Lecture 3: Density matrices

Recommended reading: Nielsen, chapter 2.4

Feynmann, chapter 2.

Mathematical warm-up: before we start with the physics, let us review the concept of trace of an operator

1.10 The Trace

The trace of an operator is defined as the sum of its diagonal entries:

$$\operatorname{tr}(A) = \sum_i \langle i|A|i\rangle. \tag{1.71}$$

It turns out that the trace is the same no matter which basis you use. You can see that using completeness: for instance, if $|a\rangle$ is some other basis then

$$\sum_i \langle i|A|i\rangle = \sum_i \sum_a \langle i|a\rangle\langle a|A|i\rangle = \sum_i \sum_a \langle a|A|i\rangle\langle i|a\rangle = \sum_a \langle a|A|a\rangle.$$

Thus, we conclude that

$$\operatorname{tr}(A) = \sum_i \langle i|A|i\rangle = \sum_a \langle a|A|a\rangle. \tag{1.72}$$

The trace is a property of the operator, not of the basis you choose. Since it does not matter which basis you use, let us choose the basis $|\lambda_i\rangle$ which diagonalizes the operator A. Then $\langle\lambda_i|A|\lambda_i\rangle = \lambda_i$ will be an eigenvalue of A. Thus, we also see that

$$\boxed{\operatorname{tr}(A) = \sum_i \lambda_i = \text{sum of all eigenvalues of } A}. \tag{1.73}$$

Perhaps the most useful property of the trace is that it is *cyclic*:

$$\operatorname{tr}(AB) = \operatorname{tr}(BA). \tag{1.74}$$

I will leave it for you to demonstrate this. Simply insert a convenient completeness relation in the middle of AB. Using the cyclic property (1.74) you can also move around an arbitrary number of operators, but only in cyclic permutations. For instance:

$$\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA). \tag{1.75}$$

Note how I am moving them around in a specific order: $\operatorname{tr}(ABC) \neq \operatorname{tr}(BAC)$. An example that appears often is a trace of the form $\operatorname{tr}(UAU^\dagger)$, where U is unitary operator. In this case, it follows from the cyclic property that

$$\operatorname{tr}(UAU^\dagger) = \operatorname{tr}(AU^\dagger U) = \operatorname{tr}(A)$$

Thus, the trace of an operator is invariant by unitary transformations. This is also in line with the fact that the trace is the sum of the eigenvalues and unitaries preserve eigenvalues.

Finally, let $|\psi\rangle$ and $|\phi\rangle$ be arbitrary kets and let us compute the trace of the outer product $|\psi\rangle\langle\phi|$:

$$\operatorname{tr}(|\psi\rangle\langle\phi|) = \sum_i \langle i|\psi\rangle\langle\phi|i\rangle = \sum_i \langle\phi|i\rangle\langle i|\psi\rangle$$

The sum over $|i\rangle$ becomes a 1 due to completeness and we conclude that

$$\boxed{\mathrm{tr}(|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle.} \tag{1.76}$$

Notice how this follows the same logic as Eq. (1.74), so you can pretend you just used the cyclic property. This formula turns out to be extremely useful, so it is definitely worth remembering.

Density Matrices and thermal states

Let us review a bit the Gibbs formalism. We consider a system described by a Hamiltonian H with eigenstuff

$$H = \sum_n E_n |n\rangle\langle n| \tag{1}$$

If this system is in equilibrium, then the probability to find it in state $|n\rangle$ will be given by

$$P_n = \frac{e^{-\beta E_n}}{Z} \tag{2}$$

Moreover, in equilibrium the expectation value of any observable A is given by

$$\langle A\rangle = \sum_n \langle n|A|n\rangle P_n \tag{3}$$

We may now ask the following question: is there a quantum state $|\psi\rangle$ which would allow us to write (3) as

$$\langle A\rangle = \langle\psi|A|\psi\rangle \qquad \text{(Not possible!)} \tag{4}$$

The answer is no. For instance, one could naively try a state like

$$|\psi\rangle = \sum_n \sqrt{P_n}\,|n\rangle \tag{5}$$

But then (4) gives

$$\langle\psi|A|\psi\rangle = \sum_{n,m} \langle n|A|m\rangle \sqrt{P_n P_m}$$

which is not like (3). It will coincide with (3) if A happens to be diagonal in the basis $|n\rangle$. But only in that case.

The moral of the story is that it is not possible to attribute a ket $|\psi\rangle$ to a thermal state. The only exception, of course, is at zero temperature, where the system tends to the ground state.

In order to attribute a quantum state to the system, we have to generalize the notion of kets. The generalization is called a density matrix, which we usually write as ρ. The density matrix of the thermal state is defined as

$$\boxed{\rho = \sum_n P_n |n\rangle\langle n|} \tag{6}$$

the expectation value of $\langle A\rangle$ in (3) can then be written using the trace as

$$\boxed{\langle A\rangle = \mathrm{tr}(A\rho)} \tag{7}$$

which is the same as $\mathrm{tr}(\rho A)$.

Below I will try to better motivate why an object such as (6) makes sense. But before doing so, let me just show you something neat. the decomposition of H as in (1) makes it convenient to study functions of H. For instance

$$H^2 = \sum_n E_n^2 |n\rangle\langle n|$$

$$H^3 = \sum_n E_n^3 |n\rangle\langle n|$$

and so on. thus, if $f(x)$ is some arbitrary function, then

$$f(H) = \sum_n f(E_n) |n\rangle\langle n| \qquad (8)$$

we just apply $f(\,)$ to the eigenvalues and multiply by the projectors $|n\rangle\langle n|$.

If we now look at (6), we see that

$$\boxed{\rho = \frac{1}{Z} \sum_n e^{-\beta E_n} |n\rangle\langle n| = \frac{e^{-\beta H}}{Z}} \qquad (9)$$

This is a very neat and very powerful way of writing the thermal state. what is nice about it is that it makes no reference to any specific basis: it simply writes the state directly as a function of the Hamiltonian.

