

Physics 125c
Course Notes
Density Matrix Formalism
040511 Frank Porter

1 Introduction

In this note we develop an elegant and powerful formulation of quantum mechanics, the “density matrix” formalism. This formalism provides a structure in which we can address such matters as:

- We typically assume that it is permissible to work within an appropriate subspace of the Hilbert space for the universe. Is this all right?
- In practice we often have situations involving statistical ensembles of states. We have not yet addressed how we might deal with this.

2 The Density Operator

Suppose that we have a state space, with a denumerable orthonormal basis $\{|u_n\rangle, n = 1, 2, \dots\}$. If the system is in state $|\psi(t)\rangle$ at time t , we have the expansion in this basis:

$$|\psi(t)\rangle = \sum_n a_n(t)|u_n\rangle. \quad (1)$$

We'll assume that $|\psi(t)\rangle$ is normalized, and hence:

$$\begin{aligned} \langle\psi(t)|\psi(t)\rangle = 1 &= \sum_n \sum_m a_n(t)a_m^*(t)\langle u_m|u_n\rangle \\ &= \sum_n |a_n(t)|^2 \end{aligned} \quad (2)$$

Suppose that we have an observable (self-adjoint operator) Q . The matrix elements of Q in this basis are:

$$Q_{mn} = \langle u_m|Qu_n\rangle = \langle Qu_m|u_n\rangle = \langle u_m|Q|u_n\rangle. \quad (3)$$

The average (expectation) value of Q at time t , for the system in state $|\psi(t)\rangle$ is:

$$\langle Q \rangle = \langle\psi(t)|Q\psi(t)\rangle = \sum_n \sum_m a_m^*(t)a_n(t)Q_{mn}. \quad (4)$$

We see that $\langle Q \rangle$ is an expansion quadratic in the $\{a_n\}$ coefficients.

Consider the operator $|\psi(t)\rangle\langle\psi(t)|$. It has matrix elements:

$$\langle u_m|\psi(t)\rangle\langle\psi(t)|u_n\rangle = a_m(t)a_n^*(t). \quad (5)$$

These matrix elements appear in the calculation of $\langle Q \rangle$. Hence, define

$$\rho(t) \equiv |\psi(t)\rangle\langle\psi(t)|. \quad (6)$$

We call this the **density operator**. It is a Hermitian operator, with matrix elements

$$\rho_{mn}(t) = \langle u_m|\rho(t)u_n\rangle = a_m(t)a_n^*(t). \quad (7)$$

Since $\psi(t)$ is normalized, we also have that

$$1 = \sum_n |a_n(t)|^2 = \sum_n \rho_{nn}(t) = \text{Tr}[\rho(t)]. \quad (8)$$

We may now re-express the expectation value of observable Q using the density operator:

$$\begin{aligned} \langle Q \rangle(t) &= \sum_m \sum_n a_m^*(t)a_n^*(t)Q_{mn} \\ &= \sum_m \sum_n \rho_{nm}(t)Q_{mn} \\ &= \sum_n [\rho(t)Q]_{nn} \\ &= \text{Tr}[\rho(t)Q]. \end{aligned} \quad (9)$$

The time evolution of a state is given by the Schrödinger equation:

$$i\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \quad (10)$$

where $H(t)$ is the Hamiltonian. Thus, the time evolution of the density operator may be computed according to:

$$\begin{aligned} \frac{d}{dt}\rho(t) &= \frac{d}{dt} [|\psi(t)\rangle\langle\psi(t)|] \\ &= \frac{1}{i}H(t)|\psi(t)\rangle\langle\psi(t)| - \frac{1}{i}|\psi(t)\rangle\langle\psi(t)|H(t) \\ &= \frac{1}{i}[H(t), \rho(t)] \end{aligned} \quad (11)$$

Suppose we wish to know the probability, $P(\{q\})$, that a measurement of Q will yield a result in the set $\{q\}$. We compute this probability by projecting out of $|\psi(t)\rangle$ that portion which lies in the eigensubspace associated with observables in the set $\{q\}$. Let $P_{\{q\}}$ be the projection operator. Then:

$$\begin{aligned} P(\{q\}) &= \langle \psi(t) | P_{\{q\}} \psi(t) \rangle \\ &= \text{Tr} [P_{\{q\}} \rho(t)]. \end{aligned} \quad (12)$$

We note that the density operator, unlike the state vector, has no phase ambiguity. The same state is described by $|\psi(t)\rangle$ and $|\psi'(t)\rangle = e^{i\theta} |\psi(t)\rangle$. Under this phase transformation, the density operator transforms as:

$$\begin{aligned} \rho(t) \rightarrow \rho'(t) &= e^{i\theta} |\psi(t)\rangle \langle \psi(t)| e^{-i\theta} \\ &= \rho(t). \end{aligned} \quad (13)$$

Furthermore, expectation values are quadratic in $|\psi(t)\rangle$, but only linear in $\rho(t)$.

For the density operators we have been considering so far, we see that:

$$\begin{aligned} \rho^2(t) &= |\psi(t)\rangle \langle \psi(t)| |\psi(t)\rangle \langle \psi(t)| \\ &= \rho(t). \end{aligned} \quad (14)$$

That is, $\rho(t)$ is an **idempotent** operator. Hence,

$$\text{Tr} \rho^2(t) = \text{Tr} \rho(t) = 1. \quad (15)$$

Finally, notice that:

$$\langle u_n | \rho(t) u_n \rangle = \rho_{nn}(t) = |a_n(t)|^2 \geq 0 \quad \forall n. \quad (16)$$

Thus, for an arbitrary state $|\phi\rangle$, $\langle \phi | \rho(t) \phi \rangle \geq 0$, as may be demonstrated by expanding $|\phi\rangle$ in the $|u\rangle$ basis. We conclude that ρ is a non-negative definite operator.

We postulate, in quantum mechanics, that the states of a system are in one-to-one correspondence with the non-negative definite density operators of trace 1 (defined on the Hilbert space).

3 Statistical Mixtures

We may wish to consider cases where the system is in any of a number of different states, with various probabilities. The system may be in state $|\psi_1\rangle$ with probability p_1 , state $|\psi_2\rangle$ with probability p_2 , and so forth (more generally, we could consider states over some arbitrary, possibly non-denumerable, index set). We must have $1 \geq p_i \geq 0$ for $i \in \{\text{index set}\}$, and $\sum_i p_i = 1$. Note that this situation is *not* the same thing as supposing that we are in the state $|\psi\rangle = p_1|\psi_1\rangle + p_2|\psi_2\rangle + \dots$ (or even with $\sqrt{p_1}$, *etc.*). Such statistical mixtures might occur, for example, when we prepare a similar system (an atom, say) many times. In general, we will not be able to prepare the same exact state every time, but will have some probability distribution of states.

