PHYSICS 391
FALL 2017

QUANTUM INFORMATION THEORY I

Jean Weill

INDEPENDENT STUDY PAPER

Professor S.G. Rajeev
10^{th} December 2017

1 Introduction

A method is more important than a discovery, since the right method will lead to new and even more important discoveries.

Lev Landau (1908-1968)

The objective of this course was to introduce ourselves to quantum information theory, a booming topic in physics with tremendous practical applications. This paper is based on a lecture I gave for the Kapitza Society.

In the first section we will learn about the density operator: a powerful computing tool in quantum mechanics. We will then have a brief overview of Lagrange multipliers and how to use them to maximize entropy and retrieve the canonical ensemble's properties. Finally, we will revealed a surprising relation between statistical and quantum mechanics.

2 The Density Operator

2.1 Quantum Mechanical Ensembles

We generally may not have perfect knowledge of a prepared quantum state. Suppose a third party, Bob, prepares a state for us and only gives a probabilistic description of it, i.e., Bob selects $|\psi_x\rangle$ with probability $p_X(x)$, where p_X is the probability distribution for the random variable X, and x is in some alphabet χ. We can summarize this information by defining an ensemble $\mathcal{E}$ of quantum states

$$\mathcal{E} \equiv \{p_X(x), |\psi_x\rangle\}_{x \epsilon \chi} \tag{1}$$

For example, $\mathcal{E} = \left\{ \left(\frac{1}{3}, |1\rangle\right); \left(\frac{2}{3}, |3\rangle\right) \right\}$. If this ensemble is span by $\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$, then the state is $|0\rangle$ with probability 0, $|1\rangle$ with probability $\frac{1}{3}$, $|2\rangle$ with probability 0, and $|3\rangle$ with probability $\frac{2}{3}$. Of course, the sum of the probabilities adds up to 1.

Consider a system with

$$\mathcal{E} = \{p_i, |\psi_i\rangle\}_{i \epsilon \mathcal{N}} \tag{2}$$

with $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. Suppose $i = 1$, then $\mathcal{E} = \{p_1, |\psi_1\rangle\} = \{1, |\psi_1\rangle\}$. If we measure an observable $\hat{A}$ in $\mathcal{E}$, then

$$\begin{aligned} \left\langle \hat{A} \right\rangle &= \langle\psi_1| \hat{A} |\psi_1\rangle \\ &= 1 \langle\psi_1| \hat{A} |\psi_1\rangle \\ &= p_1 \langle\psi_1| \hat{A} |\psi_1\rangle \end{aligned} \tag{3}$$

Now, what if $\mathcal{E} = \{(p_1, |\psi_1\rangle); (p_2, |\psi_2\rangle)\}$? If you think as p_1 and p_2 as giving weight to their corresponding state's expectation value, you can guess

$$\left\langle \hat{A} \right\rangle = p_1 \langle\psi_1| \hat{A} |\psi_1\rangle + p_2 \langle\psi_2| \hat{A} |\psi_2\rangle$$

and finally, if $\mathcal{E} = \{p_i, |\psi_i\rangle\}_{i \epsilon \mathcal{N}}$, then

$$\left\langle \hat{A} \right\rangle = \sum_i p_i \langle\psi_i| \hat{A} |\psi_i\rangle \tag{4}$$

Claim:

$$\mathcal{I} = \sum_n |n\rangle \langle n| \tag{5}$$

where $\{|n\rangle\}_{n\epsilon\mathcal{N}}$ is a complete set, i. e., $\delta_{nm} = \langle n|m\rangle$, and $\mathcal{I}$ is the identity operator.

Proof:

$\mathcal{I}\,|m\rangle = \sum_n |n\rangle \langle n|m\rangle = \sum_n |n\rangle \delta_{nm} = |m\rangle$

$\mathcal{QED}$

Thus,

$$\begin{aligned}
\left\langle \hat{A} \right\rangle &= \sum_i p_i \langle\psi_i|\,\hat{A}\,|\psi_i\rangle \\
&= \sum_i p_i \langle\psi_i|\,\mathcal{I}\hat{A}\,|\psi_i\rangle \\
&= \sum_i p_i \langle\psi_i| \left(\sum_n |n\rangle\langle n|\right) \hat{A}\,|\psi_i\rangle \\
&= \sum_i p_i \sum_n \langle n|\,\hat{A}\,|\psi_i\rangle \langle\psi_i|n\rangle \\
&= \sum_n \langle n|\,\hat{A} \left(\sum_i p_i |\psi_i\rangle\langle\psi_i|\right) |n\rangle \\
&= \sum_n \langle n|\,\hat{A}\hat{\rho}\,|n\rangle
\end{aligned} \tag{6}$$

where

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle \langle\psi_i| \tag{7}$$

is the density operator.