The partition function can also be written in a neat way as

$$\boxed{Z = \mathrm{tr}(e^{-\beta H})} \tag{10}$$

because

$$\mathrm{tr}(e^{-\beta H}) = \sum_{n} \langle n| e^{-\beta H} |n\rangle = \sum_{n} e^{-\beta E_n} = Z$$

Thus, the expectation value (7) may also be written as

$$\boxed{\langle A\rangle = \frac{\mathrm{tr}(A\, e^{-\beta H})}{\mathrm{tr}(e^{-\beta H})}} \tag{11}$$

The results above are for thermal states. but the idea of a density matrix is absolutely general. To motivate the reasoning behind this I introduce below a construction of the density matrix which is more general. Hopefully it will clarify the logic behind these objects.

Chapter 2

Density matrix theory

2.1 The density matrix

A ket $|\psi\rangle$ is actually not the most general way of defining a quantum state. To motivate this, consider the state $|\boldsymbol{n}_+\rangle$ in Eq. (1.47) and the corresponding expectation values computed in Eq. (1.48). This state is always poiting *somewhere*: it points at the direction $\boldsymbol{n}$ of the Bloch sphere. It is impossible, for instance, to find a quantum ket which is isotropic. That is, where $\langle\sigma_x\rangle = \langle\sigma_y\rangle = \langle\sigma_z\rangle = 0$. That sounds strange. The solution to this conundrum lies in the fact that we need to also introduce some *classical uncertainty* to the problem. Kets are only able to encompass quantum uncertainty.

The most general representation of a quantum system is written in terms of an operator ρ called the *density operator*, or *density matrix*. It is built in such a way that it naturally encompasses both quantum and classical probabilities. But that is not all. We will also learn next chapter that density matrices are intimately related to **entanglement**. So even if we have *no* classical uncertainties, we will also eventually find the need for dealing with density matrices. For this reason, the density matrix is *the most important concept in quantum theory*. I am not exaggerating. You started this chapter as a kid. You will finish it as an adult. :)

To motivate the idea, imagine we have a machine which prepares quantum systems in certain states. For instance, this could be an oven producing spin 1/2 particles, or a quantum optics setup producing photons. But suppose that this apparatus is imperfect, so it does not always produces the same state. That is, suppose that it produces a state $|\psi_1\rangle$ with a certian probability p_1 or a state $|\psi_2\rangle$ with a certain probability p_2 and so on. Notice how we are introducing here a *classical uncertainty*. The $|\psi_i\rangle$ are quantum states, but we simply *don't know* which states we will get out of the machine. We can have as many p's as we want. All we need to assume is that satisfy the properties expected from a probability:

$$p_i \in [0, 1], \qquad \text{and} \qquad \sum_i p_i = 1 \tag{2.1}$$

Now let A be an observable. If the state is $|\psi_1\rangle$, then the expectation value of A will be $\langle\psi_1|A|\psi_1\rangle$. But if it is $|\psi_2\rangle$ then it will be $\langle\psi_2|A|\psi_2\rangle$. To compute the actual expectation value of A we must therefore perform an **average of quantum averages**:

$$\langle A \rangle = \sum_i p_i \langle \psi_i | A | \psi_i \rangle \tag{2.2}$$

We simply weight the possible expectation values $\langle \psi_i | A | \psi_i \rangle$ by the relative probabilities p_i that each one occurs.

What is important to realize is that this type of average *cannot* be writen as $\langle \phi | A | \phi \rangle$ for some ket $|\phi\rangle$. If we want to attribute a "state" to our system, then we must generalize the idea of ket. To do that, we use Eq. (1.76) to write

$$\langle \psi_i | A | \psi_i \rangle = \mathrm{tr}\Big[A |\psi_i\rangle\langle\psi_i| \Big]$$

Then Eq. (2.2) may be written as

$$\langle A \rangle = \sum_i p_i \, \mathrm{tr}\Big[A |\psi_i\rangle\langle\psi_i| \Big] = \mathrm{tr}\Big\{ A \sum_i p_i |\psi_i\rangle\langle\psi_i| \Big\}$$

This motivates us to define the **density matrix** as

$$\boxed{\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|} \tag{2.3}$$

Then we may finally write Eq. (2.2) as

$$\boxed{\langle A \rangle = \mathrm{tr}(A\rho)} \tag{2.4}$$

which, by the way, is the same as $\mathrm{tr}(\rho A)$ since the trace is cyclic [Eq. (1.74)].

With this idea, we may now recast *all* of quantum mechanics in terms of density matrices, instead of kets. If it happens that a density matrix can be written as $\rho = |\psi\rangle\langle\psi|$, we say we have a **pure state**. And in this case it is not necessary to use ρ at all. One may simply continue to use $|\psi\rangle$. For instance, Eq. (2.4) reduces to the usual result: $\mathrm{tr}(A\rho) = \langle \psi | A | \psi \rangle$. A state which is not pure is usually called a **mixed state**. In this case kets won't do us no good and we *must* use ρ.

Examples

Let' s play with some examples. To start, suppose a machine tries to produce qubits in the state $|0\rangle$. But it is not very good so it only produces $|0\rangle$ with probability p. And, with probability $1 - p$ it produces the state $|1\rangle$. The density matrix would then be.

$$\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}.$$

Or it could be such that it produces either $|0\rangle$ or $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Then,

$$\rho = p|0\rangle\langle 0| + (1-p)|+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1+p & 1-p \\ 1-p & 1-p \end{pmatrix}.$$

Maybe if our device is not completely terrible, it will produce most of the time $|0\rangle$ and every once in a while, a state $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle$, where θ is some small angle. The density matrix for this system will then be

$$\rho = p|0\rangle\langle 0| + (1-p)|\psi\rangle\langle\psi| = \begin{pmatrix} p + (1-p)\cos^2\frac{\theta}{2} & (1-p)\sin\frac{\theta}{2}\cos\frac{\theta}{2} \\ (1-p)\sin\frac{\theta}{2}\cos\frac{\theta}{2} & (1-p)\sin^2\frac{\theta}{2} \end{pmatrix}$$

Of course, the machine can very well produce more than 2 states. But you get the idea.