We may ask, for such a system, for the probability $P(\{q\})$ that a measurement of Q will yield a result in the set $\{q\}$. For each state in our mixture, we have

$$\begin{aligned} P_n(\{q\}) &= \langle \psi_n | P_{\{q\}} | \psi_n \rangle \\ &= \text{Tr}(\rho_n P_{\{q\}}), \end{aligned} \tag{17}$$

where $\rho_n = |\psi_n\rangle\langle\psi_n|$. To determine the overall probability, we must sum over the individual probabilities, weighted by p_n :

$$\begin{aligned} P(\{q\}) &= \sum_n p_n P_n(\{q\}) \\ &= \sum_n p_n \text{Tr}(\rho_n P_{\{q\}}) \\ &= \text{Tr}\left(\sum_n p_n \rho_n P_{\{q\}}\right) \\ &= \text{Tr}(\rho P_{\{q\}}), \end{aligned} \tag{18}$$

where

$$\rho \equiv \sum_n p_n \rho_n. \tag{19}$$

Now ρ is the density operator of the system, and is a simple linear combination of the individual density operators. Note that ρ is the “average” of the ρ_n ’s with respect to probability distribution p_n .

Let us investigate this density operator:

- Since ρ_n are Hermitian, and p_n are real, ρ is Hermitian.

- $\text{Tr}\rho = \text{Tr}(\sum_n p_n \rho_n) = \sum_n p_n \text{Tr}\rho_n = \sum_n p_n = 1$.
- ρ is non-negative-definite: $\langle \phi | \rho \phi \rangle = \sum_n p_n \langle \phi | \rho_n \phi \rangle \geq 0$.
- Let Q be an operator with eigenvalues q_n . In the current situation, $\langle Q \rangle$ refers to the average of Q over the statistical mixture. We have:

$$\begin{aligned}
\langle Q \rangle &= \sum_n q_n P(\{q_n\}) = \sum_n q_n \text{Tr}(\rho P_{\{q_n\}}) \\
&= \text{Tr}\left(\rho \sum_n q_n P_{\{q_n\}}\right) \\
&= \text{Tr}(\rho Q), \quad \text{since } Q = \sum_n q_n P_{\{q_n\}}. \tag{20}
\end{aligned}$$

- We may determine the time evolution of ρ . For $\rho_n(t) = |\psi_n(t)\rangle\langle\psi_n(t)|$ we know (Eqn. 11) that

$$i \frac{d\rho_n(t)}{dt} = [H(t), \rho_n(t)]. \tag{21}$$

Since $\rho(t)$ is linear in the ρ_n , $\rho(t) = \sum_n p_n \rho_n(t)$, we have

$$i \frac{d\rho(t)}{dt} = [H(t), \rho(t)]. \tag{22}$$

- Now look at

$$\begin{aligned}
\rho^2 &= \sum_m \sum_n p_m p_n \rho_m \rho_n \\
&= \sum_m \sum_n p_m p_n |\psi_m\rangle\langle\psi_m| \psi_n\rangle\langle\psi_n| \\
&\neq \rho, \quad \text{in general.} \tag{23}
\end{aligned}$$

What about the trace of ρ^2 ? Let

$$|\psi_m\rangle = \sum_j (a_m)_j |u_j\rangle. \tag{24}$$

Then

$$\begin{aligned}
\rho^2 &= \sum_m \sum_n p_m p_n |\psi_m\rangle\langle\psi_m| \psi_n\rangle\langle\psi_n| \\
&= \sum_m \sum_n p_m p_n \left[\sum_i \sum_j (a_m)_i^* (a_n)_j \delta_{ij} \right] \left[\sum_k \sum_\ell (a_m)_k (a_n)_\ell^* |u_k\rangle\langle u_\ell| \right] \\
&= \sum_{m,n,i,k,\ell} p_m p_n (a_m)_i^* (a_n)_i (a_m)_k (a_n)_\ell^* |u_k\rangle\langle u_\ell|. \tag{25}
\end{aligned}$$

Let's take the trace of this. Notice that $\text{Tr}(|u_k\rangle\langle u_\ell|) = \delta_{k\ell}$, so that

$$\text{Tr}(\rho^2) = \sum_{m,n,i,k} p_m p_n (a_m)_i^* (a_n)_i (a_m)_k (a_n)_k^*. \quad (26)$$

But $\langle \psi_m | \psi_n \rangle = \sum_i (a_m)_i^* (a_n)_i$, and thus:

$$\begin{aligned} \text{Tr}(\rho^2) &= \sum_m \sum_n p_m p_n |\langle \psi_m | \psi_n \rangle|^2 \\ &\leq \sum_m \sum_n p_m p_n \langle \psi_m | \psi_m \rangle \langle \psi_n | \psi_n \rangle, \quad (\text{Schwarz inequality}) \\ &\leq \sum_m p_m \sum_n p_n \\ &\leq 1. \end{aligned} \quad (27)$$

The reader is encouraged to check that equality holds if and only if the system can be in only one physical state (that is, all but one of the p_n 's corresponding to independent states must be zero).

Note that, if $\text{Tr}(\rho^2) = 1$, then $\rho = |\psi\rangle\langle\psi|$, which is a projection operator. We encapsulate this observation into the definition:

Def: A state of a physical system is called a **pure state** if $\text{Tr}(\rho^2) = 1$; the density operator is a projection. Otherwise, the system is said to be in a **mixed state**, or simply a mixture.

The diagonal matrix elements of ρ have a simple physical interpretation:

$$\begin{aligned} \rho_{nn} &= \sum_j p_j (\rho_j)_{nn} \\ &= \sum_j p_j \langle u_n | \psi_j \rangle \langle \psi_j | u_n \rangle \\ &= \sum_j p_j |(a_j)_n|^2. \end{aligned} \quad (28)$$

This is just the probability to find the system in state $|u_n\rangle$. Similarly, the off-diagonal elements are

$$\rho_{mn} = \sum_j p_j (a_j)_m (a_j)_n^*. \quad (29)$$

The off-diagonal elements are called **coherences**. Note that it is possible to choose a basis in which ρ is diagonal (since ρ is Hermitian). In such a basis, the coherences are all zero.

4 Measurements, Statistical Ensembles, and Density Matrices

Having developed the basic density matrix formalism, let us now revisit it, filling in some motivational aspects. First, we consider the measurement process. It is useful here to regard an experiment as a two-stage process:

1. Preparation of the system.
2. Measurement of some physical aspect(s) of the system.

For example, we might prepare a system of atoms with the aid of spark gaps, magnetic fields, laser beams, *etc.*, then make a measurement of the system by looking at the radiation emitted. The distinction between preparation and measurement is not always clear, but we'll use this notion to guide our discussion.

We may further remark that we can imagine any measurement as a sort of “counter” experiment: First, consider an experiment as a repeated preparation and measurement of a system, and refer to each measurement as an “event”. Think of the measuring device as an array of one or more “counters” that give a response (a “count”) if the variables of the system are within some range. For example, we might be measuring the gamma ray energy spectrum in some nuclear process. We have a detector which absorbs a gamma ray and produces an electrical signal proportional to the absorbed energy. The signal is processed and ultimately sent to a multichannel analyzer (MCA) which increments the channel corresponding to the detected energy. In this case, the MCA is functioning as our array of counters.

The process is imagined to be repeated many times, and we are not concerned with issues of the statistics of finite counting here. The result of such an experiment is expressed as the probability that the various counters will register, given the appropriate preparation of the system. These probabilities may include correlations.

Let us take this somewhat hazy notion and put it into more concrete mathematical language: Associate with each counter a **dichotomic variable**, D , as follows:

If the counter registers in an event, $D = 1$.

If the counter does not register in an event, $D = 0$.

We assert that we can, in principle, express all physical variables in terms of dichotomic ones, so this appears to be a sufficiently general approach.