For some matrix B, we have $B_{ij} = \langle i|\,B\,|j\rangle$, thus $\left\langle \hat{A} \right\rangle = \sum_n (\hat{A}\hat{\rho})_{nn} = Tr(\hat{A}\hat{\rho})$, which gives the very useful and important result

$$\left\langle \hat{A} \right\rangle = Tr(\hat{A}\hat{\rho}) \tag{8}$$

Note that the trace is independent of which complete set you use. Suppose $\{|\phi_i\rangle\}_{i\epsilon\mathcal{N}}$ is a complete set, then using (5) we get

$$\begin{aligned}
Tr(B) = \sum_i \langle i|\,B\,|i\rangle &= \sum_i \langle i|\,B\mathcal{I}\,|i\rangle \\
&= \sum_i \langle i|\,B \left(\sum_j |\phi_j\rangle\langle\phi_j|\right) |i\rangle \\
&= \sum_j \langle\phi_j| \left(\sum_i |i\rangle\langle i|\right) B\,|\phi_j\rangle \\
&= \sum_j \langle\phi_j|\,B\,|\phi_j\rangle
\end{aligned} \tag{9}$$

thus, just think of the complete basis $\{|\delta_i\rangle\}_{i\epsilon\mathcal{N}}$, where for δ_i the i^{th} component is 1 and the other components are 0, to build your intuition.

Why is this density operator so important to us? We already saw that it makes finding the expectation of an observable trivial, but that's not it. It also allows you to find the probability of measuring a particular eigenvalue from your ensemble! To see this, consider an observable $\hat{A}$, so that $\hat{A}\,|a\rangle = a\,|a\rangle$, with $\langle a|a'\rangle = \delta_{aa'}$. What is the probability of measuring a particular eigenvalue a? From the law of total probability, we get

$$p(a) = \textstyle\sum_i \text{ (probability of getting } |\psi_i\rangle) \times (\text{probability of getting } |a\rangle \text{ given that you got } |\psi_i\rangle)$$

but the probability of getting $|\psi_i\rangle$ is just p_i, and the probability of getting a given that you get $|\psi_i\rangle$ is $|\langle a|\psi_i\rangle|^2$, which comes from the Bohr interpretation. If you're not sure why this is true, consider the complete set $\{|n_i\rangle\}_{i\epsilon\mathcal{N}}$ and some normalized quantum state $|\psi\rangle$, then

$$|\psi\rangle = \sum_i n_i\,|n_i\rangle \tag{10}$$

but $\langle\psi|\psi\rangle = 1$, so

$$\begin{aligned} 1 = \sum_i \sum_j n_i^* n_j \langle n_i|n_j\rangle &= \sum_i \sum_j n_i^* n_j \delta_{ij} \\ &= \sum_i n_i^* n_i = \sum_i |n_i|^2 \end{aligned} \tag{11}$$

but we also have $\sum_i p_i = 1$, where in this case p_i is the probability of getting (measuring) $|\psi\rangle$. The idea is to say that $|n_i|^2 = p_i$, which is reasonable since the n_i terms definitely seem to give some weight to how big a role a given $|n_i\rangle$ plays in the construction of $|\psi\rangle$. Finally, you can see from (10) that $n_j = \langle n_j|\psi\rangle$, thus the probability of measuring $|n_j\rangle$ given that your initial state is $|\psi\rangle$ would be $|n_j|^2 = |\langle n_j|\psi\rangle|^2$.

Going back to our $p(a)$, we now understand why we have from the law of total probability

$$\begin{aligned} p(a) = \sum_i p_i |\langle a|\psi_i\rangle|^2 &= \sum_i p_i \langle a|\psi_i\rangle \langle\psi_i|a\rangle \\ &= \langle a| \left(\sum_i p_i\,|\psi_i\rangle\langle\psi_i| \right) |a\rangle \\ &= \langle a|\,\hat{\rho}\,|a\rangle \end{aligned} \tag{12}$$

Therefore, if you have ρ, you can find some of the most useful quantities in quantum mechanics trivially.