Next let's talk about something really cool (and actually quite deep), called the **ambiguity of mixtures**. The idea is quite simple: if you mix stuff, you generally loose information, so you don't always know where you started at. To see what I mean, consider a state which is a 50-50 mixture of $|0\rangle$ and $|1\rangle$. The corresponding density matrix will then be

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Alternatively, consider a 50-50 mixture of the states $|\pm\rangle$ in Eq. (1.11). In this case we get

$$\rho = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that both are identical. Hence, we have *no way to tell* if we began with a 50-50 mixture of $|0\rangle$ and $|1\rangle$ or of $|+\rangle$ and $|-\rangle$. By mixing stuff, we have lost information.

2.2 Properties of the density matrix

The density matrix satisfies a bunch of very special properties. We can figure them out using only the definition (2.3) and recalling that $p_i \in [0,1]$ and $\sum_i p_i = 1$ [Eq. (2.1)]. First, the density matrix is a Hermitian operator:

$$\rho^\dagger = \rho. \tag{2.5}$$

Second,

$$\mathrm{tr}(\rho) = \sum_i p_i \,\mathrm{tr}(|\psi_i\rangle\langle\psi_i|) = \sum_i p_i\langle\psi_i|\psi_i\rangle = \sum_i p_i = 1. \tag{2.6}$$

This is the *normalization* condition of the density matrix. Another way to see this is from Eq. (2.4) by choosing $A = 1$. Then, since $\langle 1\rangle = 1$ we again get $\mathrm{tr}(\rho) = 1$.

We also see from Eq. (2.8) that $\langle\phi|\rho|\phi\rangle$ is a sum of quantum probabilities $|\langle\phi|\psi_i\rangle|^2$ averaged by classical probabilities p_i. This entails the following interpretation: for an arbitrary state $|\phi\rangle$,

$$\boxed{\langle\phi|\rho|\phi\rangle = \text{Prob. of finding the system at state } |\phi\rangle \text{ given that it's state is } \rho} \tag{2.7}$$

Besides normalization, the other big property of a density matrix is that it is **positive semi-definite**, which we write symbolically as $\rho \geq 0$. What this means is that *its sandwich in any quantum state is always non-negative*. In symbols, if $|\phi\rangle$ is an arbitrary quantum state then

$$\langle\phi|\rho|\phi\rangle = \sum_i p_i |\langle\phi|\psi_i\rangle|^2 \geq 0. \tag{2.8}$$

Of course, this makes sense in view of the probabilistic interpretation of Eq. (2.7). Please note that this does *not* mean that all entries of ρ are non-negative. Some of them may be negative. It does mean, however, that the diagonal entries are always non-negative, no matter which basis you use.

Another equivalent definition of a positive semi-definite operator is one whose **eigenvalues are always non-negative**. In Eq. (2.3) it already looks as if ρ is in diagonal form. However, we need to be a bit careful because the $|\psi_i\rangle$ are arbitrary states and do not necessarily form a basis (which can be seen explicitly in the examples given above). Thus, in general, the diagonal structure of ρ will be different. Notwithstanding, ρ is Hermitian and may therefore be diagonalized by some orthonormal basis $|\lambda_k\rangle$ as

$$\rho = \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|, \tag{2.9}$$

for certain eigenvalues λ_k. Since Eq. (2.8) must be true for any state $|\phi\rangle$ we may choose, in particular, $|\phi\rangle = |\lambda_k\rangle$, which gives

$$\lambda_k = \langle\lambda_k|\rho|k\rangle \geq 0.$$

Thus, we see that the statement of positive semi-definiteness is equivalent to saying that the eigenvalues are non-negative. In addition to this, we also have that $\text{tr}(\rho) = 1$, which implies that $\sum_k \lambda_k = 1$. Thus we conclude that the eigenvalues of ρ behave like probabilities:

$$\lambda_k \in [0,1], \qquad \sum_k \lambda_k = 1. \tag{2.10}$$

But they are not the same probabilities p_i. They just behave like a set of probabilities, that is all.

For future reference, let me summarize what we learned in a big box: the basic properties of a density matrix are

$$\boxed{\text{Defining properties of a density matrix:} \qquad \text{tr}(\rho) = 1 \quad \text{and} \quad \rho \geq 0.} \tag{2.11}$$

Any normalized positive semi-definite matrix is a valid candidate for a density matrix.

I emphasize again that the notation $\rho \geq 0$ in Eq. (2.11) means the matrix is positive semi-definite, not that the entries are positive. For future reference, let me list here some properties of positive semi-definite matrices:

- $\langle\phi|\rho|\phi\rangle \geq 0$ for any state $|\phi\rangle$;
- The eigenvalues of ρ are always non-negative.
- The diagonal entries are always non-negative.
- The off-diagonal entries in any basis satisfy $|\rho_{ij}| \leq \sqrt{\rho_{ii}\rho_{jj}}$.

2.3 Purity

Next let us look at ρ^2. The eigenvalues of this matrix are λ_k^2 so

$$\operatorname{tr}(\rho^2) = \sum_k \lambda_k^2 \leq 1 \tag{2.12}$$

The only case when $\operatorname{tr}(\rho^2) = 1$ is when ρ is a pure state. In that case it can be written as $\rho = |\psi\rangle\langle\psi|$ so it will have one eigenvalue $p_1 = 1$ and all other eigenvalues equal to zero. Hence, the quantity $\operatorname{tr}(\rho^2)$ represents the **purity** of the quantum state. When it is 1 the state is pure. Otherwise, it will be smaller than 1:

$$\boxed{\text{Purity} = \mathcal{P} := \operatorname{tr}(\rho^2) \leq 1} \tag{2.13}$$

As a side note, when the dimension of the Hilbert space d is finite, it also follows that $\operatorname{tr}(\rho^2)$ will have a lower bound:

$$\frac{1}{d} \leq \operatorname{tr}(\rho^2) \leq 1 \tag{2.14}$$

This lower bound occurs when ρ is the **maximally disordered state**

$$\rho = \frac{\mathbb{I}_d}{d} \tag{2.15}$$

where $\mathbb{I}_d$ is the identity matrix of dimension d.