By repeatedly preparing the system and observing the counter D , we can determine the probability that D registers: The average value of D , $\langle D \rangle$, is the probability that D registers in the experiment. We refer to the particular preparation of the system in the experiment as a **statistical ensemble** and call $\langle D \rangle$ the average of the dichotomic variable D with respect to this ensemble.

If we know the averages of all possible dichotomic variables, then the ensemble is completely known. The term “statistical ensemble” is synonymous with a suitable set of averages of dichotomic variables (*i.e.*, probabilities). Let us denote a statistical ensemble with the letter ρ . The use of the same symbol as we used for the density matrix is not coincidental, as we shall see. The quantity $\langle D \rangle_\rho$ explicitly denotes the average of D for the ensemble ρ . Clearly:

$$0 \leq \langle D \rangle_\rho \leq 1. \quad (30)$$

D is precisely known for ensemble ρ if $\langle D \rangle_\rho = 0$ or $\langle D \rangle_\rho = 1$. Otherwise, variable D possesses a statistical spread. Note that we may prepare a system (for example, an atom) many times according to a given ensemble. However, this does not mean that the system is always in the same *state*.

We have the important concept of the superposition of two ensembles: Let ρ_1 and ρ_2 be two distinct ensembles. An ensemble ρ is said to be an **incoherent superposition** of ρ_1 and ρ_2 if there exists a number θ such that $0 < \theta < 1$, and for every dichotomic variable D we have:

$$\langle D \rangle_\rho = \theta \langle D \rangle_{\rho_1} + (1 - \theta) \langle D \rangle_{\rho_2}. \quad (31)$$

This is expressed symbolically as:

$$\rho = \theta \rho_1 + (1 - \theta) \rho_2, \quad (32)$$

“ ρ is a superposition of ρ_1 and ρ_2 with probabilities θ and $1 - \theta$.”

We assume that if ρ_1 and ρ_2 are physically realizable, then any coherent superposition of them is also physically realizable. For example, we might prepare a beam of particles from two independent sources, each of which may hit our counter: ρ_1 corresponds to source 1, ρ_2 corresponds to source 2. When both sources are on, the beam hitting the counter is an incoherent mixture of ρ_1 and ρ_2 . We may compute the probability, $P(1|\text{hit})$, that a particle hitting the counter is from beam 1. Using Bayes’ theorem:

$$P(1|\text{hit}) = \frac{P(\text{hit}|1)P(1)}{P(\text{hit})}$$

$$\begin{aligned}
&= \frac{\langle D \rangle_{\rho_1} \theta}{\theta \langle D \rangle_{\rho_1} + (1 - \theta) \langle D \rangle_{\rho_2}} \\
&= \theta \frac{\langle D \rangle_{\rho_1}}{\langle D \rangle_{\rho}}.
\end{aligned} \tag{33}$$

$$\tag{34}$$

The generalization to an incoherent superposition of an arbitrary number of ensembles is clear: Let ρ_1, ρ_2, \dots be a set of distinct statistical ensembles, and let $\theta_1, \theta_2, \dots$ be a set of real numbers such that

$$\theta_n > 0, \quad \text{and} \quad \sum_n \theta_n = 1. \tag{35}$$

The incoherent sum of these ensembles, with probabilities $\{\theta_n\}$ is denoted

$$\rho = \sum_n \theta_n \rho_n. \tag{36}$$

This is to be interpreted as meaning that, for *every* dichotomic variable D :

$$\langle D \rangle_{\rho} = \sum_n \theta_n \langle D \rangle_{\rho_n}. \tag{37}$$

A particular prepared system is regarded as an **element** of the statistical ensemble. We have the intuitive notion that our level of information about an element from an ensemble $\rho = \theta \rho_1 + (1 - \theta) \rho_2$, which is an incoherent superposition of distinct ensembles ρ_1 and ρ_2 , is less than our information about an element in either ρ_1 or ρ_2 . For example, consider D a dichotomic variable such that $\langle D \rangle_{\rho_1} \neq \langle D \rangle_{\rho_2}$. Such a variable must exist, since $\rho_1 \neq \rho_2$. We have:

$$\langle D \rangle_{\rho} = \theta \langle D \rangle_{\rho_1} + (1 - \theta) \langle D \rangle_{\rho_2}. \tag{38}$$

Consider

$$\langle D \rangle_{\rho} - \frac{1}{2} = \theta \left(\langle D \rangle_{\rho_1} - \frac{1}{2} \right) + (1 - \theta) \left(\langle D \rangle_{\rho_2} - \frac{1}{2} \right). \tag{39}$$

We find:

$$\begin{aligned}
\left| \langle D \rangle_{\rho} - \frac{1}{2} \right| &\leq \theta \left| \langle D \rangle_{\rho_1} - \frac{1}{2} \right| + (1 - \theta) \left| \langle D \rangle_{\rho_2} - \frac{1}{2} \right| \\
&< \max \left(\left| \langle D \rangle_{\rho_1} - \frac{1}{2} \right|, \left| \langle D \rangle_{\rho_2} - \frac{1}{2} \right| \right).
\end{aligned} \tag{40}$$

What does this result tell us? The quantity $|\langle D \rangle_\rho - \frac{1}{2}| \in [0, \frac{1}{2}]$ can be regarded as a measure of the information we have about variable D for ensemble ρ . For example, if $|\langle D \rangle_\rho - \frac{1}{2}| = \frac{1}{2}$, then $\langle D \rangle_\rho = 1$ or 0 , and D is precisely known for ensemble ρ . On the other hand, if $|\langle D \rangle_\rho - \frac{1}{2}| = 0$, then $\langle D \rangle_\rho = 1/2$, and each of the possibilities $D = 0$ and $D = 1$ is equally likely, corresponding to maximal ignorance about D for ensemble ρ . Thus, our inequality says that, for at least one of ρ_1 and ρ_2 , we know more about D than for the incoherent superposition ρ .

We may restate our definition of pure and mixed states:

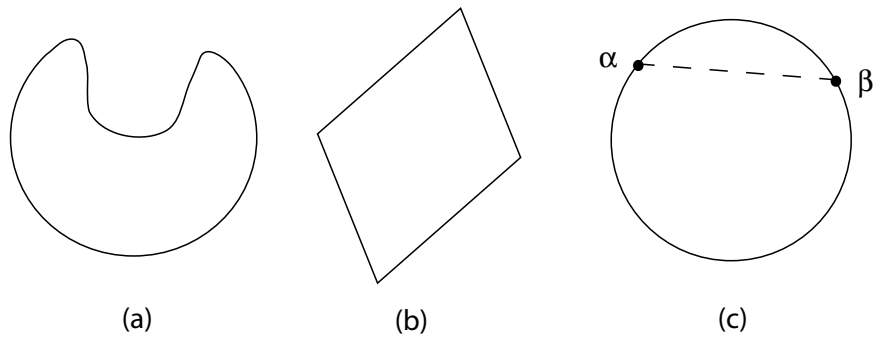
Def: A **pure ensemble** (or **pure state**) is an ensemble which is not an incoherent superposition of any other two distinct ensembles. A **mixed ensemble** (or **mixed state**) is an ensemble which is not pure.

Intuitively, a pure ensemble is a more carefully prepared ensemble – we have more (in fact, maximal) information about the elements – than a mixed ensemble.

