

Willoughby Seago

Theoretical Physics

Statistical Physics

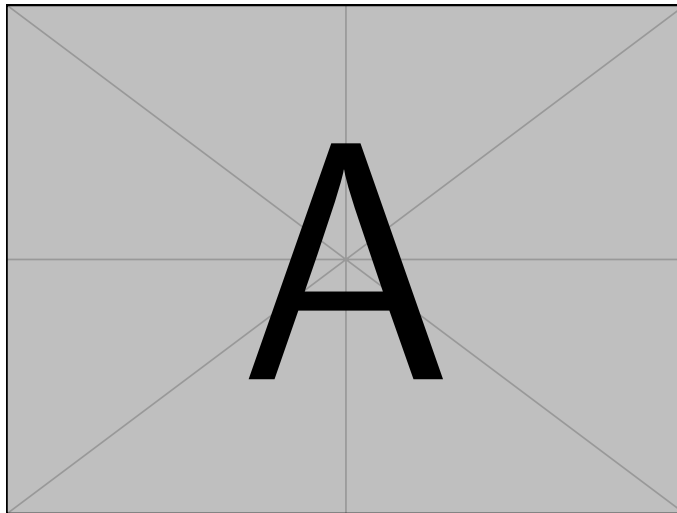
COURSE NOTES

Statistical Physics

Willoughby Seago

These are my notes from the course statistical physics. I took this course as a part of the theoretical physics degree at the University of Edinburgh.

These notes were last updated at 22:42 on February 14, 2022. For notes on other topics see <https://github.com/WilloughbySeago/Uni-Notes>.



Chapters

	Page
Chapters	ii
Contents	iii
1 Missing Information	1
2 Formulating Statistical Mechanics	7
3 Identifying the Lagrange Multipliers	13
4 Thermodynamic Potentials	19
5 Fluctuations	24
6 Weakly Interacting Systems	29
7 Dilute Limit	33
8 Density of States	35
9 Statistical Mechanics and Quantum Mechanics	38
10 Vibrations in a Solid	44
Index	52

Contents

	Page
Chapters	ii
Contents	iii
1 Missing Information	1
1.1 Probability	1
1.2 Entropy	1
1.2.1 Thermodynamics Definition	2
1.2.2 Boltzmann Entropy	2
1.2.3 Gibbs Entropy	2
1.3 Missing Information	3
2 Formulating Statistical Mechanics	7
2.1 Assignment of Probability	7
2.1.1 Lagrange Multipliers	7
2.2 Application to Statistical Mechanics	10
2.2.1 Microcanonical Ensemble	10
2.2.2 Canonical Ensemble	10
2.2.3 Grand Canonical Ensemble	11
3 Identifying the Lagrange Multipliers	13
3.1 Thermodynamics Review	13
3.2 Identifying the Lagrange Multipliers	14
3.2.1 Canonical	14
3.2.2 Grand Canonical	17
4 Thermodynamic Potentials	19
4.1 Thermodynamic Potentials	19
4.1.1 Legendre Transforms	20
4.2 Gibbs–Duhem Relation	21
4.3 Ensembles	23
5 Fluctuations	24
5.1 Energy Fluctuations	24
5.2 Magnetic Fluctuations	25

5.3	Density Fluctuations	26
5.4	General Theory	27
6	Weakly Interacting Systems	29
6.1	Localised Particles	29
6.2	Non-Localised Particles	30
6.3	Fermions and Bosons	31
6.3.1	Fermions	31
6.3.2	Bosons	32
6.3.3	Both	32
7	Dilute Limit	33
7.1	Dilute Limit	33
7.2	Semi-Classical Approximation	33
8	Density of States	35
8.1	Density of States for a Particle in a Box	35
8.1.1	Two Dimensions	36
8.1.2	Three Dimensions	36
9	Statistical Mechanics and Quantum Mechanics	38
9.1	Many Particle Schrödinger Equation	38
9.2	A Tale of Two Probabilities	40
9.3	Quantum Mechanics Recap	40
9.4	Density Matrix	41
9.4.1	Von Neumann Entropy	42
10	Vibrations in a Solid	44
10.1	Einstein Model	44
10.2	Debye Model	47
10.2.1	Debye Theory	48
10.2.2	The Integral	50
	Index	52

One

Missing Information

1.1 Probability

There are two common interpretations of probability. The first is the **frequentist** approach, in which the probability of getting outcome i , denoted p_i , is defined by

$$p_i := \lim_{N \rightarrow \infty} \frac{\text{number of times outcome } i \text{ occurs}}{N} \quad (1.1.1)$$

where N is the number of trials that we make. In this approach the probability of i is the proportion of trials that we expect to give i as a result. The frequentist definition is often the simplest, but it doesn't apply very well to one-off events.

The second approach is **Bayesian** statistics, in which p_i is a quantitative measure of the degree of belief that a rational observer has that the result of a given process will be i . This approach can apply to a one-off event.