We can make this a little bit more general.

<u>Claim:</u>

If the ensemble is given by $\mathcal{E} = \{p_X(x), |\psi_x\rangle\}_{x\epsilon\mathcal{X}}$, we have

$$\hat{\rho} = \sum_{x\epsilon\mathcal{X}} p_X(x)\,|\psi_x\rangle\langle\psi_x| \tag{13}$$

and

$$p_J(j) = Tr\,(\Pi_j\hat{\rho}) \tag{14}$$

Proof:

(13) is trivial to see from (7), but (14) is a bit harder. Recall that $\Pi_j = |j\rangle \langle j|$, we can show the equivalence between (14) and (12)

$$\begin{aligned} p_J(j) = Tr\left(\Pi_j \hat{\rho}\right) &= \sum_i \langle\phi_i| \Pi_j \hat{\rho} |\phi_i\rangle \\ &= \sum_i \langle\phi_i|j\rangle \langle j| \hat{\rho} |\phi_i\rangle \\ &= \sum_i \langle j| \hat{\rho} |\phi_i\rangle \langle\phi_i|j\rangle \\ &= \langle j| \hat{\rho} \left(\sum_i |\phi_i\rangle \langle\phi_i|\right) |j\rangle \\ &= \langle j| \hat{\rho}\mathcal{I} |j\rangle \\ &= \langle j| \hat{\rho} |j\rangle \end{aligned} \tag{15}$$

QED

Let's see how all of this theory works in an actual example. Suppose we are working with electrons. Bob tells us that an electron has a .5 chance of being in the state $|z, \uparrow\rangle$, and .5 chance of being in the state $|z, \downarrow\rangle$. Then $\mathcal{E} = \left\{\left(\frac{1}{2}, |z, \uparrow\rangle\right); \left(\frac{1}{2}, |z, \downarrow\rangle\right)\right\}$. Thus, from (7), we get

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle \langle\psi_i| = \frac{1}{2} |z, \uparrow\rangle \langle z, \uparrow| + \frac{1}{2} |z, \downarrow\rangle \langle z, \downarrow| = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{16}$$

What if I ask you to find the probability of getting $|x, \uparrow\rangle$? Or the expectation value of $\hat{S}_x$? Doing this without using the density operator is not too difficult, you should try it before reading the rest.

Did you try yet? I'm going to put my trust in humankind's righteousness and continue.

The probability of getting $|x, \uparrow\rangle$ is given by

$$\hat{\rho}(|x, \uparrow\rangle) = \langle x, \uparrow| \hat{\rho} |x, \uparrow\rangle = \frac{1}{4} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \tag{17}$$

Hopefully you got the same thing doing it the long way. Now what about $\left\langle \hat{S}_x \right\rangle = \frac{\hbar}{2} \langle \hat{\sigma}_x \rangle$?

$$\langle \hat{\sigma}_x \rangle = Tr(\hat{\sigma}_x \hat{\rho}) = Tr\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\right\} = Tr\left\{\begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}\right\} = 0 \tag{18}$$

2.2 Properties of the Density Operator

Claim 1: $\hat{\rho}$ is Hermitian

Proof:

$$\begin{aligned} \hat{\rho}^\dagger = \left(\sum_i p_i |\psi_i\rangle \langle\psi_i|\right)^\dagger &= \sum_i p_i^* \langle\psi_i|^\dagger |\psi_i\rangle^\dagger \\ &= \sum_i p_i |\psi_i\rangle \langle\psi_i| \\ &= \hat{\rho} \end{aligned} \tag{19}$$

QED

Claim 2: $Tr(\hat{\rho}) = 1$

Proof:

$$\begin{aligned} Tr(\hat{\rho}) = \sum_n \langle n| \hat{\rho} |n\rangle &= \sum_n \langle n| \left(\sum_i p_i |\psi_i\rangle \langle\psi_i| \right) |n\rangle \\ &= \sum_i p_i \langle\psi_i| \left(\sum_n |n\rangle \langle n| \right) |\psi_i\rangle \\ &= \sum_i p_i \langle\psi_i| \mathcal{I} |\psi_i\rangle \\ &= \sum_i p_i \langle\psi_i|\psi_i\rangle \\ &= \sum_i p_i = 1 \end{aligned} \tag{20}$$

$\mathcal{QED}$

where we used the normalization property of the wavefunction, and (5).