2.4 Bloch's sphere and coherence

The density matrix for a qubit will be 2 × 2 and may therefore be parametrized as

$$\rho = \begin{pmatrix} p & q \\ q^* & 1-p \end{pmatrix}, \tag{2.16}$$

where $p \in [0,1]$ and I used $1-p$ in the last entry due to the normalization $\operatorname{tr}(\rho^2) = 1$. If the state is pure then it can be written as $|\psi\rangle = a|0\rangle + b|1\rangle$, in which case the density matrix becomes

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{pmatrix}. \tag{2.17}$$

This is the density matrix of a system which is in a *superposition* of $|0\rangle$ and $|1\rangle$. Conversely, we could construct a state which can be in $|0\rangle$ **or** $|1\rangle$ with different probabilities. According to the very definition of the density matrix in Eq. (2.3), this state would be

$$\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}. \tag{2.18}$$

Pauli matrices: $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

This is a classical state, obtained from classical probability theory. The examples in Eqs. (2.17) and (2.18) reflect well the difference between quantum superpositions and classical probability distributions.

Another convenient way to write the state (2.16) is as

$$\boxed{\rho = \frac{1}{2}(1 + \boldsymbol{s} \cdot \boldsymbol{\sigma}) = \frac{1}{2}\begin{pmatrix} 1 + s_z & s_x - is_y \\ s_x + is_y & 1 - s_z \end{pmatrix}.} \tag{2.19}$$

where $\boldsymbol{s} = (s_x, s_y, s_z)$ is a vector. The physical interpretation of $\boldsymbol{s}$ becomes evident from the following relation:

$$s_i = \langle \sigma_i \rangle = \mathrm{tr}(\sigma_i \rho). \tag{2.20}$$

The relation between these parameters and the parametrization in Eq. (2.16) is

$$\begin{aligned} \langle \sigma_x \rangle &= q + q^*, \\ \langle \sigma_y \rangle &= i(q - q^*), \\ \langle \sigma_z \rangle &= 2p - 1. \end{aligned}$$

Next we look at the purity of a qubit density matrix. From Eq. (2.19) one readily finds that

$$\mathrm{tr}(\rho^2) = \frac{1}{2}(1 + \boldsymbol{s}^2). \tag{2.21}$$

Thus, due to Eq. (2.12), it also follows that

$$\boldsymbol{s}^2 = s_x^2 + s_y^2 + s_z^2 \leq 1. \tag{2.22}$$

When $\boldsymbol{s}^2 = 1$ we are in a pure state. In this case the vector $\boldsymbol{s}$ lies on the surface of the Bloch sphere. For mixed states $\boldsymbol{s}^2 < 1$ and the vector is inside the Bloch sphere. Thus, we see that the purity can be directly associated with the radius in the sphere. This is pretty cool! The smaller the radius, the more mixed is the state. In particular, the maximally disordered state occurs when $\boldsymbol{s} = 0$ and reads

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.23}$$

In this case the state lies in the center of the sphere. A graphical representation of pure and mixed states in the Bloch sphere is shown in Fig. 2.1.

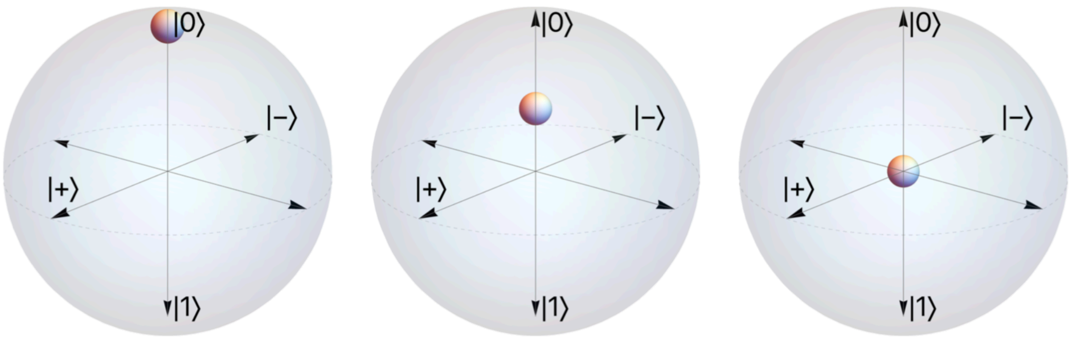


Figure 2.1: Examples of pure and mixed states in the z axis. Left: a pure state. Center: an arbitrary mixed state. Right: the maximally mixed state (2.23).

The Bloch sphere helps us understand why thermal state have to be mixed state. Take, for instance,

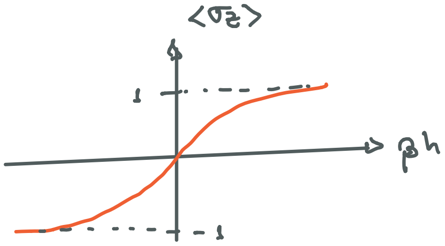
$$H = -\frac{h}{2}\sigma_z = \begin{pmatrix} -h/2 & 0 \\ 0 & h/2 \end{pmatrix}$$

The thermal state is then

$$\rho = \frac{e^{-\beta H}}{Z} = \frac{1}{2\cosh(\beta h/2)} \begin{pmatrix} e^{\beta h/2} & 0 \\ 0 & e^{-\beta h/2} \end{pmatrix}$$

and, as we saw in lecture 1,

$$\langle \sigma_z \rangle = \tanh\left(\frac{\beta h}{2}\right)$$



If $\beta h \to +\infty$ then $\langle \sigma_z \rangle \to 1$. This is the north pole of Bloch's sphere. But when $\beta h \to 0$, then $\langle \sigma_z \rangle \to 0$, which is the equator. Now comes the key point: if the state was pure, then when $\langle \sigma_z \rangle = 0$, the spin would have to be pointing somewhere on the xy plane. A pure state of a qubit is always pointing somewhere. This is something that would be easily detectable experimentally. But it's not what we find. On the contrary, what we find is that when $\beta h \to 0$ the spin doesn't point anywhere, which is only possible if the spin is in the center of Bloch's sphere. This neatly shows, I think, why thermal states are naturally mixed.