Whichever approach we take there are some rules regarding the values that p_i can take. In particular if there are r mutually exclusive outcomes each with probability p_i for $i \in \{1, \dots, r\}$ then

- $p_i \in [0, 1]$, with 0 representing no chance of happening and 1 representing certainty.
- $p_i \text{ or } j = p_i + p_j$, this only applies to mutually exclusive outcomes.
- $\sum_{i=1}^r p_i = 1$, this is simply $p_1 \text{ or } 2 \text{ or } \dots \text{ or } r$, which is to say that the probability that one of the outcomes occurs is certain, since something must happen.
- $\langle y \rangle = \bar{y} = \sum_{i=1}^r p_i y_i$, this is the mean value of y , which is a quantity that takes the value y_i when outcome i occurs.

1.2 Entropy

There are various definitions of entropy, including, but not limited to

- a measure of disorder, and
- a measure of uncertainty.

The second definition here is more useful to us and by the end of this chapter we will have made it rigorous.

1.2.1 Thermodynamics Definition

We start with the entropy as defined in thermodynamics:

$$dS = \frac{dQ_{\text{rev}}}{T} \quad (1.2.1)$$

where dQ_{rev} is the heat absorbed by the system during a reversible process and T is the temperature. This gives the change in entropy of the system, dS . In thermodynamics we almost exclusively deal with changes in entropy rather than the absolute value of entropy. One useful thing that we find from this is that entropy has units of $[\text{energy}][\text{temperature}]^{-1}$, so typically we measure entropy in units of JK^{-1} .

Recall that the second law of thermodynamics states that the entropy of an isolated system can only increase or stay the same, that is

$$\frac{\partial S}{\partial t} \geq 0. \quad (1.2.2)$$

1.2.2 Boltzmann Entropy

The **Boltzmann entropy** is defined as

$$S_B := k_B \ln \Omega. \quad (1.2.3)$$

Here $k_B = 1.380\,649 \times 10^{-23} \text{JK}^{-1}$ (exact) is Boltzmann's constant. Recall that a macrostate describes the bulk properties of a system, such as temperature, pressure, or volume. A microstate describes the microscopic properties of a system, such as the velocity or energy of all the particles. In general for each macrostate there are multiple microstates which the system could be in and still have the same macroscopic properties. The quantity Ω is then the **weight** of the macrostate, which is the number of microstates that correspond to the macrostate.

We expect that the entropy is an extrinsic property, that is the more matter we have the greater the entropy. In particular we expect that the entropy is of the form $S = sN$, where N is the number of constituents in the system, for example the number of particles. We then have

$$\Omega = \exp\left(\frac{sN}{k_B}\right). \quad (1.2.4)$$

So Ω is exponential in N . Therefore the larger N is the larger Ω is, and hence the larger S_B is. This means we can interpret S_B as a measure of the uncertainty about which microstate the system is in.

1.2.3 Gibbs Entropy

The **Gibbs entropy** is defined as

$$S_G := -k_B \sum_{\text{microstates}, i} p_i \ln p_i \quad (1.2.5)$$

where the sum is over microstates and p_i is the probability that the system is in microstate i .

We can show that both the Boltzmann and Gibbs entropies agree when we have an isolated system in a given macrostate with Ω microstates. With no further information the only sensible choice is $p_i = 1/\Omega$, this is called the principle of equal

a priori probabilities. The sum for the Gibbs entropy then runs over all microstates associated with the given macrostate, meaning that it goes from 1 to Ω . Hence we have

$$S_G = -k_B \sum_{i=1}^{\Omega} \frac{1}{\Omega} \ln \frac{1}{\Omega} \quad (1.2.6)$$

$$= k_B \sum_{i=1}^{\Omega} \frac{1}{\Omega} \ln \Omega \quad (1.2.7)$$

$$= k_B \ln \Omega \quad (1.2.8)$$

$$= S_B. \quad (1.2.9)$$

Here we have used the fact that the terms in the sum are constant so the sum from 1 to Ω is just Ω times this constant.

The Gibbs entropy uses the probability of microstates, p_i , whereas the Boltzmann entropy uses the weight of a macrostate. This means that the Gibbs entropy is a microscopic picture of entropy whereas the Boltzmann entropy is a macroscopic picture. This makes the Gibbs entropy a more appropriate definition for developing a fundamental understanding of entropy.

The Gibbs entropy also has the advantage of applying to systems which are not large, this allows us to break systems up into smaller subsystems. The Gibbs entropy also applies to systems which are not in thermal equilibrium.

1.3 Missing Information

Information is a quantity measured in bits. A bit is a quantity that can take one of two value, usually we think of these as 0 or 1, or in physics we may consider spins being up or down. If a system has a single bit that can take either value then we have missing information, the value of the bit. Suppose we then apply some process which fixes the value of the bit, such as applying a magnetic field such that the spin aligns with the field. We then have gained some information since we now know the value of the bit, we have gained 1 bit of information. If we had started off knowing the value of the bit then we would not have gained any information.

We wish to generalise this idea of missing some information to systems of more than two states where the probabilities of different states are not necessarily equal.

In order to find the correct function to determine the missing information we now list some requirements of this function. To do so we consider the missing information of a system with r mutually exclusive outcomes, each with probability p_i of occurring.

1. The missing information should be a continuous function of the probabilities, p_i . This means that changing one of the probabilities by only a very small amount should change the missing information by only a small amount, which makes intuitive sense. This means we are looking for a continuous function of the form

$$S: [0, 1]^r \rightarrow \mathbb{R}. \quad (1.3.1)$$

2. The missing information should be symmetric in the probabilities. This means if we label the probabilities differently the missing information doesn't change.