Claim 3: $\hat{\rho}$ can be diagonalized to the following

$$\hat{\rho} = \begin{pmatrix} p_1 & 0 & ... & & 0 \\ 0 & p_2 & 0 & ... & 0 \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & p_{|\chi|} \end{pmatrix} \tag{21}$$

where $|\chi|$ is the size of the set.

Proof:

We have that $\hat{\rho} = \sum_i p_i |\psi_i\rangle \langle\psi_i|$, thus

$$\hat{\rho} |\psi_k\rangle = \sum_i p_i |\psi_i\rangle \delta_{ik} = p_k |\psi_k\rangle \tag{22}$$

so $\hat{\rho}$ as eigenvectors $\{|\psi\rangle_k\}_k$ with eigenvalues $\{p_k\}_k$. Let

$$M = \begin{pmatrix} |\psi_1\rangle & |\psi_2\rangle & ... & |\psi_{|\chi|}\rangle \end{pmatrix} \tag{23}$$

Then

$$\hat{\rho}M = \begin{pmatrix} \hat{\rho} |\psi_1\rangle & \hat{\rho} |\psi_2\rangle & ... & \hat{\rho} |\psi_{|\chi|}\rangle \end{pmatrix} = \begin{pmatrix} p_1 |\psi_1\rangle & p_2 |\psi_2\rangle & ... & p_{|\chi|} |\psi_{|\chi|}\rangle \end{pmatrix} \tag{24}$$

which we can rewrite in a more enlightening way

$$\hat{\rho}M = \begin{pmatrix} |\psi_1\rangle & |\psi_2\rangle & ... & |\psi_{|\chi|}\rangle \end{pmatrix} \begin{pmatrix} p_1 & 0 & ... & & 0 \\ 0 & p_2 & 0 & ... & 0 \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & p_{|\chi|} \end{pmatrix} = M \begin{pmatrix} p_1 & 0 & ... & & 0 \\ 0 & p_2 & 0 & ... & 0 \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & p_{|\chi|} \end{pmatrix} \tag{25}$$

if M is invertible, we get

$$M^{-1}\hat{\rho}M = \begin{pmatrix} p_1 & 0 & ... & & 0 \\ 0 & p_2 & 0 & ... & 0 \\ & . & & & \\ & & . & & \\ & & & . & \\ & & & & p_{|\chi|} \end{pmatrix} \tag{26}$$

$\mathcal{QED}$

2.3 Pure vs. Mixed Quantum States

Pure: We know the system is in a particular state $|\psi\rangle \Rightarrow \mathcal{E} = \{1, |\psi\rangle\}$.
Mixed: We don't know the system is in a particular state $|\psi\rangle$, i.e., the system could be in several states $\{|\psi_i\rangle\}_{i\epsilon\mathcal{N}} \Rightarrow \mathcal{E} = \{p_i, |\psi_i\rangle\}_{i\epsilon\mathcal{N}}$.

Using Claim 3, we can define pure and mixed quantum states using the density operator. A mixed state would be given by (21), and a pure state would be given by

$$\hat{\rho} = \begin{pmatrix} 0 & 0 & ... & & 0 \\ 0 & 0 & 0 & ... & 0 \\ & . & & & \\ & & 1 & & \\ & & & . & \\ & & & & 0 \end{pmatrix} \tag{27}$$

since $p_k = 1$ and $p_i = 0$ for all $i \neq k$.
We will come back to this when we will be discussing statistical mechanics and entropy.