The $T \to \infty$ limit

Remember our previous discussions about infinite temperatures. If the system has dimension d, when $T \to \infty$ all states become equally likely with

$$P_n = \frac{1}{d} \tag{12}$$

This now becomes much more obvious in terms of density matrices:

$$\rho = \frac{e^{-\beta H}}{tr(e^{-\beta H})} \tag{13}$$

If $T \to \infty$, $\beta \to 0$ and thus

$$e^{-\beta H} \to I = \text{identity matrix} \tag{14}$$

Moreover, for a system with dimension d,

$$tr(I) = d \tag{15}$$

Thus

$$\lim_{T \to \infty} \rho = \frac{I}{d} \tag{16}$$

which is the maximally mixed state.

The prob. of finding the system in an arbitrary state $|\phi\rangle$ [Eq. (2.8) above] now becomes

$$\langle \phi | \rho | \phi \rangle = \frac{1}{d} \langle \phi | \phi \rangle = \frac{1}{d} \tag{17}$$

At $T \to \infty$ the system is equally likely to be found in any state of the Hilbert space.

Example: qubits vs. qutrits.

The density matrix for a 2-level and 3-level system, assuming they are diagonal, would have the form

$$\rho_2 = \begin{pmatrix} p_0 & 0 \\ 0 & p_1 \end{pmatrix} \qquad \rho_2 = \begin{pmatrix} p_0 & 0 & 0 \\ 0 & p_1 & 0 \\ 0 & 0 & p_2 \end{pmatrix}$$

If we think about it, a diagonal state of a 2-level system is always thermal: that is, given p_0 and $p_1 = 1 - p_0$, we can always find a value of β such that

$$\frac{e^{-\beta E_0}}{e^{-\beta E_0} + e^{-\beta E_1}} = p_0 \qquad \frac{e^{-\beta E_1}}{e^{-\beta E_0} + e^{-\beta E_1}} = p_1 = 1 - p_0$$

If $p_1 > p_0$ we may need $\beta < 0$. But still, we can always view a qubit diagonal state as thermal. For qutrits, however, this is not in general the case. We now have 2 independent probabilities (only 2 because $p_2 = 1 - p_0 - p_1$) and in general we cannot find a single number β which fits both.

I'm just saying all this to call attention to the special structure of thermal states: thermal states are special because the populations appear in a very specific proportion, related to the value of β.

Example : QHO

The QHO is characterized by creation and annihilation operators $a^\dagger$ and a. They are such that

$$\boxed{a^\dagger a|n\rangle = n|n\rangle \quad , \; n = 0, 1, 2, 3, \ldots} \tag{18}$$

and

$$\begin{aligned} a^\dagger |n\rangle &= \sqrt{n+1}\,|n+1\rangle \\ a|n\rangle &= \sqrt{n}\,|n-1\rangle \end{aligned} \tag{19}$$

The Hamiltonian then reads

$$H = \hbar\omega\,(a^\dagger a + 1/2) \tag{20}$$

The partition function was computed in lecture 1 and reads

$$Z = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}} \tag{21}$$

Thus the thermal density matrix becomes

$$\boxed{\rho = (1 - e^{-\beta\hbar\omega})\, e^{-\beta\hbar\omega a^\dagger a}} \tag{22}$$

The generalization of the Shannon entropy to density matrices is called the von Neumann entropy.

2.8 The von Neumann Entropy

The concept of entropy plays a central role in classical and quantum information theory. In its simplest interpretation, entropy is a measure of the disorder (or mixedness) of a density matrix, quite like the purity $\mathrm{tr}(\rho^2)$. But with entropy this disorder acquires a more informational sense. We will therefore start to associate entropy with questions like "how much information is stored in my system". We will also introduce another extremely important concept, called **relative entropy** which plays the role of a "distance" between two density matrices.

Given a density matrix ρ, the von Neumann entropy is defined as

$$\boxed{S(\rho) = -\,\mathrm{tr}(\rho \log \rho) = -\sum_k \lambda_k \log \lambda_k,} \tag{2.53}$$

In terms of λ_k we simply get the Shannon entropy

where λ_k are the eigenvalues of ρ. Working with the logarithm of an operator can be awkward. That is why in the last equality I expressed $S(\rho)$ in terms of them. The logarithm in Eq. (2.53) can be either base 2 or base e. It depends if the application is more oriented towards information theory or physics (respectively). The last expression in (2.53), in terms of a sum of probabilities, is also called the **Shannon entropy**.

The entropy is seen to be a sum of functions of the form $-p\log(p)$, where $p \in [0, 1]$. The behavior of this function is shown in Fig. 2.3. It tends to zero both when $p \to 0$ and $p \to 1$, and has a maximum at $p = 1/e$. Hence, any state which has $p_k = 0$ *or* $p_k = 1$ will not contribute to the entropy (even though $\log(0)$ alone diverges, $0\log(0)$ is well behaved). States that are too deterministic therefore contribute little to the entropy. Entropy likes randomness.