This makes sense since the missing information should be independent of our choices, such as the way we order outcomes.

3. For the case where $p_1 = \dots = p_r = 1/r$ the missing information should reduce to an increasing function of r . That is for equally likely outcomes the more possible outcomes there are the more information will be missing, since there is more uncertainty.
4. The missing information should not change based on how we group the outcomes.

This last point is non-trivial and we will expand upon it. We can divide the outcomes into n groups labelled $j = 1, \dots, n$, such that each group has r_j outcomes and the probability of an outcome in group j is w_j . We should then be able to write the missing information

$$S(\{p\}_r) = S(\{w\}_n) + w_1 S\left(\frac{p_1}{w_1}, \dots, \frac{p_{r_1}}{w_1}\right) + w_2 S\left(\frac{p_{r_1+1}}{w_2}, \dots, \frac{p_{r_1+r_2}}{w_2}\right) + \dots \quad (1.3.2)$$

$$= S(\{w\}_n) + \sum_{j=1}^n w_j S\left(\frac{p_{r_1+\dots+r_{j-1}+1}}{w_j}, \dots, \frac{p_{r_1+\dots+r_j}}{w_j}\right). \quad (1.3.3)$$

The interpretation here is that the first term, $S(\{w\}_n)$ deals with the missing information about which group the outcome is in, and then the relevant second term deals with the missing information about which element within that group the outcome is.

This becomes clearer with an example. Suppose that we have three outcomes, say possible energies of a particle, and these outcomes have associated probabilities p_1 , p_2 , and p_3 . Suppose that outcomes 1 and 2 differ only due to the spin of the particle, and that there is a second observer who can't make measurements as precisely and therefore can't differentiate between the first two outcomes. In this case all they can do is ascribe a single probability, $p_1 + p_2$, that one of the first two outcomes occurs. The missing information is then

$$S(p_1, p_2, p_3) = S(w_1, w_2) + w_1 S\left(\frac{p_1}{w_1} + \frac{p_2}{w_1}\right) + S\left(\frac{p_3}{w_2}\right) \quad (1.3.4)$$

$$= S(w_1, w_2) + w_1 S\left(\frac{p_1}{w_1} + \frac{p_2}{w_1}\right) \quad (1.3.5)$$

where $w_1 = p_1 + p_2$ and $w_2 = p_3$. We have then used $S(p_3/w_2) = S(1) = 0$.

For an even more explicit example consider a procedure with four possible outcomes with probabilities $p_1 = 1/6$, $p_2 = 1/3$, and $p_3 = p_4 = 1/4$. Suppose that the first two outcomes can be grouped together as can the second two. Then we expect

$$S\left(\frac{1}{6}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4}\right) = S\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} S\left(\frac{1/6}{1/2}, \frac{1/3}{1/2}\right) + \frac{1}{2} S\left(\frac{1/4}{1/2}, \frac{1/4}{1/2}\right) \quad (1.3.6)$$

$$= S\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} S\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{1}{2} S\left(\frac{1}{2}, \frac{1}{2}\right). \quad (1.3.7)$$

We make the following ansatz for the form of S :

$$S(\{p\}_r) = \sum_{i=1}^r \varphi(p_i) \quad (1.3.8)$$

for some function φ . The fact that addition is commutative takes care of the requirement that S be symmetric. Requiring φ to be continuous also takes care of our continuity conditions.

Noticing that adding in outcomes with zero probability cannot change the information content, since we know for certain that these outcomes won't occur, we see that we must have $\varphi(0) = 0$. Similarly if one outcome is certain, so $p_i = 1$ for some fixed value of i , and all other outcomes cannot occur, so $p_j = \delta_{ij}$, then we also have no missing information since the outcome is certain and so we must have $\varphi(1) = 0$ also so that the sum is zero.

Considering the case where $p_i = 1/r$ we have

$$S(\{p\}_r) = \sum_{i=1}^r \varphi\left(\frac{1}{r}\right) = r\varphi\left(\frac{1}{r}\right). \quad (1.3.9)$$

Now dividing the outcome into n groups with m outcomes, so $r = mn$, we have that the probability of being in the j th group is $w_j = 1/n = m/r$. We then have

$$S(\{w\}_n) = \sum_{j=1}^n \varphi\left(\frac{1}{n}\right) = n\varphi\left(\frac{1}{n}\right) \quad (1.3.10)$$

and

$$\sum_{j=1}^n w_j S\left(\frac{\{p\}_m}{w_j}\right) = \sum_{j=1}^n \frac{1}{n} \varphi\left(\frac{n}{r}\right) \quad (1.3.11)$$

$$= n \frac{1}{n} \varphi\left(\frac{n}{r}\right) \quad (1.3.12)$$

$$= m\varphi\left(\frac{1}{m}\right). \quad (1.3.13)$$

Hence Equation (1.3.3) becomes

$$r\varphi\left(\frac{1}{r}\right) = \frac{r}{m} \varphi\left(\frac{m}{r}\right) + m\varphi\left(\frac{1}{m}\right). \quad (1.3.14)$$

This isn't simple to solve, fortunately others have solved it and we can just check their solution:

$$\varphi(x) = -kx \ln x \quad (1.3.15)$$

for some constant k . With this form the left hand side becomes

$$r\varphi\left(\frac{1}{r}\right) = r \left(-k \frac{1}{r} \ln \frac{1}{r} \right) = k \ln r. \quad (1.3.16)$$