The set of all physical statistical ensembles is a convex set,¹ with an incoherent superposition of two ensembles a convex combination of two elements of the convex set. Pure states are the extreme points of the set – *i.e.*, points which are not convex combinations of other points.

So far, this discussion has been rather general, and we have not made any quantum mechanical assumptions. In fact, let us think about classical

¹Convex set: A subset $K \subset C^n$ of n -dimensional complex Euclidean space is **convex** if, given any two points $\alpha, \beta \in K$, the straight line segment joining α and β is entirely contained in K :



(a) Not a convex set. (b) A convex set. (c) A convex set: Any **convex combination** of α, β , $x = \theta\alpha + (1 - \theta)\beta$, where $0 < \theta < 1$ is an element of the set.

mechanics first. In classical physics, the pure states correspond to a complete absence of any statistical spread in the dichotomic variables. If a preparation yields a pure state, then a repeated measurement of any variable will always yield the same result, either 0 or 1. Experimentally, this does not seem to be the case. Instead, no matter how carefully we prepare our ensemble, there will always be at least one dichotomic variable D such that $\langle D \rangle = 1/2$, corresponding to maximal statistical spread. Quantum mechanics (ignoring now issues of superselection rules) also deals with the nature of dichotomic variables and the set of ensembles, in a way which agrees so far with experiment. Let us restate some earlier postulates of quantum mechanics, modified and expanded in this context:

1. To every physical system we associate a Hilbert space H . The pure ensembles of the system are in 1:1 correspondence with the set of all one-dimensional projections in H . Such a projection, P , is an operator on the Hilbert space satisfying (A):

$$(A) \quad \begin{cases} P^2 = P & \text{idempotent,} \\ P^\dagger = P & \text{Hermitian,} \\ \text{Tr}(P) = 1 & \text{“primitive”, or one-dimensional.} \end{cases} \quad (41)$$

The set of all such projections is in one-to-one correspondence with the set of all rays² in H . Alternatively, we say that there is a one-to-one correspondence between the rays and the pure states.

Given any ray R , we can pick a unit vector $\phi \in R$, and the idempotent P associated with R is

$$(B) \quad P = |\phi\rangle\langle\phi|. \quad (42)$$

Conversely, any idempotent with the properties (A) can also be written in the form (B).

Proof: We assume (see Exercises) that it has been demonstrated that any linear operator in an n -dimensional Euclidean space may be expressed as an n -term dyad, and that the extension of this idea to an infinite-dimensional separable space has been made. Hence, we may write:

$$P = \sum_i |a_i\rangle\langle b_i|. \quad (43)$$

²A **ray** is the set of all non-zero multiples of a given non-zero vector. Such a multiple is called an **element** of the ray.

Note that in some orthonormal basis $\{|e_i\rangle\}$, the matrix elements of P are $P_{ij} = \langle e_i|P|e_j\rangle$, and hence,

$$P = \sum_{i,j} |e_i\rangle P_{ij} \langle e_j|. \quad (44)$$

In the present case, P is Hermitian and therefore diagonalizable. Let $\{|e_i\rangle\}$ be a basis in which P is diagonal:

$$P = \sum_i |e_i\rangle P_{ii} \langle e_i|. \quad (45)$$

Since $P^\dagger = P$, the P_{ii} are all real. Calculate:

$$\begin{aligned} P^2 &= \sum_{i,j} |e_i\rangle P_{ii} \langle e_i|e_j\rangle P_{jj} \langle e_j| \\ &= \sum_i |e_i\rangle P_{ii}^2 \langle e_i| \\ &= P, \end{aligned} \quad (46)$$

where the latter equality can be true if and only if $P_{ii}^2 = P_{ii}$ for all i . That is, for each i we must either have $P_{ii} = 1$ or $P_{ii} = 0$. But we must also have $\text{Tr}(P) = \sum_i P_{ii} = 1$, which holds if exactly one $P_{ii} \neq 0$, say P_{aa} . In this basis,

$$P = |e_a\rangle \langle e_a| \quad (47)$$

The ray R associated with P is then $\{c|e_a\rangle; c \neq 0\}$.

2. To every dichotomic variable D there corresponds a projection on some subspace of H . That is, such a variable is represented by an *operator* D on H satisfying:

$$D^\dagger = D \quad (48)$$

$$D^2 = D \neq 0, \quad (49)$$

the latter since the eigenvalues of D are 0 and 1.

3. The average of D in pure ensemble P (corresponding to projection P) is:

$$\langle D \rangle_P = \text{Tr}(DP) \quad (50)$$

(if $P = |\phi\rangle \langle \phi|$, then $\langle D \rangle_P = \langle \phi|D|\phi\rangle$).

4. An arbitrary ensemble ρ is represented by a **statistical operator**, or **density matrix**, which we also denote by symbol ρ . This is a Hermitian operator on H with spectral decomposition,

$$\rho = \sum_i r_i P_i, \quad (51)$$

where

$$P_i P_j = \delta_{ij} \quad (52)$$

$$\sum_i P_i = I \quad (53)$$

$$r_i \geq 0 \quad (54)$$

$$\sum_i r_i = 1. \quad (55)$$

The set $\{r_i\}$ is the set of eigenvalues of the operator ρ . The properties of this density matrix are precisely as in our earlier discussion.

Our symbolic equation for the incoherent superposition of two ensembles, $\rho = \theta\rho_1 + (1 - \theta)\rho_2$, can be interpreted as an equation for the corresponding density matrices represented by the same symbols. Hence, the density matrix ρ describing the superposition of ρ_1 and ρ_2 with probabilities θ and $1 - \theta$ is $\rho = \theta\rho_1 + (1 - \theta)\rho_2$. Thus, if ρ is any density matrix, and D any dichotomic variable, then:

$$\langle D \rangle_\rho = \text{Tr}(D\rho). \quad (56)$$

For example,

$$\begin{aligned} \langle D \rangle_\rho &= \langle D \rangle_{\theta\rho_1 + (1-\theta)\rho_2} \\ &= \theta\langle D \rangle_{\rho_1} + (1 - \theta)\langle D \rangle_{\rho_2} \\ &= \text{Tr}(D\theta\rho_1) + \text{Tr}[D(1 - \theta)\rho_2] \\ &= \text{Tr}\{D[\theta\rho_1 + (1 - \theta)\rho_2]\} \\ &= \text{Tr}(D\rho) \end{aligned} \quad (57)$$

5. We regard every projection as corresponding to an observable, *i.e.*, every primitive Hermitian idempotent P corresponds to an observable. If ρ is a density matrix, then

$$\rho = P \Leftrightarrow \text{Tr}(P\rho) = 1. \quad (58)$$

Proof: Suppose $\rho = P$. Then $\text{Tr}(PP) = \text{Tr}(P)$, since $P^2 = P$. But $\text{Tr}P = 1$, since P is primitive. Now suppose $\text{Tr}(P\rho) = 1$. Then

$$\begin{aligned} 1 &= \text{Tr} \left(P \sum_i r_i P_i \right) \\ &= \sum_i r_i \text{Tr}(PP_i). \end{aligned} \quad (59)$$

Expand the one-dimensional projection operator in the basis in which $P_i = |e_i\rangle\langle e_i|$:

$$P = \sum_{j,k} |e_j\rangle\langle e_j|P|e_k\rangle\langle e_k|. \quad (60)$$

Then:

$$\begin{aligned} 1 &= \sum_i r_i \text{Tr} \left(\sum_{j,k} |e_j\rangle\langle e_j|P|e_k\rangle\langle e_k|e_i\rangle\langle e_i| \right) \\ &= \sum_i r_i \sum_j \langle e_j|P|e_i\rangle \text{Tr}(|e_j\rangle\langle e_i|) \\ &= \sum_i r_i \langle e_i|P|e_i\rangle. \end{aligned} \quad (61)$$

But we also have $\sum_i \langle e_i|P|e_i\rangle = 1$ and $\sum_i r_i = 1$, with $0 \leq r_i \leq 1$. Thus, $\sum_i r_i \langle e_i|P|e_i\rangle < 1$, unless there is a k such that $r_k = 1$, and all of the other $r_i = 0$, $i \neq k$. Hence, $\langle e_k|P|e_k\rangle = 1$, or $P = |e_k\rangle\langle e_k| = \rho$.