Since each $-p\log(p)$ is always non-negative, the same must be true for $S(\rho)$:

$$S(\rho) \geq 0. \tag{2.54}$$

Moreover, if the system is in a pure state, $\rho = |\psi\rangle\langle\psi|$, then it will have one eigenvalue $p_1 = 1$ and all others zero. Consequently, in a pure state the entropy will be zero:

$$\boxed{\text{The entropy of a pure state is zero.}} \tag{2.55}$$

In information theory the quantity $-\log(p_k)$ is sometimes called the *surprise*. When an "event" is rare ($p_k \sim 0$) this quantity is big ("surprise!") and when an event is common ($p_k \sim 1$) this quantity is small ("meh"). The entropy is then interpreted as the *average surprise* of the system, which I think is a little bit funny.

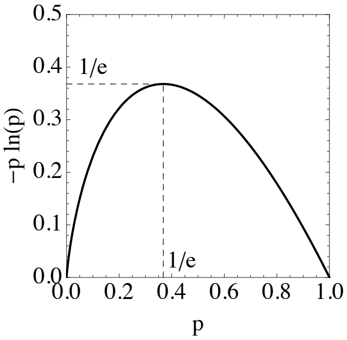


Figure 2.3: The function $-p\log(p)$, corresponding to each term in the von Neumann entropy (2.53).

As we have just seen, the entropy is bounded from below by 0. But if the Hilbert space dimension d is finite, then the entropy will also be bounded from above. I will leave this proof for you as an exercise. What you need to do is maximize Eq. (2.53) with respect to the p_k, but using Lagrange multipliers to impose the constraint $\sum_k p_k = 1$. Or, if you are not in the mood for Lagrange multipliers, wait until Eq. (**??**) where I will introduce a much easier method to demonstrate the same thing. In any case, the result is

$$\max(S) = \log(d). \qquad \text{Occurs when } \rho = \frac{\mathbb{I}}{d}. \tag{2.56}$$

The entropy therefore varies between 0 for pure states and $\log(d)$ for maximally disordered states. Hence, it clearly serves as a measure of how mixed a state is.

Another very important property of the entropy (2.53) is that it is invariant under unitary transformations:

$$S(U\rho U^\dagger) = S(\rho). \tag{2.57}$$

This is a consequence of the infiltration property of the unitaries $Uf(A)U^\dagger = f(UAU^\dagger)$ together with the cyclic property of the trace. Since the time evolution of closed systems are implemented by unitary transformations, this means that the entropy is a constant of motion. We have seen that the same is true for the purity: unitary evolutions do not change the mixedness of a state. Or, in the Bloch sphere picture, unitary evolutions keep the state on the same spherical shell. For open quantum systems this will no longer be the case.

As a quick example, let us write down the formula for the entropy of a qubit. Recall the discussion in Sec. 2.4: the density matrix of a qubit may always be written as in Eq. (2.19). The eigenvalues of ρ are therefore $(1 \pm s)/2$ where $s = \sqrt{s_x^2 + s_y^2 + s_z^2}$ represents the radius of the state in Bloch's sphere. Hence, applying Eq. (2.53) we get

$$S = -\left(\frac{1+s}{2}\right)\log\left(\frac{1+s}{2}\right) - \left(\frac{1-s}{2}\right)\log\left(\frac{1-s}{2}\right). \tag{2.58}$$

For a pure state we have $s = 1$ which then gives $S = 0$. On the other hand, for a maximally disordered state we have $s = 0$ which gives the maximum value $S = \log 2$, the log of the dimension of the Hilbert space. The shape of S is shown in Fig. 2.4.

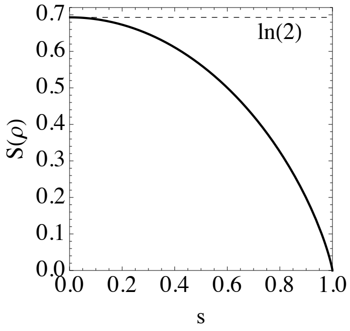


Figure 2.4: The von Neumann entropy for a qubit, Eq. (2.58), as a function of the Bloch-sphere radius s.

The quantum relative entropy

Another very important quantity in quantum information theory is the *quantum relative entropy* or *Kullback-Leibler divergence*. Given two density matrices ρ and σ, it is defined as

$$\boxed{S(\rho||\sigma) = \text{tr}(\rho\log\rho - \rho\log\sigma).} \tag{2.59}$$

This quantity is important for a series of reasons. But one in particular is that it satisfies the *Klein inequality*:

$$\boxed{S(\rho||\sigma) \geq 0, \qquad S(\rho||\sigma) = 0 \text{ iff } \rho = \sigma.} \tag{2.60}$$

The proof of this inequality is really boring and I'm not gonna do it here. You can find it in Nielsen and Chuang or even in Wikipedia.

Eq. (2.60) gives us the idea that we could use the relative entropy as a measure of the *distance* between two density matrices. But that is not entirely precise since the relative entropy does not satisfy the triangle inequality

$$d(x,z) \leq d(x,y) + +d(y,z). \tag{2.61}$$

This is something a true measure of distance must always satisfy. If you are wondering what quantities are actual distances, the *trace distance* is one of them[2]

$$\mathcal{T}(\rho,\sigma) = ||\rho - \sigma||_1 := \text{tr}\left[\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}\right]. \tag{2.62}$$

But there are others as well.

Entropy and information

Define the maximally mixed state $\pi = I_d/d$. This is the state we know absolutely nothing about. We have zero information about it. Motivated by this, we can define the amount of information in a state ρ as the "distance" between ρ and π; viz,

$$\mathcal{I}(\rho) = S(\rho||\pi).$$

But we can also open this up as

$$S(\rho||1/d) = \text{tr}(\rho\log\rho) - \text{tr}(\rho\log(I_d/d)) = -S(\rho) + \log(d).$$

I know it is a bit weird to manipulate I_d/d here. But remember that the identity matrix satisfies *exactly* the same properties as the number one, so we can just use the usual algebra of logarithms in this case.

We see from this result that the information contained in a state is nothing but

$$\boxed{\mathcal{I}(\rho) = S(\rho||\pi) = \log(d) - S(\rho).} \tag{2.63}$$

This shows how information is connected with entropy. The larger the entropy, the less information we have about the system. For the maximally mixed state $S(\rho) = \log(d)$ and we get zero information. For a pure state $S(\rho) = 0$ and we have the maximal information $\log(d)$.