The right hand side becomes

$$\frac{r}{m} \varphi\left(\frac{m}{r}\right) + m\varphi\left(\frac{1}{m}\right) = \frac{r}{m} \left(-k \frac{m}{r} \ln \frac{m}{r} \right) + m \left(-k \frac{1}{m} \ln \frac{1}{m} \right) \quad (1.3.17)$$

$$= k \ln \frac{r}{m} + k \ln \frac{1}{m} = k \ln r. \quad (1.3.18)$$

So we see that this is indeed a solution, We also have

$$\varphi(1) = -k \ln 1 = 0 \quad (1.3.19)$$

and

$$\lim_{x \rightarrow 0} -kx \ln x = 0 \quad (1.3.20)$$

also. This therefore gives all the properties required for the missing information function:

$$S(\{p\}_r) = -k \sum_{i=1}^r p_i \ln p_i. \quad (1.3.21)$$

It can be shown that this is the only function satisfying the necessary properties, up to the value of k . We take k to be positive so that S is non-negative, since $p_i \in [0, 1]$ and so $\ln p_i \leq 0$. It can be shown that S is additive, that S is maximised when all probabilities are equal, which is when we know the least about the system, and that $S = 0$ when $p_j = \delta_{ij}$ for some fixed i , which is when we know everything about the system.

To uniquely specify S all we need to do is pick a value for k . When Claude Shannon first found this function he was studying information theory. For this reason he chose $k = 1/\ln 2$. His reasoning being if he had a string of B bits, each being 0 or 1, so some element of $\{0, 1\}^B$, then the total number of possible states for this string is 2^B . If all states are equally likely then we have

$$S = -k \sum_{i=1}^{2^B} \frac{1}{2^B} \ln \frac{1}{2^B} = k \sum_{i=1}^{2^B} \frac{1}{2^B} \ln 2^B = k 2^B \frac{1}{2^B} \ln 2^B = k \ln 2^B = Bk \ln 2 \quad (1.3.22)$$

and so the choice of $k = 1/\ln 2$ gives the missing information as $S = B$, which is measured in bits. In this case we call this the **Shannon entropy**.

On the other hand in statistical mechanics we typically take $k = k_B$, which has units of $[k_B] = [\text{energy}]/[\text{temperature}]$, most commonly J K^{-1} . This gives

$$S = -k_B \sum_{i=1}^r p_i \ln p_i = S_G. \quad (1.3.23)$$

That is with this choice we have that the missing information is exactly the Gibbs entropy. This is the precise meaning in the statement “entropy is a measure of disorder”, entropy is a measure of how much information we are missing from a system, which relates to the disorder since the more ordered a system is the more we know and the less information we are missing.

Two

Formulating Statistical Mechanics

2.1 Assignment of Probability

Suppose that we have a system with some constraints, for example the energy may be fixed, or the average of some value might be fixed. We want to assign probabilities, $\{p\}_r$, to the states. The values of these probabilities should reflect only the information that we have available. Otherwise there would be some bias in the assignment and we cannot then argue that it is a rational assignment of the probabilities. To ensure that we aren't accidentally including more information in the assignment than we actually know we should aim to maximise the missing information. This leads to the following principle.

Probabilities should be assigned to states such as to maximise S subject to known constraints.

For our purposes constraints usually take the form of fixed expectation values of some observables. We also always have the constraint that

$$\sum_{i=1}^r p_i = 1. \quad (2.1.1)$$

We can maximise a quantity subject to constraint's using Lagrange multipliers.

2.1.1 Lagrange Multipliers

R For more details about Lagrange multipliers see the notes from the Lagrangian dynamics course.

Suppose we wish to extremise some function, f , which is a function of independent variables, x_i . We want to find a point where f is stationary so we need

$$df = \sum_{i=1}^r \frac{\partial f}{\partial x_i} dx_i = 0. \quad (2.1.2)$$

Suppose also that we have the constraint that $g(\{x\}_r) = g_0 = \text{constant}$ for some function g . We cannot therefore assume that $\partial f / \partial x_i = 0$ gives the desired point.

What we do is construct the function $h(\{x\}_r) = f(\{x\}_r) - \lambda g(\{x\}_r)$ where λ is some constant. We then have

$$dh = df - \lambda dg = \sum_{i=1}^r \left(\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} \right) dx_i = 0. \quad (2.1.3)$$

We can choose λ such that $\partial f / \partial x_r - \lambda \partial g / \partial x_r = 0$. Since x_i are independent variables we can vary them separately and the sum must still give zero, it follows then that for this same value of λ we will have $\partial f / \partial x_i - \lambda \partial g / \partial x_i = 0$.

The general method then to minimise f subject to the constraint that $g(\{x\}_r) = g_0$ is to define

$$h(\{x\}_r) = f(\{x\}_r) - \lambda g(\{x\}_r). \quad (2.1.4)$$

Then extremise h by requiring that

$$\frac{\partial h}{\partial x_i} = 0 \quad (2.1.5)$$

for all x_i .

We apply this method with x_i being the probabilities p_i , and f being the entropy, S . We then choose λ such that the constrained quantities have the desired values. If there are several constraints then each constraint gets its own Lagrange multiplier.