Thus, P is the observable which tests whether an element of the statistical ensemble is in the state corresponding to ray “ P ”.

6. In addition to the projection operators, we regard general self-adjoint operators as observables, and the laws of nature deal with these observables. For example, we may consider operators with spectral resolutions of the form:

$$Q = \sum_i q_i P_i = \sum_i q_i |e_i\rangle\langle e_i|, \quad (62)$$

where $P_i P_j = \delta_{ij} P_i$, and where the eigenvalues q_i are real. We may regard this as expressing the physical variable Q in terms of the dichotomic variables P_i (noting that the eigenvalues of P_i are 0 and 1).

Hence it is natural to define the ensemble average of Q in an ensemble ρ by:

$$\begin{aligned}\langle Q \rangle_\rho &= \langle \sum_i q_i P_i \rangle_\rho \\ &= \sum_i q_i \text{Tr}(\rho P_i) \\ &= \text{Tr}(\rho Q).\end{aligned}\tag{63}$$

This completes our picture of the mathematical structure and postulates of quantum mechanics in this somewhat new language. We see that we need not discuss “state vectors” in quantum mechanics, we can talk about “ensembles” instead. In fact, the latter description has a more “physical” aspect, in the sense that experimentally we seem to be able to prepare systems as statistical ensembles, but not so readily as pure states.

Of course, we have no proof that our experimental ensembles and dichotomic variables must obey the above postulates. It may be that there is some other theory which is more correct. However, there is so far no experimental conflict with our orthodox theory, and we shall continue in this vein.

5 Coherent Superpositions

Theorem: Let P_1, P_2 be two primitive Hermitian idempotents (*i.e.*, rays, or pure states, with $P^\dagger = P$, $P^2 = P$, and $\text{Tr}P = 1$). Then:

$$1 \geq \text{Tr}(P_1 P_2) \geq 0.\tag{64}$$

If $\text{Tr}(P_1 P_2) = 1$, then $P_2 = P_1$. If $\text{Tr}(P_1 P_2) = 0$, then $P_1 P_2 = 0$ (vectors in ray 1 are orthogonal to vectors in ray 2).

More generally, if ρ is a density matrix, and Q is any projection, then

$$1 \geq \text{Tr}(Q\rho) \geq 0,\tag{65}$$

$$\text{Tr}(Q\rho) = 1 \quad \Leftrightarrow \quad Q\rho = \rho Q = \rho,\tag{66}$$

$$\text{Tr}(Q\rho) = 0 \quad \Leftrightarrow \quad Q\rho = 0.\tag{67}$$

Suppose we have orthogonal pure states, $P_1 P_2 = 0$. There then exists a unique two parameter family of pure states $\{P\}$ such that

$$\text{Tr}(PP_1) + \text{Tr}(PP_2) = 1.\tag{68}$$

Any member P of this family is a ray corresponding to any vector in the two-dimensional subspace defined by the projection $P_1 + P_2 = S$. We say that P is a **coherent superposition** of the pure states P_1 and P_2 .

Let's give an explicit construction of the operators P : Pick unit vector $|e_1\rangle$ from ray P_1 and $|e_2\rangle$ from ray P_2 . Construct the following four operators:

$$S = P_1 + P_2 = |e_1\rangle\langle e_1| + |e_2\rangle\langle e_2| \quad (69)$$

$$\sigma_1 = |e_1\rangle\langle e_2| + |e_2\rangle\langle e_1| \quad (70)$$

$$\sigma_2 = i(|e_2\rangle\langle e_1| - |e_1\rangle\langle e_2|) \quad (71)$$

$$\sigma_3 = |e_1\rangle\langle e_1| - |e_2\rangle\langle e_2|. \quad (72)$$

These operators satisfy the algebraic relations (noting the obvious similarities with the Pauli matrices):

$$S^2 = S \quad (73)$$

$$S\sigma_i = \sigma_i \quad (74)$$

$$\sigma_i^2 = S \quad (75)$$

$$[\sigma_i, \sigma_j] = i\epsilon_{ijk}\sigma_k. \quad (76)$$

Let $\mathbf{u} = (u_1, u_2, u_3)$ be a unit vector in three-dimensional Euclidean space. Define

$$P(\mathbf{u}) \equiv \frac{1}{2}(S + \mathbf{u} \cdot \boldsymbol{\sigma}). \quad (77)$$

The reader should demonstrate that $P(\mathbf{u})$ is the most general coherent superposition of pure states P_1 and P_2 . This set is parameterized by the two-parameter unit vector \mathbf{u} . This, of course, is very characteristic of quantum mechanics: If we have a “two-state” system we may form arbitrary superpositions $|\psi\rangle = \alpha|\psi_1\rangle + \beta|\psi_2\rangle$ (assume $\langle\psi_1|\psi_2\rangle = 0$). The overall phase is arbitrary, and the normalization constraint $|\alpha|^2 + |\beta|^2 = 1$ uses another degree of freedom, hence two parameters are required to describe an arbitrary state. Note that the coherent superposition of pure states is itself a pure state, unlike an incoherent superposition.

6 Density Matrices in a Finite-Dimensional Hilbert Space

Consider a finite-dimensional Hilbert space H . The set of Hermitian operators on H defines a real vector space (real, so that aQ is Hermitian if Q is

Hermitian). Call this vector space \mathcal{O} (for vector space of Operators). Define a positive definite $[(X, X) > 0 \text{ unless } X = 0]$ symmetric $[(X, Y) = (Y, X)]$ scalar product on \mathcal{O} by:

$$(X, Y) \equiv \text{Tr}(XY), \quad (78)$$

for any two vectors (*i.e.*, Hermitian operators) $X, Y \in \mathcal{O}$. The set of all density matrices forms a convex subset of \mathcal{O} , with norm ≤ 1 .

Consider a complete orthonormal basis in \mathcal{O} :

$$\{B\} = \{B_1, B_2, \dots\} \subset \mathcal{O} \text{ such that } \text{Tr}(B_i B_j) = \delta_{ij}. \quad (79)$$

Expand any vector $X \in \mathcal{O}$ in this basis according to

$$X = \sum_i B_i \text{Tr}(B_i X). \quad (80)$$

For a density matrix ρ this expansion is

$$\rho = \sum_i B_i \text{Tr}(B_i \rho), \quad (81)$$

but, as we have seen before, $\text{Tr}(B_i \rho) = \langle B_i \rangle_\rho$ is just the ensemble average of observable B_i in the ensemble ρ . Hence, the density matrix may be determined through measurements, uniquely, if we measure the ensemble averages of a complete set of operators.