As I mentioned above, the relative entropy is very useful in proving some mathematical relations. For instance consider the result in Eq. (2.56). If we look at Eq. (2.63) and remember that $S(\rho||\sigma) \geq 0$, this result becomes kind of obvious: $S(\rho) \leq \log(d)$ and $S(\rho) = \log(d)$ iff $\rho = 1/d$, which is precisely Eq. (2.56).

---

[2] The fact that $\rho - \sigma$ is Hermitian can be used to simplify this a bit. I just wanted to write it in a more general way, which also holds for non-Hermitian operators.

Response functions

A very common problem in statistical mechanics is that of measuring the response of a system to a certain external stimulus. This is usually formulated as follows. We have a system with Hamiltonian H_0. We then apply a stimulus λ which couples to some operator A, such that the Hamiltonian changes to

$$\boxed{H = H_0 - \lambda A} \qquad (23)$$

Of course, the coupling doesn't have to be like this. But this kind of coupling turns out to be quite common. The most important example is that of a spin system coupled to a magnetic field, so that

$$H = H_0 - h S_z \qquad (24)$$

I will work with (23) for now though, because I want to emphasize that the results are also valid beyond spin systems.

The quantity we now want to evaluate is the response of the system to the stimulus λ,

$$\langle A \rangle = \frac{\operatorname{tr}(A e^{-\beta H})}{\operatorname{tr}(e^{-\beta H})} \qquad (25)$$

What I want to show is that this can be written as

$$\boxed{\langle A \rangle = -\frac{\partial F}{\partial \lambda}} \qquad (26)$$

where $F = -T \ln Z$.

You already showed (26) in the problem set for the case of spin. But here I would like to go through a more careful proof, because there is actually a subtlety involved, related to whether or not A and Ho commute.

The reason is that in order to write something like (26) we need to compute

$$\frac{\partial}{\partial \lambda} e^{-\beta H}$$

If $[H_0, A] = 0$ then we can write

$$e^{-\beta H} = e^{-\beta H_0} e^{\beta \lambda A}$$

In this case

$$\frac{\partial}{\partial \lambda} e^{-\beta H} = e^{-\beta H_0} \frac{\partial}{\partial \lambda} e^{\beta \lambda A}$$

and

$$\frac{\partial}{\partial \lambda} e^{\beta \lambda A} = \beta A e^{\beta \lambda A} = e^{\beta \lambda A} \beta A$$

since A commutes with itself. Then

$$\frac{\partial}{\partial \lambda} e^{-\beta H} = \beta A e^{-\beta H} \qquad (\text{when } [H_0, A] = 0) \qquad (27)$$

so that going from (25) to (26) is easy

$$\langle A \rangle = \frac{1}{\beta} \frac{1}{Z} \frac{\partial Z}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left(\frac{1}{\beta} \ln Z\right) = -\frac{\partial F}{\partial \lambda}$$

But if $[H, A] \neq 0$, computing $\partial e^{-\beta H}/\partial \lambda$ is not trivial. Quite surprisingly though, Eq (26) continues to hold. Let's see why.

It is useful to have the following BCH expansion in mind: let $K(\lambda)$ denote an arbitrary operator. then

$$\boxed{\frac{d}{d\lambda} e^{K(\lambda)} = \left\{ \frac{dK}{d\lambda} + \frac{1}{2}\left[K, \frac{dK}{d\lambda}\right] + \frac{1}{3!}\left[K,\left[K, \frac{dK}{d\lambda}\right]\right] + \dots \right\} e^{K(\lambda)}} \tag{28}$$

You can verify this formula by series expanding the exponentials in both sides in a Taylor series.

What (28) shows is that to compute $\frac{de^K}{d\lambda}$, it matters whether or not $\frac{dK}{d\lambda}$ commutes with K. If they do, then it's easy. If they don't, then we get a (in principle) infinite series.

For our problem $K(\lambda) = -\beta H_0 + \beta\lambda A$ so $\frac{\partial K}{\partial \lambda} = \beta A$. We thus get

$$\frac{\partial}{\partial\lambda} e^{-\beta H} = \left\{ \beta A + \frac{1}{2}[K, \beta A] + \frac{1}{3!}[K,[K,\beta A]] + \dots \right\} e^{-\beta H} \tag{29}$$

But now comes the fun part! let's compute

$$\mathrm{tr}\left\{\frac{\partial}{\partial\lambda} e^{-\beta H}\right\} = \underbrace{\mathrm{tr}\left\{\beta A\, e^{-\beta H}\right\}}_{\text{what we want!}} + \frac{1}{2}\mathrm{tr}\left\{[K, \beta A]\, e^{-\beta H}\right\} + \frac{1}{3!}\mathrm{tr}\left\{[K,[K,\beta A]]\, e^{-\beta H}\right\} + \dots \tag{30}$$

Using the cyclic property of the trace, we can write the 2^{nd} term as

$$\mathrm{tr}\left\{[K, \beta A]\, e^{K}\right\} = \mathrm{tr}\left\{K\, \beta A\, e^{K}\right\} - \mathrm{tr}\left\{\beta A\, K e^{K}\right\}$$

$$= \mathrm{tr}\left\{\beta A\, e^{K} K\right\} - \mathrm{tr}\left\{\beta A\, K e^{K}\right\}$$

$$= 0$$

since K and e^{K} commute. The same is also true for all other terms in (28). Thus we reach the quite nice conclusion that

$$\boxed{\frac{\partial}{\partial \lambda}\, \mathrm{tr}\left\{e^{-\beta H}\right\} = \beta\, \mathrm{tr}\left\{A\, e^{-\beta H}\right\}} \qquad (31)$$

the derivation of (26) therefore remains unaltered, even if A and H_0 do not commute.