2.4 Why is ρ called the Density Operator

$$\left\langle \hat{A} \right\rangle = Tr(\hat{A}\hat{\rho}) = \textstyle\sum_{m,n} \langle n| \hat{A} |m\rangle \langle m| \hat{\rho} |n\rangle$$

which becomes for a continuous basis

$$\left\langle \hat{A} \right\rangle = \int d^3x \int d^3x' \langle x| \hat{A} \left|x'\right\rangle \left\langle x'\right| \hat{\rho} |x\rangle \tag{28}$$

Let's focus on $\left\langle x'\right| \hat{\rho} |x\rangle$.

$$\begin{aligned} \left\langle x'\right| \hat{\rho} |x\rangle &= \left\langle x'\right| \left(\sum_i p_i |\psi_i\rangle \langle\psi_i| \right) |x\rangle \\ &= \sum_i p_i \left\langle x' \middle| \psi_i \right\rangle \langle\psi_i|x\rangle \\ &= \sum_i p_i \psi_i(x') \psi_i^*(x) \end{aligned} \tag{29}$$

If $x = x'$, then $\langle x| \hat{\rho} |x\rangle = \sum_i p_i |\psi_i(x)|^2$. Thus, the diagonal matrix elements are the weighted sums of the probability densities of the states in the ensemble.

3 The Canonical Ensemble

3.1 Lagrange Multipliers

Let $\mathcal{S}$ be a surface given by $f(x, y, z) = c$ for some constant c. What if you want to find which point on this surface is the closest to the origin? To do this we need some background. A curve on this surface is defined by $r(t) = (x(t), y(t), z(t))$. Now, the derivative of this curve, $r^{'}(t)$, is tangent to the curve at any point $P = (x(t), y(t), z(t))$. Let's take one of them $P_0 = (x(0), y(0), z(0)) = (x_0, y_0, z_0)$. At P_0, $r^{'}(t)$ is tangent to the curve and thus tangent to $\mathcal{S}$, which means it lies on the tangent plane of $f(x, y, z)$ at P_0. This means that if a vector is perpendicular to $r^{'}(t)$ at P_0, then it is perpendicular to the tangent plane of $\mathcal{S}$ at P_0.

Claim: The gradient of f is perpendicular to $r^{'}(t)$ at any point.

If we can prove that $\boldsymbol{\nabla} f|_{P_0} \cdot r^{'}(t_0) = 0$ for some arbitrary point P_0 we would be done. But

$$\frac{df}{dt}|_{P_0} = \frac{\partial f}{\partial x}|_{P_0}\frac{dx}{dt}|_{t_0} + \frac{\partial f}{\partial y}|_{P_0}\frac{dy}{dt}|_{t_0} + \frac{\partial f}{\partial z}|_{P_0}\frac{dz}{dt}|_{t_0} = \boldsymbol{\nabla} f|_{P_0} \cdot r^{'}(t_0) \tag{30}$$

but $f(x.y.z) = c$ where c is constant, thus $\frac{df}{dt}|_{P_0} = 0$.

$\mathcal{QED}$

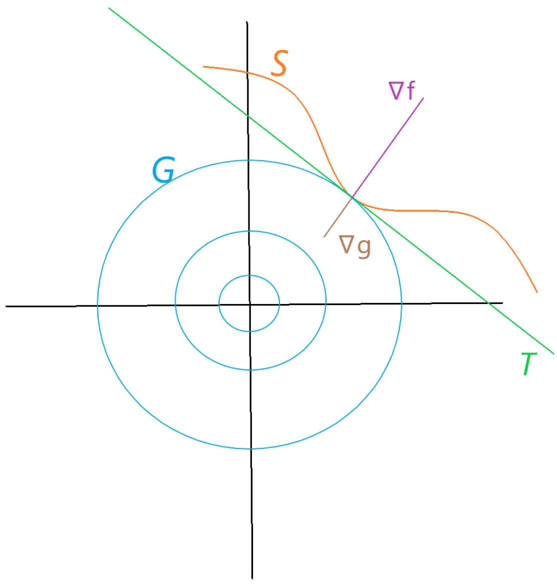

FIG. 1. Lagrange Multipliers

Let's go back to our minimization problem. Try the problem in two dimension, where $\mathcal{S}$ is a curve. Take a compass and start drawing circles with a common center at the origin. Start from a small one and increase the radius of the next one. Continue increasing the radius until a circle enters in contact with $\mathcal{S}$. This point on $\mathcal{S}$ is the closest one to the origin. We of course have the exact same procedure in three dimension with circles being replaced by spheres. Let's call this last sphere $\mathcal{G}$, defined by $g(x, y, z) = x^2 + y^2 + z^2 - r^2 = 0$. Note that g contains the minimization constraint at the heart of the problem. Now, here is the trick, the tangent plane $\mathcal{T}$, of $\mathcal{G}$ and $\mathcal{S}$ at the point of contact is the same one! Therefore, recalling the claim we just proved, this means that $\boldsymbol{\nabla} g$ and $\boldsymbol{\nabla} f$ are both perpendicular to this tangent plane, and so they are parallel to each other. This means that

