} = \hat{I}_1 \otimes \hat{I}_2 \otimes \cdots \otimes \hat{I}_{i-1} \otimes \hat{O}_i \otimes \hat{I}_{i+1} \otimes \cdots \otimes \hat{I}_N. \quad (3.19)$$

The measurement statistics of \hat{O} is identical whether we use the pure state density matrix $\hat{\rho}$ or the reduced density matrix of the i -th system $\hat{\rho}_i$, i.e.

$$\langle \hat{O} \rangle = \text{Tr}(\hat{\rho}\hat{O}) = \text{Tr}_i(\hat{\rho}_i\hat{O}_i), \quad \text{where } \hat{\rho}_i = \text{Tr}_{1,\dots,i-1,i+1,\dots,N}(\hat{\rho}). \quad (3.20)$$

■ **Example 3.6** Deduce that the reduced density matrix of system \mathcal{A} , from a pure state $|\Psi\rangle = \frac{1}{\sqrt{2}} \sum_{i=1}^2 |a_i\rangle |b_i\rangle$ describing a composite system of two qubits, is given by

$$\hat{\rho}_A = \frac{1}{2} \sum_{i,j=1}^2 |a_i\rangle \langle a_j| \langle b_j| b_i\rangle. \quad (3.21)$$

■

Generalizing to a composite system of dimensionality N , such that a pure state of the composite system is described by $|\Psi\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^N |a_i\rangle |b_i\rangle$, we have the reduced density matrix being

$$\hat{\rho}_A = \frac{1}{N} \sum_{i,j=1}^N |a_i\rangle \langle a_j| \langle b_j| b_i\rangle. \quad (3.22)$$

Eq. (3.22) provides a very important result which will be elaborated upon in the next chapter on decoherence.

Local Measureability of Interference and Distinguishability

The degree of interference of $\hat{\rho}_A$, represented by off-diagonal elements $|a_i\rangle \langle a_j|$ ($i \neq j$), is quantified by the amount of overlap (or distinguishability) between corresponding states $|b_i\rangle, |b_j\rangle$, given by $\langle b_i|b_j\rangle$.

In the case of zero overlap, the reduced density matrix is completely diagonal in the $\{|a_i\rangle\}$ basis, and no local observable will allow us to measure interference between states $|a_i\rangle, |a_j\rangle$.

3.4 Reduced Density Matrices and Schmidt Decomposition

In Chapter 2.2, we provided the Schmidt decomposition theorem. Here, we complete that discussion since it relates to density operators.

The Schmidt decomposition theorem is very useful for bipartite systems. Consider the following. Suppose a pure state $|\psi\rangle$ of a composite system \mathcal{AB} has the Schmidt decomposition,

$$|\psi\rangle = \sum_{i=1}^d \alpha_i |\chi_i\rangle_A |\chi_i\rangle_B, \quad \text{where } d = \min(d_A, d_B) \text{ and } \alpha_i > 0, \alpha_i \in \mathbb{R}. \quad (3.23)$$

The density operator is

$$\hat{\rho} = |\psi\rangle \langle \psi| = \sum_{i=1}^d \sum_{j=1}^d \alpha_i \alpha_j (|\chi_i\rangle_A \langle \chi_j|) \otimes (|\chi_i\rangle_B \langle \chi_j|). \quad (3.24)$$

The reduced density operators of the subsystems are then

$$\hat{\rho}_A = \sum_{i=1}^d \alpha_i^2 |\chi_i\rangle_A \langle \chi_i|, \quad (3.25)$$

$$\hat{\rho}_B = \sum_{i=1}^d \alpha_i^2 |\chi_i\rangle_B \langle \chi_i|. \quad (3.26)$$

Therefore, the eigenvalues of $\hat{\rho}_A$ and $\hat{\rho}_B$ identical, namely α_i^2 for both density operators, with eigenvectors $|\chi_i\rangle_A$ and $|\chi_i\rangle_B$. Therefore, we can find the Schmidt decomposition of a pure bipartite state by finding the eigen-decomposition of the reduced density matrices.

Since many important properties of quantum systems are completely determined by the eigenvalues of the reduced density operator of the system, so for a *pure state* of a composite system such properties will be the same for both systems.