7 Entropy, Mixing, Correlations

For this discussion, we need to first define the concept of a function of an operator. Consider a self-adjoint operator Q , with a pure point spectrum consisting of (real) eigenvalues $\{q_i; i = 1, 2, \dots\}$ and no finite point of accumulation.³ Let $|k\rangle$ denote the eigenvector corresponding to eigenvalue q_k , and assume it has been normalized. Then $\{|k\rangle\}$ forms a complete orthonormal set, *i.e.*:

$$\langle k|j\rangle = \delta_{kj}; \quad I = \sum_k |k\rangle\langle k|. \quad (82)$$

³Abstractly, a **point of accumulation** (or a **limit point**) is a point $x \in S \subset T$, where T is a topological space, if every neighborhood $N(x)$ contains a point of S distinct from x .

The **spectral resolution** of Q is given by:

$$Q = \sum_k q_k |k\rangle\langle k|. \quad (83)$$

Let $\Sigma(Q)$ denote the spectrum $\{q\}$ of Q . If $f(q)$ is any function defined on $\Sigma(Q)$, we define the operator $f(Q)$ by:

$$f(Q) \equiv \sum_k f(q_k) |k\rangle\langle k|. \quad (84)$$

For example,

$$Q^2 = \sum_k q_k^2 |k\rangle\langle k|, \quad (85)$$

which may be compared with

$$\begin{aligned} Q^2 &= \sum_{k,j} q_k q_j |k\rangle\langle k|j\rangle\langle j| \\ &= \sum_k q_k^2 |k\rangle\langle k|, \end{aligned} \quad (86)$$

which is what we hope should happen. In particular, we may perform Taylor series expansions of functions of operators.

We wish to define a measure of the amount of (or lack of) information concerning the elements of a statistical ensemble ρ . Thus, define the **entropy** $s = s(\rho)$ by:

$$s \equiv -\text{Tr}(\rho \ln \rho) \quad (= -\langle \ln \rho \rangle_\rho). \quad (87)$$

Note that, with an expansion (spectral decomposition) of ρ according to

$$\rho = \sum_i r_i P_i = \sum_i r_i |e_i\rangle\langle e_i|, \quad (88)$$

we have

$$\ln \rho = \sum_i (\ln r_i) P_i, \quad (89)$$

and hence

$$\begin{aligned} s &= -\text{Tr} \left[\sum_i (\ln r_i) \rho P_i \right] \\ &= -\sum_i \ln r_i \text{Tr}(\rho P_i) \\ &= -\sum_i \ln r_i \text{Tr}(\sum_j r_j P_j P_i) \\ &= -\sum_i r_i \ln r_i. \end{aligned} \quad (90)$$

Since $0 \leq r_i \leq 1$, we always have $s \geq 0$, and also $s = 0$ if and only if the ensemble is a pure state. Roughly speaking, the more non-zero r_i 's there are, that is the more the number of pure states involved, the greater the entropy.

Consistent with our classical thermodynamic notion that entropy increases with “mixing”, we have the “von Neumann mixing theorem”:

Theorem: If $0 < \theta < 1$, and $\rho_1 \neq \rho_2$, then:

$$s[\theta\rho_1 + (1 - \theta)\rho_2] > \theta s(\rho_1) + (1 - \theta)s(\rho_2). \quad (91)$$

8 Combination of Systems

Consider the situation where the system of interest may be regarded as the “combination” of two subsystems, 1 and 2. For example, perhaps the system consists of two atoms. For simplicity of illustration, assume that the states of system 1 alone form a finite-dimensional Hilbert space H_1 , and the states of system 2 alone form another finite-dimensional Hilbert space H_2 . The combined system is then associated with Hilbert space $H = H_1 \otimes H_2$. For example, we may have a two-dimensional space H_1 and a three-dimensional space H_2 , with sets of vectors:

$$\left\{ \begin{pmatrix} a \\ b \end{pmatrix} \right\} \quad \text{and} \quad \left\{ \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \right\}, \quad (92)$$

respectively. Then the product space consists of direct product vectors of the form:

$$\begin{pmatrix} a\alpha \\ b\alpha \\ a\beta \\ b\beta \\ a\gamma \\ b\gamma \end{pmatrix}. \quad (93)$$

The operators on H which refer only to subsystem 1 are of the form $X \otimes I$, and the operators on H which refer only to subsystem 2 are of the form $I \otimes Y$

(X is an operator on H_1 and Y is an operator on H_2). For example:

$$X \otimes I = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & 0 & 0 \\ x_3 & x_4 & 0 & 0 \\ 0 & x_1 & x_2 & 0 \\ 0 & x_3 & x_4 & 0 \\ 0 & 0 & x_1 & x_2 \\ 0 & 0 & x_3 & x_4 \end{pmatrix}. \quad (94)$$

We see that this operator does not mix up the components α, β, γ vectors in H_2 .

Consider now an operator on H of the special form $Z = X \otimes Y$. Define “partial traces” for such an operator according to the mappings:

$$\text{Tr}_1(Z) = \text{Tr}_1(X \otimes Y) \equiv Y \text{Tr}(X) \quad (95)$$

$$\text{Tr}_2(Z) = \text{Tr}_2(X \otimes Y) \equiv X \text{Tr}(Y) \quad (96)$$

For our example:

$$Z = X \otimes Y = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} \otimes \begin{pmatrix} y_1 & y_2 & y_3 \\ y_4 & y_5 & y_6 \\ y_7 & y_8 & y_9 \end{pmatrix} \quad (97)$$

$$= \begin{pmatrix} x_1 y_1 & x_2 y_1 & x_1 y_2 & x_2 y_2 & x_1 y_3 & x_2 y_3 \\ x_3 y_1 & x_4 y_1 & x_3 y_2 & x_4 y_2 & x_3 y_3 & x_4 y_3 \\ x_1 y_4 & x_2 y_4 & x_1 y_5 & x_2 y_5 & x_1 y_6 & x_2 y_6 \\ x_3 y_4 & x_4 y_4 & x_3 y_5 & x_4 y_5 & x_3 y_6 & x_4 y_6 \\ x_1 y_7 & x_2 y_7 & x_1 y_8 & x_2 y_8 & x_1 y_9 & x_2 y_9 \\ x_3 y_7 & x_4 y_7 & x_3 y_8 & x_4 y_8 & x_3 y_9 & x_4 y_9 \end{pmatrix}, \quad (98)$$

and thus, for example,

$$\text{Tr}_1(Z) = (x_1 + x_4) \begin{pmatrix} y_1 & y_2 & y_3 \\ y_4 & y_5 & y_6 \\ y_7 & y_8 & y_9 \end{pmatrix}, \quad (99)$$

and also

$$\text{Tr}[\text{Tr}_1(Z)] = (x_1 + x_4)(y_1 + y_5 + y_9) = \text{Tr}(Z). \quad (100)$$

These mappings thus map operators on H of this form into operators on H_1 or on H_2 .