It is easiest to see why this works geometrically. Suppose we want to extremise $f(\mathbf{x})$ subject to the constraint $g(\mathbf{x}) = 0$, if instead the constraint is $g(\mathbf{x}) = g_0$ we can redefine g , so $g(\mathbf{x}) \rightarrow g(\mathbf{x}) - g_0$, to get the desired form. Recall that $df = d\mathbf{x} \cdot \nabla f$. We must also choose $d\mathbf{x}$ such that g doesn't change. This means that $d\mathbf{x}$ must be along the level surfaces of g . The level surfaces of g are perpendicular to ∇g , therefore $d\mathbf{x}$ must be perpendicular to ∇g . This means that we must have $d\mathbf{x} \cdot \nabla g = 0$.

Therefore at the extremum we have both ∇g and ∇f parallel, which means that for some value of λ we must have $\nabla f - \lambda \nabla g = 0$. In component form this is

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} = 0, \quad (2.1.6)$$

which is exactly what we had before.

■ **Example 2.1.7** Suppose we have no constraints other than the mandatory $\sum_i p_i = 1$. We then have

$$h(\{p\}_r) = -k \sum_i p_i \ln p_i - \lambda \sum_i p_i. \quad (2.1.8)$$

Taking the derivative we use

$$\frac{\partial}{\partial x} x \ln x = \ln x + 1 \quad (2.1.9)$$

and so

$$\frac{\partial h}{\partial p_j} = -k \ln p_j - k - \lambda = 0 \quad (2.1.10)$$

since all terms with p_i for $i \neq j$ vanish in the derivative. Rearranging we have

$$p_j = \exp \left[-1 - \frac{\lambda}{k} \right]. \quad (2.1.11)$$

Noticing that the right hand side here is constant we have

$$1 = \sum_{i=1}^r p_i = \sum_{i=1}^r \exp \left[-1 - \frac{\lambda}{k} \right] = r \exp \left[-1 - \frac{\lambda}{k} \right]. \quad (2.1.12)$$

Rearranging this we have

$$\frac{1}{r} = \exp \left[-1 - \frac{\lambda}{k} \right] = p_i. \quad (2.1.13)$$

So we see that we recover the principle of equal *a priori* probabilities.

■ **Example 2.1.14** Suppose we have two constraints. Namely that the expectation values of two observables, y and z , are fixed. That is

$$\langle y \rangle = \sum_i p_i y_i, \quad \text{and} \quad \langle z \rangle = \sum_i p_i z_i \quad (2.1.15)$$

are fixed values. As well as this we have the requirement that $\sum_i p_i = 1$. We then define

$$h(\{p\}_r) = -k \sum_i p_i \ln p_i - \lambda_1 \sum_i p_i - \lambda_y \sum_i p_i y_i - \lambda_z \sum_i p_i z_i. \quad (2.1.16)$$

Differentiating we have

$$\frac{\partial h}{\partial p_j} = -k \ln p_j - k - \lambda_1 - \lambda_y y_j - \lambda_z z_j = 0. \quad (2.1.17)$$

Thus,

$$p_j = \exp \left[-1 - \frac{\lambda_1}{k} - \frac{\lambda_y y_j}{k} - \frac{\lambda_z z_j}{k} \right]. \quad (2.1.18)$$

Now considering the requirement that $\sum_j p_j = 1$ we have

$$1 = \sum_j p_j \quad (2.1.19)$$

$$= \sum_j \exp \left[-1 - \frac{\lambda_1}{k} - \frac{\lambda_y y_j}{k} - \frac{\lambda_z z_j}{k} \right] \quad (2.1.20)$$

$$= \exp \left[-1 - \frac{\lambda_1}{k} \right] \sum_j \exp \left[-\frac{\lambda_y y_j}{k} - \frac{\lambda_z z_j}{k} \right] \quad (2.1.21)$$

$$= Z \exp \left[-1 - \frac{\lambda_1}{k} \right] \quad (2.1.22)$$

where we have defined

$$Z := \sum_j \exp \left[-\frac{\lambda_y y_j}{k} - \frac{\lambda_z z_j}{k} \right]. \quad (2.1.23)$$

Rearranging Equation (2.1.22) we have

$$\exp \left[-1 - \frac{\lambda_1}{k} \right] = \frac{1}{Z}. \quad (2.1.24)$$

We then have

$$p_j = \exp \left[-1 - \frac{\lambda_1}{k} - \frac{\lambda_y y_j}{k} - \frac{\lambda_z z_j}{k} \right] \quad (2.1.25)$$

$$= \exp \left[-1 - \frac{\lambda_1}{k} \right] \exp \left[-\frac{\lambda_y y_j}{k} - \frac{\lambda_z z_j}{k} \right] \quad (2.1.26)$$

$$= \frac{1}{Z} \exp \left[-\frac{\lambda_y y_j}{k} - \frac{\lambda_z z_j}{k} \right]. \quad (2.1.27)$$

We can fix the values of λ_y and λ_z using the constraints that $\langle y \rangle$ and $\langle z \rangle$ are known, fixed values.

2.2 Application to Statistical Mechanics

Recall that a microstate is the most detailed description possible of a system. Typically this will correspond to the i th microstate being the i th solution to the Schrödinger equation. Let E_i be the energy of the i th microstate, this will typically be a function of extensive thermodynamic properties, such as volume.