■ **Example 3.7** Compute the Schmidt decomposition of the following two-qubit states. For the simplest cases, you can perform the decompose by inspection or factorization.

- (i) $|\psi\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$.
- (ii) $|\psi\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |01\rangle + |10\rangle + |11\rangle)$.
- (iii) $|\psi\rangle = \frac{1}{3} (|00\rangle + |01\rangle + |11\rangle)$.

PH3406 (2024) T.S.KOH

3.4.1 Reduced Density Matrices and Decoherence

The loss of coherence, or decoherence, of an open system arises from interactions between the open system and its environment. System-environment interactions, e.g. photons scattering off an object or magnetic interactions between background nuclear spins and the spin of a system electron, typically lead to entanglement between the system and the environment. Typically, an experimenter has access to information about the system of interest, e.g. vibrational state of an ion array in an ion trap, or the spin state of an electron in a quantum dot. Information about the environment, e.g. ambient photons or background charges in a semiconductor device, is inaccessible because it contains a huge number of degrees of freedom, or there are practical limitations to measurements of the environment, or is simply of no interest.

This is where the reduced density matrix comes in – by tracing over the degrees of freedom of the environment of the combined system–environment density matrix, we obtain a complete description of the measurement statistics in terms of the reduced density matrix of the open system. All influences of the environment on local measurements performed on the system will automatically be encapsulated in this reduced density matrix. However, because the system is entangled with its environment, no individual pure state can be attributed to the system itself. Therefore, the reduced density matrix is, in general, a mixed state due to the system–environment entanglement.

In summary, even if the system started in a pure and coherent state, after interacting with its environment, and tracing out the environment degrees of freedom, the description of the system is one of a mixed state where the coherence terms are lost. We will elaborate upon the result of Eq. (3.22) and see how it can be so in the next chapter on decoherence.

3.4.2 Conceptual Notes on Mixed State Density Matrices

Classical vs quantum origin

Mixed state density matrices may arise in two fundamentally distinct ways.

The first way is through a state preparation procedure that produces different pure states. The density matrix is then a reflection of the observer's ignorance of which pure state was prepared in a particular experiment, which is then described using a statistical distribution of pure states. The classical probabilities here represent subjective ignorance because, in principle, one could retrace the preparation procedure to determine which pure state was prepared. This statistical mixture is called a “proper” mixture.

The second way in which a mixed state may be obtained is when a system S is entangled with an environment E . As we saw, the reduced density matrix of the system $\hat{\rho}_S = \text{Tr}_E(\hat{\rho}_{SE})$ will be mixed. This is called an “improper” mixture. This is a reflection of the loss of information about a particular pure state of the larger system-environment composite, and is quantum-mechanical in nature (arising from entanglement). Therefore, this situation is distinct from the first; it is not one in which the system is in a pure state but the observer does not know which. One cannot attribute subjective ignorance to such a mixed state.

Take the canonical example of Ex. 3.6, where a qubit (A) is entangled with another system, say an environment (B). In the limit of vanishing overlap $\langle b_i | b_j \rangle = \delta_{ij}$, the reduced density matrix of a qubit (A) becomes $\hat{\rho}_A = \frac{1}{2} (|0\rangle\langle 0| + |1\rangle\langle 1|)$. Although this is mathematically identical to the density matrix corresponding to a situation where the system is in either one of the pure states $|0\rangle$ or $|1\rangle$ with equal probability, these are distinct situations representing improper and proper mixtures. In general, density matrices are mathematical tools for calculating the probability distribution of outcomes of measurements, but they do not specify the state of the system.

Physical ensembles vs mixed-states

An important distinction in the description of physical ensembles and mixed-states is needed.