An arbitrary linear operator on H may be expressed as a linear combination of operators of this form, and we extend the definition of Tr_1 and Tr_2 by

linearity to all operators on H . For example, suppose $Z = X_1 \otimes Y_1 + X_2 \otimes Y_2$. Then

$$\begin{aligned}\mathrm{Tr}_1(Z) &= \mathrm{Tr}_1(X_1 \otimes Y_1 + X_2 \otimes Y_2) \\ &= \mathrm{Tr}_1(X_1 \otimes Y_1) + \mathrm{Tr}_1(X_2 \otimes Y_2) \\ &= Y_1 \mathrm{Tr}(X_1) + Y_2 \mathrm{Tr}(X_2),\end{aligned}\tag{101}$$

and the result is an operator on H_2 .

Now let ρ be a density matrix on H , describing a statistical ensemble of the combined system. Define “reduced density matrices” for subsystems 1 and 2:

$$\rho_1 \equiv \mathrm{Tr}_2(\rho), \quad \rho_2 \equiv \mathrm{Tr}_1(\rho).\tag{102}$$

The interpretation is that ρ_1 summarizes all of the information contained in ρ about the variables of subsystem 1 *alone*, and similarly for ρ_2 . For example, if X is any operator on system 1 alone:

$$\begin{aligned}\langle X \rangle_\rho &= \mathrm{Tr}[\rho(X \otimes I)] \\ &= \langle X \rangle_{\rho_1} = \mathrm{Tr}(X \rho_1).\end{aligned}\tag{103}$$

From the reduced density matrices ρ_1 and ρ_2 we can form a new density matrix on H :

$$\rho_{12} = \rho_1 \otimes \rho_2.\tag{104}$$

It contains the same information which ρ_1 and ρ_2 contain together — ρ_{12} describes a statistical ensemble for which the variables of subsystem 1 are completely **uncorrelated** with the variables of subsystem 2. If ρ is not of this form ($\rho \neq \rho_{12}$), then ρ describes an ensemble for which there is some correlation between the variables of the two subsystems.

For the entropy in particular, we have

$$s(\rho_{12}) = s(\rho_1 \otimes \rho_2) = s(\rho_1) + s(\rho_2).\tag{105}$$

Proof: We can choose a basis in which ρ_1 and ρ_2 are diagonal, and in this basis $\rho_{12} = \rho_1 \otimes \rho_2$ is also diagonal. Denote the diagonal elements of ρ_1 as d_i , *i.e.*, $d_i \equiv (\rho_1)_{ii}$, and the diagonal elements of ρ_2 as δ_j . Then the diagonal elements of ρ_{12} are given by all products of the form $d_i \delta_j$, where $i = 1, 2, \dots, n_1$, and $j = 1, 2, \dots, n_2$, and where n_1 and n_2 are

the dimensions of H_1 and H_2 , respectively. Thus,

$$\begin{aligned} s(\rho_{12}) &= -\text{Tr}(\rho_{12} \ln \rho_{12}) \\ &= -\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (d_i \delta_j) \ln(d_i \delta_j). \end{aligned} \quad (106)$$

We compare this with (noting that $\text{Tr} \rho_1 = \text{Tr} \rho_2 = 1$):

$$\begin{aligned} s(\rho_1) + s(\rho_2) &= -\left(\sum_{i=1}^{n_1} d_i \ln d_i + \sum_{j=1}^{n_2} \delta_j \ln \delta_j \right) \\ &= -\left(\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \delta_j d_i \ln d_i + \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} d_i \delta_j \ln \delta_j \right) \\ &= -\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} d_i \delta_j (\ln d_i + \ln \delta_j) \\ &= s(\rho_{12}). \end{aligned} \quad (107)$$

Thus, the entropy for an ensemble (ρ_{12}) for which the subsystems are uncorrelated is just equal to the sum of the entropies of the reduced ensembles for the subsystems. When there are correlations, we should expect an inequality instead, since in this case ρ contains additional information concerning the correlations, which is not present in ρ_1 and ρ_2 ($\rho_{12} = \rho_1 \otimes \rho_2 \neq \rho$). Then:

$$s(\rho_{12}) = s(\rho_1) + s(\rho_2) \geq s(\rho), \quad (108)$$

where equality holds if and only if $\rho = \rho_{12}$, that is, if there are no correlations.

It is interesting that this inequality is specific for $-x \ln x$, in the following sense: Let $s(\rho) = \text{Tr}[f(\rho)]$. If this inequality, including the condition for equality, holds for all finite-dimensional Hilbert spaces H_1 and H_2 , and all density matrices ρ on $H = H_1 \otimes H_2$, then $f(x) = -kx \ln x$, where $k > 0$ (and we may take $k = 1$). Since this inequality appears to be determined by physical considerations, this becomes a strong argument for the form $s(\rho) = -\text{Tr}(\rho \ln \rho)$ for the entropy.

9 Some Statistical Mechanics

Consider a Hamiltonian H with point spectrum $\{\omega_i; i = 1, 2, \dots\}$, bounded below. The **partition function**, $Z(T)$, for temperature $T > 0$ is defined by:

$$Z(T) \equiv \sum_{k=1}^{\infty} e^{-\omega_k/T}. \quad (109)$$

We are assuming that this sum converges. The **density matrix** (or **statistical operator**) for the **canonical distribution** is given by:

$$\rho = \rho(T) = \frac{e^{-H/T}}{Z(T)} \quad (110)$$

$$= \frac{1}{Z(T)} \sum_{k=1}^{\infty} |k\rangle\langle k| e^{-\omega_k/T}. \quad (111)$$

This makes intuitive sense – our canonical, thermodynamic distribution consists of a mixture of states, with each state receiving a “weight” of $\exp(-\omega_k/T)$. Note that

$$Z(T) = \sum_{k=1}^{\infty} e^{-\omega_k/T} = \text{Tr} \left(\sum_{k=1}^{\infty} |k\rangle\langle k| e^{-\omega_k/T} \right)$$

$$= \text{Tr} \left(e^{-H/T} \right). \quad (112)$$

$$(113)$$

Hence, $\text{Tr} [\rho(T)] = 1$.

The ensemble average of any observable (self-adjoint operator), Q , in the canonical ensemble is:

$$\langle Q \rangle_{\rho} = \text{Tr} [Q\rho(T)]. \quad (114)$$

For example, the mean energy is:

$$\begin{aligned} U &= \langle H \rangle_{\rho} \\ &= \frac{1}{Z(T)} \text{Tr} \left(H e^{-H/T} \right) \\ &= \frac{T^2}{Z(T)} \partial_T \text{Tr} \left(e^{-H/T} \right) \\ &= \frac{T^2}{Z(T)} \partial_T \text{Tr} Z(T) \end{aligned} \quad (115)$$

$$= T^2 \partial_T \ln [Z(T)]. \quad (116)$$

The entropy is:

$$\begin{aligned}
 S &= -\text{Tr}(\rho \ln \rho) \\
 &= -\text{Tr} \left\{ \frac{e^{-H/T}}{Z(T)} \left[-\frac{H}{T} - \ln Z(T) \right] \right\} \\
 &= \frac{U}{T} + \ln [Z(T)].
 \end{aligned} \tag{117}$$

If we define the **Helmholtz free energy**, $F = -T \ln Z$, then $S = -\partial_T F$. Alternatively, $U = TS + F$.