$$\boldsymbol{\nabla} f = \lambda \boldsymbol{\nabla} g \tag{31}$$

where g is the function containing the constraint, and λ is the famous Lagrange multipliers. The general case is given by

$$\boldsymbol{\nabla} f = \sum_i \lambda \boldsymbol{\nabla} g_i \tag{32}$$

3.2 Mixed States and Entropy

We know that the entropy of a system is given by

$$S = -k \sum_i p_i ln p_i = -kTr(\hat{\rho} ln \hat{\rho}) \tag{33}$$

$\hat{\rho}$ is diagonalized. What are the $\{p_i\}$ in the minimum entropy state?
From (33) it's easy to see that S is a of positive quantities $\Rightarrow S \geq 0 \Rightarrow S_{min} = 0$. Recall that we need $\sum_i p_i = 1$, which means that $p_k = 1$ and $p_i = 0 \ \forall i \neq k$. This might ring a bell (no pun intended), we have in this case a pure state. Thus, in order to have minimum entropy, an ensemble must be composed of a pure state.

Now what about maximizing entropy? This is where Lagrange multipliers come handy. In this case, $f = S$, and $g = \sum_j p_j - 1$. Looking back at (32), we must have

$$\boldsymbol{\nabla}\left(-k \sum_i p_i ln p_i\right) = \lambda \boldsymbol{\nabla}\left(\sum_i p_i - 1\right) \tag{34}$$

We can simplify this quite a bit

$$\begin{aligned} \boldsymbol{\nabla}\left(-k \sum_i p_i ln p_i - \lambda\left(\sum_i p_i - 1\right)\right) &= \sum_i \left(-k(\boldsymbol{\nabla} p_i) ln p_i - k p_i \frac{\boldsymbol{\nabla} p_i}{p_i} - \lambda \boldsymbol{\nabla} p_i\right) \\ &= \sum_i \left(-k(ln p_i + 1) - \lambda\right) \boldsymbol{\nabla} p_i \end{aligned} \tag{35}$$

but $\boldsymbol{\nabla} p_i$ is arbitrary, thus

$$-k(ln p_i + 1) - \lambda = 0 \Longleftrightarrow p_i = e^{-\frac{\lambda}{k} - 1} \tag{36}$$

Using our constraint $\sum_i p_i = 1$, we get

$$\sum_{i=1}^{|\chi|} e^{-\frac{\lambda}{k} - 1} = 1 \Longleftrightarrow e^{-\frac{\lambda}{k} - 1} = \frac{1}{|\chi|} \Longleftrightarrow p_i = \frac{1}{|\chi|} \tag{37}$$

This means that in order to maximize entropy, we want a uniform distribution. This gives us that a state is a maximally mixed state if every eigenvalue of the density operator is equal, i.e.,

$$\hat{\rho} = \frac{1}{|\chi|}\begin{pmatrix} 1 & 0 & ... & & 0 \\ 0 & 1 & 0 & ... & 0 \\ & & . & & \\ & & & . & \\ & & & & 1 \end{pmatrix} = \frac{1}{|\chi|}\mathcal{I} \tag{38}$$

Note that the entropy for the maximally mixed state is

$$S = -k \sum_{i=1}^{|\chi|} p_i ln p_i = -k \sum_{i=1}^{|\chi|} \frac{1}{|\chi|} ln \frac{1}{|\chi|} = k ln |\chi| \tag{39}$$

3.3 Statistical Mechanics

To some extent, statistical mechanics is an assumption about the density matrix for a macroscopic system. The assumptions (constraints) are

(a) $\sum_i p_i = 1$.
(b) $\left\langle \hat{H} \right\rangle = E$ is known.