Consider a physical ensemble of N identical systems, each prepared in one of the pure states $\{|\psi_i\rangle\}$, $i = 1, \dots, M \leq N$, with the relative fraction of systems prepared in state $|\psi_i\rangle$ being $p_i \equiv N_i/N$. This is typically, but *erroneously*, given a mixed-state density matrix description

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (3.27)$$

How such an ensemble of N individual systems should be described is by a pure tensor product state of individual states, in the tensor product Hilbert space of the N constituent state spaces, as follows:

$$|\Psi\rangle = \left(\prod_{j=1}^{N_1} |\psi_1\rangle \right) \left(\prod_{j=1}^{N_2} |\psi_2\rangle \right) \dots \left(\prod_{j=1}^{N_M} |\psi_M\rangle \right), \quad (3.28)$$

$$\mathcal{H} = \otimes_{k=1}^N \mathcal{H}_k. \quad (3.29)$$

To be precise, it is only in the sense of an *ensemble average* where the statistical concept of a single-system mixed-state density matrix can be justified. This is the case if the average statistics for the entire ensemble is constructed based on measurements on single systems in the ensemble. That is, when observable on single systems $\hat{O} = \sum_i \hat{O}_i$ is measured in the N -system pure state of Eq. (3.28), and the average is obtained,

$$\langle \hat{O} \rangle = \frac{1}{N} \sum_{i=1}^M N_i \langle \psi_i | \hat{O} | \psi_i \rangle = \sum_{i=1}^M p_i \langle \psi_i | \hat{O} | \psi_i \rangle \quad (3.30)$$

where we have assumed for simplicity that $\{|\psi_i\rangle\}$ form an orthonormal set of basis states for each individual system. Computationally, we get the same result from using the trace rule on the mixed state description, hence the conceptual confusion. The trace rule on Eq. (3.27) gives

$$\langle \hat{O} \rangle = \text{Tr}(\hat{\rho}\hat{O}) = \sum_{i=1}^M p_i \langle \psi_i | \hat{O} | \psi_i \rangle. \quad (3.31)$$

Therefore, the averaged statistics of a single-system observable are identical for the single-system mixed state (Eq. (3.27)) and the composite N -system pure state (Eq. (3.28)). However, the proper pure-state N -system description (Eq. (3.28)) of the ensemble allows one to compute the statistics for *any* arbitrary observable over one or more systems in the ensemble.

3.5 Postulates of Quantum Mechanics (Revisited)

We now restate the postulates of quantum mechanics using the more general density operator formalism, following Lidar's notes⁴.

Postulate 1 – Description of the state of a system.

- **State space.** The state space is the Hilbert-Schmidt space of linear operators $\hat{\rho}$ such that $\text{Tr}(\hat{\rho}) = 1$ and $\hat{\rho} \geq 0$ (i.e. eigenvalues are *non-negative*).
- **Inner product.** The inner product between operators in the Hilbert-Schmidt space is defined as

$$\langle \hat{A}, \hat{B} \rangle := \text{Tr}(\hat{A}^\dagger \hat{B}), \quad (3.32)$$

⁴Section III.C of “Lecture notes on open quantum systems”. <https://arxiv.org/abs/1902.00967>.

for any two operators \hat{A} and \hat{B} acting on the space. This inner product defines a length or norm in the usual way,

$$\|\hat{\rho}\| = \sqrt{\langle \hat{\rho}, \hat{\rho} \rangle} = \sqrt{\text{Tr}(\hat{\rho}^\dagger \hat{\rho})} = \sqrt{\text{Tr}(\hat{\rho}^2)} = \sqrt{\xi}, \quad (3.33)$$

where $\xi \equiv \text{Tr}(\hat{\rho}^2)$ is the purity of the state. Thus a density matrix can have “length” or norm ≤ 1 . A state is called “pure” if $\xi = 1$ and “mixed” if $\xi < 1$.

Postulate 2 — Combining quantum systems. State spaces are composed via the tensor product \otimes .

Postulate 3 — Time evolution (dynamics). Density operators evolve as

$$\frac{\partial \hat{\rho}}{\partial t} = -\frac{i}{\hbar} [\hat{H}, \hat{\rho}] \quad (3.34)$$

under a Hamiltonian \hat{H} , or equivalently as

$$\hat{\rho}(t) = \hat{U}(t)\hat{\rho}(0)\hat{U}^\dagger(t), \quad (3.35)$$

where the unitary time-evolution operator is $\hat{U}(t)$. (For time-independent Hamiltonians, $\hat{U}(t) = e^{-i\hat{H}t/\hbar}$. The usual cases of time-dependence apply, see Section 1.3.)