10 Exercises

1. Show that any linear operator in an n -dimensional Euclidean space may be expressed as an n -term dyad. Show that this may be extended to an infinite-dimensional Euclidean space.
2. Suppose we have a system with total angular momentum 1. Pick a basis corresponding to the three eigenvectors of the z -component of angular momentum, J_z , with eigenvalues $+1, 0, -1$, respectively. We are given an ensemble described by density matrix:

$$\rho = \frac{1}{4} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

- (a) Is ρ a permissible density matrix? Give your reasoning. For the remainder of this problem, assume that it is permissible. Does it describe a pure or mixed state? Give your reasoning.
 - (b) Given the ensemble described by ρ , what is the average value of J_z ?
 - (c) What is the spread (standard deviation) in measured values of J_z ?
3. Prove the first theorem in section 5.
 4. Prove the von Neumann mixing theorem.
 5. Show that an arbitrary linear operator on a product space $H = H_1 \otimes H_2$ may be expressed as a linear combination of operators of the form $Z = X \otimes Y$.

6. Let us try to improve our understanding of the discussions on the density matrix formalism, and the connections with “information” or “entropy” that we have made. Thus, we consider a simple “two-state” system. Let ρ be any general density matrix operating on the two-dimensional Hilbert space of this system.

- (a) Calculate the entropy, $s = -\text{Tr}(\rho \ln \rho)$ corresponding to this density matrix. Express your result in terms of a single real parameter. Make sure the interpretation of this parameter is clear, as well as its range.
- (b) Make a graph of the entropy as a function of the parameter. What is the entropy for a pure state? Interpret your graph in terms of knowledge about a system taken from an ensemble with density matrix ρ .
- (c) Consider a system with ensemble ρ a mixture of two ensembles ρ_1, ρ_2 :

$$\rho = \theta \rho_1 + (1 - \theta) \rho_2, \quad 0 \leq \theta \leq 1 \quad (118)$$

As an example, suppose

$$\rho_1 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad \rho_2 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad (119)$$

in some basis. Prove that VonNeuman’s mixing theorem holds for this example:

$$s(\rho) \geq \theta s(\rho_1) + (1 - \theta) s(\rho_2), \quad (120)$$

with equality iff $\theta = 0$ or $\theta = 1$.

7. Consider an N -dimensional Hilbert space. We define the real vector space, \mathcal{O} of Hermitian operators on this Hilbert space. We define a scalar product on this vector space according to:

$$(x, y) = \text{Tr}(xy), \quad \forall x, y \in \mathcal{O}. \quad (121)$$

Consider a basis $\{B\}$ of orthonormal operators in \mathcal{O} . The set of density operators is a subset of this vector space, and we may expand an arbitrary density matrix as:

$$\rho = \sum_i B_i \text{Tr}(B_i \rho) = \sum_i B_i \langle B_i \rangle_\rho. \quad (122)$$

By measuring the average values for the basis operators, we can thus determine the expansion coefficients for ρ .

- (a) How many such measurements are required to completely determine ρ ?
 - (b) If ρ is known to be a pure state, how many measurements are required?
8. Two scientists (they happen to be twins, named “Oivil” and “Livio”, but never mind...) decide to do the following experiment: They set up a light source, which emits two photons at a time, back-to-back in the laboratory frame. The ensemble is given by:

$$\rho = \frac{1}{2}(|LL\rangle\langle LL| + |RR\rangle\langle RR|), \quad (123)$$

where “ L ” refers to left-handed polarization, and “ R ” refers to right-handed polarization. Thus, $|LR\rangle$ would refer to a state in which photon number 1 (defined as the photon which is aimed at scientist Oivil, say) is left-handed, and photon number 2 (the photon aimed at scientist Livio) is right-handed.

These scientists (one of whom is of a diabolical bent) decide to play a game with Nature: Oivil (of course) stays in the lab, while Livio treks to a point a light-year away. The light source is turned on and emits two photons, one directed toward each scientist. Oivil soon measures the polarization of his photon; it is left-handed. He quickly makes a note that his brother is going to see a left-handed photon, sometime after next Christmas.

Christmas has come and gone, and finally Livio sees his photon, and measures its polarization. He sends a message back to his brother Oivil, who learns in yet another year what he knew all along: Livio’s photon was left-handed.

Oivil then has a sneaky idea. He secretly changes the apparatus, without telling his forlorn brother. Now the ensemble is:

$$\rho = \frac{1}{2}(|LL\rangle + |RR\rangle)(\langle LL| + \langle RR|). \quad (124)$$

He causes another pair of photons to be emitted with this new apparatus, and repeats the experiment. The result is identical to the first experiment.

- (a) Was Oivil just lucky, or will he get the right answer every time, for each apparatus? Demonstrate your answer explicitly, in the density matrix formalism.
- (b) What is the probability that Livio will observe a left-handed photon, or a right-handed photon, for each apparatus? Is there a problem with causality here? How can Oivil know what Livio is going to see, long before he sees it? Discuss! Feel free to modify the experiment to illustrate any points you wish to make.
9. Let us consider the application of the density matrix formalism to the problem of a spin-1/2 particle (such as an electron) in a static external magnetic field. In general, a particle with spin may carry a magnetic moment, oriented along the spin direction (by symmetry). For spin-1/2, we have that the magnetic moment (operator) is thus of the form:

$$\boldsymbol{\mu} = \frac{1}{2}\gamma\boldsymbol{\sigma}, \quad (125)$$

where $\boldsymbol{\sigma}$ are the Pauli matrices, the $\frac{1}{2}$ is by convention, and γ is a constant, giving the strength of the moment, called the gyromagnetic ratio. The term in the Hamiltonian for such a magnetic moment in an external magnetic field, \mathbf{B} is just:

$$H = -\boldsymbol{\mu} \cdot \mathbf{B}. \quad (126)$$

Our spin-1/2 particle may have some spin-orientation, or “polarization vector”, given by:

$$\mathbf{P} = \langle \boldsymbol{\sigma} \rangle. \quad (127)$$

Drawing from our classical intuition, we might expect that in the external magnetic field the polarization vector will exhibit a precession about the field direction. Let us investigate this.

Recall that the expectation value of an operator may be computed from the density matrix according to:

$$\langle A \rangle = \text{Tr}(\rho A). \quad (128)$$

Furthermore, recall that the time evolution of the density matrix is given by:

$$i\frac{\partial \rho}{\partial t} = [H(t), \rho(t)]. \quad (129)$$

What is the time evolution, $d\mathbf{P}/dt$, of the polarization vector? Express your answer as simply as you can (more credit will be given for right answers that are more physically transparent than for right answers which are not). Note that we make no assumption concerning the purity of the state.

10. Let us consider a system of N spin-1/2 particles (see the previous problem) per unit volume in thermal equilibrium, in our external magnetic field \mathbf{B} . Recall that the canonical distribution is:

$$\rho = \frac{e^{-H/T}}{Z}, \quad (130)$$

with partition function:

$$Z = \text{Tr} \left(e^{-H/T} \right). \quad (131)$$

Such a system of particles will tend to orient along the magnetic field, resulting in a bulk magnetization (having units of magnetic moment per unit volume), \mathbf{M} .

- (a) Give an expression for this magnetization (don't work too hard to evaluate).
- (b) What is the magnetization in the high-temperature limit, to lowest non-trivial order (this I want you to evaluate as completely as you can!)?