Susceptibility

The susceptibility is defined as the sensitivity of $\langle A \rangle$ to changes in λ:

$$\boxed{\chi = \frac{\partial \langle A \rangle}{\partial \lambda} = - \frac{\partial^2 F}{\partial \lambda^2}} \qquad (32)$$

Plugging in the definition of $\langle A \rangle$, together with (31), we get

$$\chi = \frac{\partial}{\partial \lambda} \frac{\mathrm{tr}(A e^{-\beta H})}{\mathrm{tr}(e^{-\beta H})}$$

$$= \frac{1}{Z} \frac{\partial}{\partial \lambda} \mathrm{tr}(A e^{-\beta H}) - \frac{\mathrm{tr}(A e^{-\beta H})}{Z^2} \frac{\partial Z}{\partial \lambda}$$

$$= \frac{1}{Z} \frac{\partial}{\partial \lambda} \mathrm{tr}(A e^{-\beta H}) - \beta \langle A \rangle^2$$

The first term is a bit tricky. If $[H_0, A] = 0$ we can use (27) to get

$$\chi = \frac{\beta}{Z} \mathrm{tr}(A^2 e^{-\beta H}) - \beta \langle A \rangle^2$$

or

$$\boxed{\chi = \beta \left[\langle A^2 \rangle - \langle A \rangle^2 \right]} \qquad (33)$$

This result is quite nice. It shows that the susceptibility (the sensitivity of $\langle A \rangle$ to changes in λ) is actually related to the variance of A; i.e. to the fluctuations of A in equilibrium.

When $[H_0, A] \neq 0$ Eq (33) no longer holds. In this case it is useful to use the following Feynmann integral tricky:

$$\frac{\partial}{\partial \lambda} e^{-\beta H} = \beta \int_0^1 ds \, e^{-\beta H s} A \, e^{\beta H s} \, e^{-\beta H} \qquad (34)$$

This identity is actually a consequence of (28). It is demonstrated in the appendix below. We then get

$$\chi = \frac{\beta}{Z} \int_0^1 ds \, \mathrm{tr}\left\{ A \, e^{-\beta H s} A \, e^{\beta H s} \, e^{-\beta H} \right\} - \beta \langle A \rangle^2$$

or

$$\boxed{\chi = \beta \left[\int_0^1 ds \, \langle A \, e^{-\beta H s} A \, e^{\beta H s} \rangle - \langle A \rangle^2 \right]} \qquad (35)$$

If $[H_0, A] = 0$ this clearly reduces to (33). But written like this, it clearly shows how $[H_0, A] \neq 0$ affects χ.

The susceptibility is defined for any λ. However, quite often, one is interested in the value of χ for $\lambda \to 0$. This is linear response theory: we apply an infinitesimal perturbation and analyze how the system responds to it. To leading order we may in this case compute χ from the unperturbed state H_0:

$$\chi = \beta \left[\int_0^1 ds \, \langle A \, e^{-\beta H_0 s} A \, e^{\beta H_0 s} \rangle_0 - \langle A \rangle_0^2 \right] \qquad (36)$$

where $\langle \rangle_0 = tr(\dots e^{-\beta H_0}/Z_0)$ is a thermal average over the unperturbed Hamiltonian H_0.

Appendix: a Feynman trick for dealing with $de^K/d\lambda$

We discussed in Eq (26) the following BCH expansion:

$$\frac{d}{d\lambda} e^{K(\lambda)} = \left\{ \frac{dK}{d\lambda} + \frac{1}{2}\left[K, \frac{dK}{d\lambda}\right] + \frac{1}{3!}\left[K, \left[K, \frac{dK}{d\lambda}\right]\right] + \cdots \right\} e^{K(\lambda)} \qquad (A.1)$$

We can rewrite this in terms of an integral identity due to Feynman which can be quite useful. To do that, recall the "traditional" BCH expansion

$$e^X Y e^{-X} = Y + [X, Y] + \frac{1}{2}[X, [X, Y]] + \cdots \qquad (A.2)$$

Comparing with (A.1) we see that the coefficients are a bit off. For instance in (A.1) the factor of $1/2$ multiplies $[K, K']$, whereas in (A.2) it multiplies $[X,[X,Y]]$.

We can fix this by introducing some auxiliary integrals: first,

$$\frac{1}{2}[K, K'] = \int_0^1 ds \, s \, [K, K']$$

since $\int_0^1 ds\, s = 1/2$. Continuing in the same way,

$$\frac{1}{3!}[K, [K, K']] = \int_0^1 ds \, \frac{s^2}{2} \, [K, [K, K']]$$

I'm not doing anything with the operators, just playing with the coefficients.

Eq (A.1) may then be rewritten as

$$\frac{d\,e^{K}}{d\lambda} = \int_0^1 ds \left\{ K + s[K, K'] + \frac{s^2}{2}[K,[K,K']] + \cdots \right\} e^{K}$$

and if we now compare this with (A.2), we see that the integrand is precisely in the usual BCH form:

$$\frac{de^{K}}{d\lambda} = \int_0^1 ds \; e^{Ks} \frac{dK}{d\lambda} e^{-Ks} \; e^{K}$$

Changing $1-s \to s$ we can alternatively write this as

$$\frac{de^{K}}{d\lambda} = e^{K} \int_0^1 ds \; e^{-Ks} \frac{dK}{d\lambda} e^{Ks}$$

Thus, to summarize:

$$\boxed{\frac{de^{K}}{d\lambda} = \int_0^1 ds \; e^{Ks} \frac{dK}{d\lambda} e^{-Ks} \; e^{K} = e^{K} \int_0^1 ds \; e^{-Ks} \frac{dK}{d\lambda} e^{Ks}} \qquad (A.3)$$

If $[K, K'] = 0$ then we can cancel out e^{Ks} with e^{-Ks}, leading to

$$\frac{de^{K}}{d\lambda} = \frac{dK}{d\lambda} e^{K}$$

Otherwise, we have to stick with (A.3).