Postulate 4 — Generalized measurement. A general measurement operation defined by elements $\{\hat{M}_k\}$ results with probability of k -th outcome given by

$$p_k = \text{Tr}(\hat{\rho}\hat{M}_k^\dagger \hat{M}_k), \text{ with } \sum_k p_k = 1, \quad (3.36)$$

and the state transforms upon measurement as

$$\hat{\rho} \rightarrow \frac{\hat{M}_k \hat{\rho} \hat{M}_k^\dagger}{p_k}. \quad (3.37)$$

3.5.1 Visualizing the density matrix of a qubit on the Bloch sphere

The density operator for a state in a two-dimensional complex space spanned by $\{|0\rangle, |1\rangle\}$ can be written as

$$\hat{\rho} = \begin{pmatrix} a & b \\ b^* & 1-a \end{pmatrix}, \quad \text{where } a \in \mathbb{R} \text{ and } b \in \mathbb{C}. \quad (3.38)$$

We can also write the density matrix of a qubit in terms of linear combinations of the Pauli matrices and the identity:

$$\hat{\rho} = \frac{1}{2}(\mathbb{I} + \sum_{i=x,y,z} v_i \sigma_i) = \frac{1}{2}(\mathbb{I} + \vec{v} \cdot \vec{\sigma}) = \frac{1}{2} \begin{pmatrix} 1+v_z & v_x - iv_y \\ v_x + iv_y & 1-v_z \end{pmatrix}. \quad (3.39)$$

We require $b = (v_x - iv_y)/2$ and $a = (1 + v_z)/2$, and call \vec{v} the Bloch vector.

■ **Example 3.8** Eigenvalues of the density matrix are non-negative. Show that this condition on $\hat{\rho}$ in Eq. (3.39) implies that $\|\vec{v}\| \leq 1$. ■

■ **Example 3.9** Calculate purity, $\xi = \text{Tr}(\hat{\rho}^2)$, and deduce that the norm of the Bloch vector, $\|\vec{v}\|^2 = 1$ represents a pure state and $\|\vec{v}\|^2 < 1$ represents a mixed state. ■

■ **Example 3.10** The Hilbert-Schmidt inner product is defined between operators in Hilbert-Schmidt space. For a two-level system, operators in this space can be represented by 2×2 matrices. There are many ways to write an operator basis. One such operator basis is the following.

$$\hat{O}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \hat{O}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \hat{O}_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \hat{O}_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.40)$$

Verify that this set of operators form an orthonormal basis, and express the density matrix of Eq. (3.39) in this basis. ■

3.5.2 Derivation of time evolution of density operator (Schrödinger picture)

Here, we shall derive Eq. (3.35) and Eq. (3.34).

$$\begin{aligned}\hat{\rho}(t) &= \sum_i p_i |\psi_i(t)\rangle \langle \psi_i(t)| \\ &= \sum_i p_i (\hat{U}(t) |\psi_i(0)\rangle) (\langle \psi_i(0)| \hat{U}^\dagger(t)) \\ &= \hat{U}(t) \left(\sum_i p_i |\psi_i(0)\rangle \langle \psi_i(0)| \right) \hat{U}^\dagger(t) \\ &= \hat{U}(t) \hat{\rho}(0) \hat{U}^\dagger(t).\end{aligned}\quad (3.41)$$