The equilibrium state is specified by the expectation values of extensive observables, such as the internal energy, the expectation value of which is

$$\bar{E} = \sum_i p_i E_i. \quad (2.2.1)$$

Note that it is common to drop the overline and simply write E for this quantity.

From here the exact physics depends on which constraints are imposed. We classify the constraints into categories of ensembles, which for now we can think of as being synonymous with probability distributions.

2.2.1 Microcanonical Ensemble

A **microcanonical ensemble** is a completely isolated system. This means that the energy is fixed and so all microstates must have the same energy, $E_i = \bar{E}$. The only constraint therefore is that $\sum_i p_i = 1$. This is analogous to [Example 2.1.7](#) and so

$$p_i = \frac{1}{\Omega} \quad (2.2.2)$$

where Ω is the number of microstates, which we called r in [Example 2.1.7](#). So maximising S in a microcanonical ensemble corresponds to the principle of equal *a priori* probabilities.

2.2.2 Canonical Ensemble

A **canonical ensemble** can explore states of different energies, E_i , we can think of it as being isolated but in contact with some heat reservoir which allows it to change energy. The observable \bar{E} specifies the equilibrium state. Therefore we need to maximise S subject to the constraint that \bar{E} is fixed, as well as the constraint that $\sum_i p_i = 1$. We therefore have

$$h(\{p\}_r) = -k \sum_i p_i \ln p_i - \lambda_1 \sum_i p_i - \lambda_E \sum_i p_i E_i. \quad (2.2.3)$$

Extremising this we have

$$\frac{\partial h}{\partial p_j} = -k \ln p_j - k - \lambda_1 - \lambda_E E_j = 0. \quad (2.2.4)$$

Rearranging this gives

$$p_j = \exp \left[-1 - \frac{\lambda_1}{k} - \frac{\lambda_E E_j}{k} \right]. \quad (2.2.5)$$

The constraint that $\sum_j p_j = 1$ gives us

$$1 = \sum_j p_j = \sum_j \exp \left[-1 - \frac{\lambda_1}{k} - \frac{\lambda_E E_j}{k} \right] \quad (2.2.6)$$

$$= \exp \left[-1 - \frac{\lambda_1}{k} \right] \sum_j \exp \left[-\frac{\lambda_E E_j}{k} \right] \quad (2.2.7)$$

$$= Z_c \exp \left[-1 - \frac{\lambda_1}{k} \right] \quad (2.2.8)$$

where we have defined

$$Z_c := \sum_j \exp \left[-\frac{\lambda_E E_j}{k} \right]. \quad (2.2.9)$$

We therefore have

$$\exp \left[-1 - \frac{\lambda_1}{k} \right] = \frac{1}{Z_c}. \quad (2.2.10)$$

It follows that

$$p_j = \exp \left[-1 - \frac{\lambda_1}{k} - \frac{\lambda_E E_j}{k} \right] \quad (2.2.11)$$

$$= \exp \left[-1 - \frac{\lambda_1}{k} \right] \exp \left[-\frac{\lambda_E E_j}{k} \right] \quad (2.2.12)$$

$$= \frac{1}{Z_c} \exp \left[-\frac{\lambda_E E_j}{k} \right]. \quad (2.2.13)$$

Here Z_c is the **canonical partition function**. We will see later that we can identify λ_E in such a way that we recover the Boltzmann distribution.

2.2.3 Grand Canonical Ensemble

A **grand canonical ensemble** can explore states of different energies, but also different numbers of particles. We can think of the system as being in contact with a heat reservoir and a reservoir of particles. We now need to label states both by i , which corresponds to the energy, E_i , but also by the number of particles, N , since in general if N changes then E_i will change also.

The constraints are then that the following are fixed

$$\bar{E} = \sum_{i,N} p_{iN} E_{iN}, \quad \text{and} \quad \bar{N} = \sum_{i,N} p_{iN} N. \quad (2.2.14)$$

As well as the normal $\sum_{i,N} p_{iN} = 1$. This corresponds to [Example 2.1.14](#) and we identify

$$p_{iN} = \frac{1}{\mathcal{Z}_{\text{gc}}} \exp \left[-\frac{\lambda_E E_{iN}}{k} - \frac{\lambda_N N}{k} \right] \quad (2.2.15)$$

where

$$\mathcal{Z}_{\text{gc}} := \sum_{i,N} \exp \left[-\frac{\lambda_E E_{iN}}{k} - \frac{\lambda_N N}{k} \right] \quad (2.2.16)$$

is the **grand canonical partition function**.

The next step is to identify the physical interpretation of the Lagrange multipliers. This will be the focus of the next chapter.

Three

Identifying the Lagrange Multipliers

3.1 Thermodynamics Review



For more details on thermodynamics see the Thermodynamics notes from the Thermal Physics course.

In order to identify the physical meaning of the Lagrange multipliers we will need to recap some basic thermodynamics. Equilibrium thermodynamic variables are variables which are stationary when the system is in equilibrium and have well defined values, we call these functions of state. The equation of state for a system relates different thermodynamic variables, for example $PV = nRT$ relates the pressure, P , volume, V , number of moles, n , and temperature, T .