Constraint (b) can be expressed in a more useful way since from Schroedinger equation $\hat{H} |\psi_k\rangle = E_k |\psi_k\rangle \Rightarrow \langle\psi_k| \hat{H} |\psi_k\rangle = E_k$, which gives

$$\left\langle \hat{H} \right\rangle = \sum_k p_k \langle\psi_k| \hat{H} |\psi_k\rangle = \sum_k p_k E_k = E \tag{40}$$

Since we have two constraints, using equation (32) we have that

$$\boldsymbol{\nabla}\left(-k\sum_i p_i ln p_i\right) = \lambda_1 \boldsymbol{\nabla}\left(\sum_i p_i - 1\right) + \lambda_2 \boldsymbol{\nabla}\left(\sum_i p_i E_i - E\right) \tag{41}$$

which we can simplify to

$$\boldsymbol{\nabla}\left(-k\sum_i p_i ln p_i - \lambda_1\left(\sum_i p_i - 1\right) - \lambda_2 \boldsymbol{\nabla}\left(\sum_i p_i E_i - E\right)\right) = \sum_i \left(-k(ln p_i + 1) - \lambda_1 - \lambda_2 E_i\right) \boldsymbol{\nabla} p_i \tag{42}$$

but $\boldsymbol{\nabla} p_i$ is arbitrary, thus

$$-k(ln p_i + 1) - \lambda_1 - \lambda_2 E_i = 0 \Longleftrightarrow p_i = e^{-\frac{\lambda_1}{k} - \frac{\lambda_2 E_i}{k} - 1} \tag{43}$$

Using our constraint $\sum_i p_i = 1$, we get

$$e^{-\frac{\lambda}{k} - 1} = \frac{1}{\mathcal{Z}} \tag{44}$$

with

$$\mathcal{Z} = \sum_i e^{-\frac{\lambda_2 E_i}{k}} \tag{45}$$

Letting $\frac{\lambda_2}{k} = \beta$, we get the well known canonical ensemble equations

$$p_i = \frac{e^{-\beta E_i}}{\mathcal{Z}} \tag{46}$$

and

$$\mathcal{Z} = \sum_i e^{-\beta E_i} \tag{47}$$

4 Quantum Mechanics and Statistical Mechanics

4.1 Mixed State, Pure State, and Temperature

The internal energy of a monoatomic gas is given by $E = \frac{3}{2}NkT \Rightarrow E \propto T$, but looking back at (47), we must have $\beta \propto \frac{1}{E}$ since the exponential must be unitless. These two relations imply that $\beta \propto \frac{1}{T}$. Thus,

$$T \to \infty \Rightarrow \beta \to 0^+ \Rightarrow \mathcal{Z} = |\chi| \Rightarrow p_i = \frac{1}{|\chi|}$$

This means that all states are equally probable, which is what we found for the maximally mixed state. What about $T \to 0^+$? I'm sure you can guess what is about to happen

$$\mathcal{Z} = \sum_{i=0} e^{-\beta E_i} = e^{-\beta E_0} \sum_{i=0} e^{-\beta(E_i - E_0)} = e^{-\beta E_0} + e^{-\beta E_0} \sum_{i=1} e^{-\beta(E_i - E_0)} \tag{48}$$

therefore since $E_i > E_0 \; \forall \; i > 0$, we get

$$T \to 0^+ \Rightarrow \beta \to \infty \Rightarrow Z \to e^{-\beta E_0} \tag{49}$$

this means that

$$T \to 0^+ \Rightarrow \beta \to \infty \Rightarrow p_k = \frac{e^{-\beta E_k}}{Z} \to e^{-\beta(E_k - E_0)} \tag{50}$$

thus, $p_k = 0 \; \forall \; k \neq 0$ and $p_0 = 1$. Therefore all the particles are in the ground state E_0. This exactly what we found for a pure state.

4.2 Imaginary Statistical Mechanics

Definition: (Function of a Hermitian Operator) Suppose that a Hermitian operator A has a spectral decomposition $A = \sum_i a_i |i\rangle \langle i|$ for some orthonormal basis $\{|i\rangle\}$. Then the operator $f(A)$ for some function f is defined as follows:

$$f(A) \equiv \sum_i f(a_i) |i\rangle \langle i| \tag{51}$$

Recall that $\hat{H} |\psi_k\rangle = E_k |\psi_k\rangle$, therefore $\hat{H} = \sum_i E_i |\psi_i\rangle \langle \psi_i|$. Using the above definition, this means that

$$e^{-\beta \hat{H}} = \sum_i e^{-\beta E_i} |\psi_i\rangle \langle \psi_i| \tag{52}$$