Taking the time derivative of the density operator,

$$\begin{aligned}\frac{\partial \hat{\rho}(t)}{\partial t} &= \frac{\partial}{\partial t} \sum_i p_i |\psi_i(t)\rangle \langle \psi_i(t)| \\ &= \sum_i p_i \left(\underbrace{\frac{\partial |\psi_i(t)\rangle}{\partial t}}_{-i\hat{H}(t)|\psi_i(t)\rangle/\hbar} \langle \psi_i(t)| + |\psi_i(t)\rangle \underbrace{\frac{\partial \langle \psi_i(t)|}{\partial t}}_{i\hat{H}^\dagger(t)\langle \psi_i(t)|/\hbar} \right) \\ &= -\frac{i}{\hbar} \sum_i p_i \left(\hat{H}(t) |\psi_i(t)\rangle \langle \psi_i(t)| - |\psi_i(t)\rangle \langle \psi_i(t)| \underbrace{\hat{H}^\dagger(t)}_{=\hat{H}(t)} \right) \\ &= -\frac{i}{\hbar} (\hat{H}(t) \hat{\rho}(t) - \hat{\rho}(t) \hat{H}(t)) \\ \therefore \frac{\partial \hat{\rho}(t)}{\partial t} &= -\frac{i}{\hbar} [\hat{H}(t), \hat{\rho}(t)].\end{aligned}\quad (3.42)$$

This is known as the Liouville-von Neumann equation.⁵ Below, we give the equation in the Schrödinger and interaction pictures.

Liouville-von Neumann Equation

Schrödinger picture:

$$\frac{\partial \hat{\rho}(t)}{\partial t} = -\frac{i}{\hbar} [\hat{H}(t), \hat{\rho}(t)] \quad (3.43)$$

$$\text{Equivalently, } \hat{\rho}(t) = \hat{U}(t) \hat{\rho}(0) \hat{U}^\dagger(t), \quad (3.44)$$

$$\text{where } \hat{U}(t) \text{ satisfies } i\hbar \frac{\partial}{\partial t} \hat{U}(t) = \hat{H}(t) \hat{U}(t).$$

Interaction picture:

$$\frac{\partial \hat{\rho}^I(t)}{\partial t} = -\frac{i}{\hbar} [\hat{V}^I(t), \hat{\rho}^I(t)]. \quad (3.45)$$

$$\text{where } \hat{V}^I(t) = \hat{U}_0^\dagger(t) \hat{V}(t) \hat{U}_0(t); \hat{U}_0(t) = e^{-iH_0 t/\hbar}; \hat{H}(t) = \hat{H}_0 + \hat{V}(t).$$

$$\text{Equivalently, } \hat{\rho}^I(t) = \hat{U}^I(t) \hat{\rho}^I(0) \hat{U}^{I\dagger}(t), \quad (3.46)$$

$$\text{where } \hat{U}^I(t) \text{ satisfies } i\hbar \frac{\partial}{\partial t} \hat{U}^I(t) = \hat{V}^I(t) \hat{U}^I(t).$$

⁵Note the commonly confused difference in the sequence of terms in the commutator on the RHS, from the Heisenberg equation of motion given in Eq. (1.77).

3.6 Discussion Set

Problem 3.1 Density matrices. Write the density matrix for these single qubit states:

- (i) $|\psi_1\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$,
- (ii) mixed state with 50% $|0\rangle$ and 50% $|1\rangle$,
- (iii) $|\psi_2\rangle = (|0\rangle + e^{i\phi_0}|1\rangle)/\sqrt{2}$,
- (iv) equal probabilities of the states

$$|\phi_1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |\phi_2\rangle = \frac{1}{2} \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix}, \quad |\phi_3\rangle = \frac{1}{2} \begin{pmatrix} 1 \\ -\sqrt{3} \end{pmatrix}. \quad (3.47)$$

Problem 3.2 Reduced density matrices. Suppose a composite of systems A and B is in the following states

- (i) $|a\rangle \otimes |b\rangle$,
- (ii) $\frac{1}{\sqrt{2}}(|a_1\rangle \otimes |b_1\rangle + |a_2\rangle \otimes |b_2\rangle)$.

where $|a\rangle, |b\rangle$ are pure states of systems A and B respectively. In which case is the reduced density operator of system A (a) a pure state, (b) a mixed state?

Problem 3.3 Hilbert-Schmidt inner product. A density matrix is given by

$$\hat{\rho} = \begin{pmatrix} 0.4 & 0.25 - 0.5i \\ 0.25 + 0.5i & 0.6 \end{pmatrix}. \quad (3.48)$$

Using the Hilbert-Schmidt inner product and the following operator basis, answer the questions below.

$$\text{operator basis : } \frac{1}{\sqrt{2}} \{ \mathbb{I}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z \}. \quad (3.49)$$

- (i) What are the Bloch vector components corresponding to this density matrix?
- (ii) Is the state pure or mixed?
- (iii) Why is the pre-factor $\frac{1}{\sqrt{2}}$ necessary for the operator basis?