In statistical mechanics we replace thermodynamic variables with their expectation values. This is valid since in the thermodynamic limit, where the number of constituents, N , tends to infinity, we get distributions which are sharply peaked about the expected value.

The zeroth law of thermodynamics is that heat flows from a hotter body to a colder body. This defines what we mean when we say hot and cold and can be used to define temperature.

The first and second laws of thermodynamics can be combined into

$$d\bar{E} = T dS - P dV \quad (3.1.1)$$

where \bar{E} is the mean (internal) energy (denoted U in thermodynamics), T is the temperature, S is the entropy, P is the pressure and V is the volume. This form of the first and second laws is valid for a PVT system, which is a system where the thermodynamic variables are pressure, volume, and temperature.

A more general form of this law is

$$d\bar{E} = T dS + \sum_Y f_Y dX_Y. \quad (3.1.2)$$

Here f_Y is what we call a **thermodynamic force**, or a **generalised force**¹. X_Y is then the **thermodynamic displacement**, also known as the **conjugate field**. These quantities have units such that their product has dimensions of energy.

We've already seen the example of a PVT system where we can identify $-P$ as a generalised force with the associated conjugate variable V . Another example would be the magnetic field, $\mu_0 H$, with the conjugate field being the magnetisation, M . In

¹for more on generalised forces and conjugate variables see the notes from the Lagrangian Dynamics course.

this case the term that contributes to $d\bar{E}$ is $\mu_0 \mathbf{H} \cdot \mathbf{M}$, that is each component of $\mu_0 \mathbf{H}$ and \mathbf{M} acts as a conjugate pair.

Importantly the forces are **intensive**, meaning they don't depend on the size of the system, and the displacements are **extensive**, meaning they scale linearly with the size of the system.

Being even more general the total change in internal energy is

$$d\bar{E} = T dS + \sum_Y f_Y dX_Y + \sum_\alpha \mu_\alpha d\bar{N}_\alpha \quad (3.1.3)$$

²this is just a fancy word for “type of particle” which allows for us to talk about atoms, molecules, colloidal particles, etc. using the same word.

where \bar{N}_α is the mean number of particles of species² α and μ_α is the associated **chemical potential**. We can take this equation to define μ_α and so we see that

$$\mu_\alpha := \left(\frac{\partial \bar{E}}{\partial \bar{N}_\alpha} \right)_{S, \{X\}} \quad (3.1.4)$$

where the notation

$$\left(\frac{\partial f}{\partial x} \right)_y \quad (3.1.5)$$

means the derivative of f with respect to x holding y constant.

The final takeaway is that the entropy, S , conjugate fields, $\{X\}$, and number of particles, $\{N\}$, are the natural variables for \bar{E} , meaning that $\bar{E} = \bar{E}(S, \{X\}, \{N\})$.

3.2 Identifying the Lagrange Multipliers

3.2.1 Canonical

Recall that for the canonical ensemble

$$p_i = \frac{1}{Z_c} \exp \left[-\frac{\lambda_E E_i}{k} \right], \quad \text{where} \quad Z_c = \sum_i \exp \left[-\frac{\lambda_E E_i}{k} \right]. \quad (3.2.1)$$

We take $k = k_B$, the Boltzmann constant, in order to get results familiar from thermodynamics. We start by calculating $d\bar{E}$. Starting with the definition

$$\bar{E} := \sum_i p_i E_i \quad (3.2.2)$$

we see that we can change \bar{E} in two ways. First, we could change p_i , second we can change E_i , to change E_i we have to change properties of the system. Say we change the system by dX_Y , then there will be an associated change in E_i , and hence \bar{E} .

Putting this together

$$d\bar{E} = d \left(\sum_i E_i p_i \right) \quad (3.2.3)$$

$$= \sum_i \frac{\partial \bar{E}}{\partial p_i} dp_i + \sum_Y \frac{\partial \bar{E}}{\partial X_Y} dX_Y. \quad (3.2.4)$$

Identifying

$$\frac{\partial \bar{E}}{\partial p_i} = \frac{\partial}{\partial p_i} \sum_j E_j p_j = E_i \quad (3.2.5)$$

we can write this as

$$d\bar{E} = \sum_i E_i dp_i + \sum_Y \frac{\partial \bar{E}}{\partial X_Y} dX_Y. \quad (3.2.6)$$

Now consider the entropy

$$S = -k_B \sum_i p_i \ln p_i. \quad (3.2.7)$$

The change in entropy for a change in probabilities is given by the chain rule as

$$dS = \sum_i \frac{\partial S}{\partial p_i} dp_i = -k_B \sum_i (\ln p_i + 1) dp_i. \quad (3.2.8)$$

Since the total probability must stay constant we must have $\sum_i dp_i = 0$ and hence the second term in this sum vanishes and we have

$$dS = -k_B \sum_i \ln(p_i) dp_i. \quad (3.2.9)$$

Substituting in

$$p_i = \frac{1}{Z_c} \exp \left[-\frac{\lambda_E E_i}{k_B} \right] \quad (3.2.10)$$

we get

$$dS = -k_B \sum_i \left[-\frac{\lambda_E E_i}{k_B} - \ln Z_c \right] dp_i. \quad (3.2.11)$$