Note that this implies

$$\langle \psi_k| e^{-\beta \hat{H}} |\psi_k\rangle = \sum_i e^{-\beta E_i} \delta_{ki} \delta_{ik} = e^{-\beta E_k} \tag{53}$$

From this it is trivial to see that

$$\hat{\rho} = \frac{1}{\mathcal{Z}} e^{-\beta \hat{H}} \tag{54}$$

and

$$Z = \sum_i e^{-\beta E_i} = Tr\left(e^{-\beta \hat{H}}\right) \tag{55}$$

In quantum mechanics, we can determine how a quantum state evolves with time using the operator $U(t)$.

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle \tag{56}$$

We can get it from looking at Schroedinger equation.

$$\hat{H}\,|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}\,|\psi(t)\rangle \Rightarrow |\psi(t)\rangle = e^{\frac{-i\hat{H}t}{\hbar}}\,|\psi(0)\rangle \tag{57}$$

thus, $U(t) = e^{\frac{-i\hat{H}t}{\hbar}}$.
By making the change

$$t \rightarrow -i\beta\hbar \tag{58}$$

we get

$$U(t) \rightarrow U(-i\beta\hbar) = e^{-\beta\hat{H}} \tag{59}$$

which gives us the startling result that

$$Tr[U(-i\beta\hbar)] = Tr\left(e^{-\beta\hat{H}}\right) = \mathcal{Z} \tag{60}$$

Therefore, statistical mechanics can be perceived as quantum mechanics in imaginary time. An incredible result that might sound familiar if you have studied special relativity.

5 Conclusion

The task is, not so much to see what no one has yet seen; but to think what nobody has yet thought, about that which everybody sees.

Erwin Schroedinger (1887-1961)

Through the use of the density operator, Lagrange multipliers, and entropy, we have discovered an incredible relationship between statistical and quantum mechanics. These types of transformations are known as Wick rotations and create strange links between many areas in physics. For example, special relativity an classical physics are linked by the transformation $t \rightarrow it$. To see this, recall that the dot product invariance between two inertial frames $\mathcal{S}$ and $\mathcal{S}'$ in special relativity give $z^2 - t^2 = (z')^2 - (t')^2$ with $\mathcal{S}'$ traveling at velocity $v\hat{\mathbf{z}}$ with respect to $\mathcal{S}$. This becomes the classical Euclidean dot product $z^2 + T^2 = (z')^2 + (T')^2$ if you make the following change $t \rightarrow iT$. It is easy to see from this last equation that the two points lie on the circumference of a circle and must therefore be equivalent through a rotation, i.e.

$$\begin{pmatrix} z' \\ T' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} z \\ T \end{pmatrix} \tag{61}$$

thus,

$$\begin{aligned} z' &= z\cos\theta - T\sin\theta \\ T' &= T\cos\theta + z\sin\theta \end{aligned} \tag{62}$$

but we must have $z' = 0$ when $z = vt$, thus the first equation gives

$$\tan\theta = \frac{vt}{T} = -iv \tag{63}$$

but then $T^{'} = \cos\theta(z\tan\theta + T)$. We can simplify this by using the fact that $\sec^2\theta + \tan^2\theta = 1$ we get $\sec\theta = \sqrt{1-v^2} = \frac{1}{cos\theta}$. Thus $T^{'} = \frac{1}{\sqrt{1-v^2}}(it - izv)$. But $T^{'} = it^{'}$, so we finally get

$$
\begin{aligned}
t^{'} =& \gamma(t - vz) \\
z^{'} =& \gamma(z - vt)
\end{aligned}
\tag{64}
$$

This approach shows even more clearly that we can see the Lorentz transformations as a rotation through an imaginary angle $i\theta$ in the $cT - z$ plane, or equivalently, the $ict - z$ plane.

References

[1] Mark M. Wilde, *Quantum Information Theory*, Cambridge Univ. Press, 2017, Second Edition, pp. 101-109.

[2] Segev Benzvi "Particle Mixtures", University of Rochester, Department of Physics & Astronomy, 2017, as part of Prof. Benzvi lecture notes for Physics 407: Quantum Mechanics.