Problem 3.4 Purity. Show that the purity is bounded as such:

$$\frac{1}{d} \leq \xi \leq 1, \quad (3.50)$$

where d is the dimension of the Hilbert space. What kind of state has a purity of (i) 1, (ii) $1/d$?

Problem 3.5 von Neumann Entropy. Write the density matrix $\hat{\rho}_{AB}$ for the Bell state $|\Psi^+\rangle$, and the reduced density matrix $\hat{\rho}_A = \text{Tr}_B(\hat{\rho}_{AB})$. Calculate and compare the von Neumann entropies $S(\hat{\rho}_{AB})$ and $S(\hat{\rho}_A)$, and purities ξ_A and ξ_B . Comment on the state described by $\hat{\rho}_A$. Give an example of a composite state that gives $S(\hat{\rho}_A) = S(\hat{\rho}_B) = 0$.

Problem 3.6 von Neumann Entropy and Fidelity. Consider the pure state

$$|\psi\rangle = \alpha|00\rangle + \beta|11\rangle \quad (3.51)$$

in the Hilbert space \mathcal{H}_{AB} spanned by the tensor product of Hilbert spaces of qubits A and B .

- (i) Write the density matrix $\hat{\rho}$ for the state $|\psi\rangle$.
- (ii) Determine an expression for the von Neumann entropy of the reduced density matrix of A , $S(\hat{\rho}_A) = -\text{Tr}(\hat{\rho}_A \log_2(\hat{\rho}_A))$, in terms of α .
- (iii) Let $\tilde{\rho}$ be a density matrix for a disentangled state on \mathcal{H}_{AB} . Find an expression for the fidelity between two quantum states, $\tilde{\rho}$ and $\hat{\rho}$, defined as $\mathcal{F}(\tilde{\rho}, \hat{\rho}) = \left| \text{Tr}(\sqrt{\sqrt{\hat{\rho}}\tilde{\rho}\sqrt{\hat{\rho}}}) \right|^2$, where $\sqrt{\hat{\rho}}$ is defined such that $(\sqrt{\hat{\rho}})^2 = \hat{\rho}$.

Problem 3.7 Distinguishability. Consider two mixed states

- $\cos \theta |0\rangle + \sin \theta |1\rangle$ and $\cos \theta |0\rangle - \sin \theta |1\rangle$, with equal probabilities, and
- $|0\rangle$ and $|1\rangle$, with equal probabilities.

Are their density matrices distinguishable?

Problem 3.8 Measurement. Suppose a mixed state $\hat{\rho}$ is given by the probabilities and pure states, $\{p_j, |\psi_j\rangle\}$. This is measured in an orthonormal basis $\{|\beta_j\rangle\}$, i.e. with projectors $|\beta_j\rangle\langle\beta_j|$. Show that the probability of measuring the k -th state $|\beta_k\rangle$ is $\langle\beta_k|\hat{\rho}|\beta_k\rangle$. What is the density matrix after measurement?

Problem 3.9 Schmidt Decomposition. Find the Schmidt decomposition of a two-qubit composite state,

$$|\psi\rangle = \frac{1}{2\sqrt{2}} (|00\rangle + \sqrt{3}|01\rangle + \sqrt{3}|10\rangle + |11\rangle).$$

Time evolution

Problem 3.10 Jaynes-Cummings Hamiltonian; interaction picture. The interaction of a two-level atom with a photon (i.e. quantized electromagnetic radiation⁶) may be described by a Jaynes-Cummings Hamiltonian,

$$\hat{H} = \hat{H}_{\text{atom}} + \hat{H}_{\text{field}} + \hat{H}_{\text{int}}, \quad (3.52)$$

$$\text{where } \hat{H}_{\text{atom}} = \frac{\hbar\omega_0}{2}\sigma_z, \hat{H}_{\text{field}} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2}), \hat{H}_{\text{int}} = \hbar g(\hat{\sigma}_+\hat{a} + \hat{\sigma}_-\hat{a}^\dagger). \quad (3.53)$$

where \hat{a}, \hat{a}^\dagger are bosonic annihilation and creation operators describing the quantized electromagnetic field of angular frequency ω , and $\hat{\sigma}_z = |e\rangle\langle e| - |g\rangle\langle g|$, $\hat{\sigma}_+ = |e\rangle\langle g| = \hat{\sigma}_-^\dagger$, with $|e\rangle, |g\rangle$ describing excited and ground atomic states. The photonic Hilbert space is spanned by the Fock states $\{|n\rangle\}$, which are eigenstates of the number operator $\hat{n} \equiv \hat{a}^\dagger\hat{a}$. In this composite system, \hat{H}_{int} describes the atom-field interaction, while $\hat{H}_0 = \hat{H}_{\text{field}} + \hat{H}_{\text{atom}}$ contains the unperturbed terms. For this problem, we go into the interaction picture rotating with \hat{H}_0 .

- Find the time-dependence of the following operators, $\hat{a}, \hat{a}^\dagger, \hat{\sigma}_+, \hat{\sigma}_-, \hat{\sigma}_+\hat{a}$, and $\hat{\sigma}_-\hat{a}^\dagger$.
- Deduce the resonant condition by which \hat{H}_{int}^I becomes time-independent in the interaction picture.

Now, consider the atom-field system to be initially separable, $\hat{\rho}(0) = \hat{\rho}_{\text{field}}(0) \otimes \hat{\rho}_{\text{atom}}(0)$, with $\hat{\rho}_{\text{atom}}(0) = |e\rangle\langle e|$. The atomic inversion factor $W(t)$ is defined as the difference between the atomic populations in the excited and ground states, and is a function of time.

- Show that the interaction picture density operator may be written in the form,

$$\hat{\rho}^I(t) = \begin{pmatrix} \hat{C}(t)\hat{\rho}_{\text{field}}(0)\hat{C}(t) & -\hat{C}(t)\hat{\rho}_{\text{field}}(0)\hat{S}'(t) \\ \hat{S}(t)\hat{\rho}_{\text{field}}(0)\hat{C}(t) & -\hat{S}(t)\hat{\rho}_{\text{field}}(0)\hat{S}'(t) \end{pmatrix}, \quad (3.54)$$

$$\text{where } \hat{C}(t) = \cos(gt\sqrt{\hat{a}\hat{a}^\dagger}), \quad \hat{C}'(t) = \cos(gt\sqrt{\hat{a}^\dagger\hat{a}}),$$

$$\hat{S}(t) = -i\hat{a}^\dagger \frac{\sin(gt\sqrt{\hat{a}\hat{a}^\dagger})}{\sqrt{\hat{a}\hat{a}^\dagger}}, \quad \hat{S}'(t) = -i\hat{a} \frac{\sin(gt\sqrt{\hat{a}^\dagger\hat{a}})}{\sqrt{\hat{a}^\dagger\hat{a}}}.$$

- Determine the reduced density operator of the atom, $\hat{\rho}_{\text{atom}}^I(t)$, from Eq. (3.54).
- If the field was initially in a pure state $|\Psi\rangle = \sum_{n=0}^{\infty} c_n |n\rangle$, determine the atomic inversion factor, $W(t)$. Make a plot of W against dimensionless time gt for $c_0 = 1$, and deduce the behaviour of the system from an interpretation of the plot.
- If the field was initially in a thermal state, determine the atomic inversion factor, $W(t)$. The initial thermal state is given by $\hat{\rho}_{\text{field}}(0) = \sum_{n=0}^{\infty} P_n |n\rangle\langle n|$, where $P_n = \frac{\bar{n}^n}{(1+\bar{n})^{n+1}}$ and $\frac{\bar{n}}{1+\bar{n}} = \exp\left(-\frac{\hbar\omega}{k_B T}\right)$. Make a numerical plot of W against gt for $\bar{n} = 1$, room temperature T and microwave frequency ω . (You should truncate the sum at a suitable n .)

⁶This is in contrast with the classical electromagnetic field modelled in Problem 1.10.