The second term vanishes again since $\ln Z_c$ is a constant and the sum of the changes in probabilities must vanish to maintain constant total probability. We therefore have

$$dS = \lambda_E \sum_i E_i dp_i. \quad (3.2.12)$$

Substituting this result into [Equation \(3.2.6\)](#) we get

$$d\bar{E} = \frac{1}{\lambda_E} dS = \sum_Y \frac{\partial \bar{E}}{\partial X_Y} dX_Y. \quad (3.2.13)$$

Comparing this to

$$d\bar{E} = T dS + \sum_Y f_Y dX_Y \quad (3.2.14)$$

we identify

$$\lambda_E = \frac{1}{T}. \quad (3.2.15)$$

We also get an expression for the thermodynamic force:

$$f_Y = \frac{\partial \bar{E}}{\partial X_Y}. \quad (3.2.16)$$

For example,

$$-P - \frac{\partial \bar{E}}{\partial V} = \sum_i p_i \frac{\partial E_i}{\partial V}. \quad (3.2.17)$$

We can use this to identify the **instantaneous pressure**

$$P_i := -\frac{\partial E_i}{\partial V} \quad (3.2.18)$$

so that the mean pressure takes the expected form

$$P = \sum_i p_i P_i. \quad (3.2.19)$$

Now that we have identified the Lagrange multiplier we can write the canonical distribution in its usual form:

$$p_i = \frac{1}{Z_c} \exp \left[-\frac{E_i}{k_B T} \right], \quad \text{where} \quad Z_c = \sum_i \exp \left[-\frac{E_i}{k_B T} \right]. \quad (3.2.20)$$

Introducing $\beta := 1/(k_B T)$, which we will use throughout this course as we see fit, we can write this as

$$p_i = \frac{1}{Z_c} e^{-\beta E_i}, \quad \text{where} \quad Z_c = \sum_i e^{-\beta E_i}. \quad (3.2.21)$$

Considering again the entropy we have

$$S = -k_B \sum_i p_i \ln p_i \quad (3.2.22)$$

$$= -k_B \sum_i p_i [-\beta E_i - \ln Z_c] \quad (3.2.23)$$

$$= -k_B \sum_i p_i \left[-\frac{E_i}{k_B T} - \ln Z_c \right] \quad (3.2.24)$$

$$= \frac{1}{T} \sum_i p_i E_i + k_B \ln Z_c \sum_i p_i. \quad (3.2.25)$$

Identifying the final sum as 1 and rearranging this we see that

$$-k_B T \ln Z_c = \bar{E} - TS =: F \quad (3.2.26)$$

where F is the **Helmholtz free energy**, which we will discuss more in the next chapter. This equation is called a **bridge equation** because the left hand side is in terms of the partition function, which is a sum over microstates, and the right hand side is in terms of macroscopic properties such as the energy. Notice that

$$dF = d\bar{E} - T dS - S dT = -S dT + \sum_Y f_Y dX_Y. \quad (3.2.27)$$

Here we used the chain rule to compute $d(ST)$ and then recognised

$$d\bar{E} - T dS = \sum_Y f_Y dX_Y \quad (3.2.28)$$

as a rearrangement of the first and second law. From this we see that T and X_Y are the natural variables to express F .

3.2.2 Grand Canonical

Recall that for the grand canonical ensemble

$$p_{iN} = \frac{1}{\mathcal{Z}_{gc}} \exp \left[-\frac{\lambda_E E_{iN}}{k} - \frac{\lambda_N N}{k} \right], \quad \text{where} \quad \mathcal{Z}_{gc} = \sum_{i,N} \exp \left[-\frac{\lambda_E E_{iN}}{k} - \frac{\lambda_N N}{k} \right]. \quad (3.2.29)$$

We take $k = k_B$, the Boltzmann constant, in order to get results familiar from thermodynamics. We start by calculating $d\bar{E}$. By the same logic as the canonical case

$$d\bar{E} = \sum_{i,N} \frac{\partial \bar{E}}{\partial p_{iN}} dp_{iN} + \sum_Y \frac{\partial \bar{E}}{\partial X_Y} dX_Y \quad (3.2.30)$$

$$= \sum_{i,N} E_{iN} dp_{iN} + \sum_Y \frac{\partial \bar{E}}{\partial X_Y} dX_Y. \quad (3.2.31)$$

Now consider the entropy. The change in entropy for a change in probabilities is

$$dS = \sum_{i,N} \frac{\partial S}{\partial p_{iN}} dp_{iN} \quad (3.2.32)$$

$$= -k_B \sum_{i,N} (\ln p_{iN} + 1) dp_{iN} \quad (3.2.33)$$

$$= -k_B \sum_{i,N} \ln(p_{iN}) dp_{iN} \quad (3.2.34)$$

$$= -k_B \sum_{i,N} \left[-\frac{\lambda_E E_{iN}}{k_B} - \frac{\lambda_N N}{k_B} - \ln \mathcal{Z}_{gc} \right] dp_{iN} \quad (3.2.35)$$

$$= \lambda_E \sum_{i,N} E_{iN} dp_{iN} + \lambda_N \sum_{i,N} N dp_{iN}. \quad (3.2.36)$$

Here we substituted in the definition of S to get to the second line. We then identified that $\sum_{i,N} p_{iN} = 1$ and so $\sum_{i,N} dp_{iN} = 0$ so we can neglect the second term in the second line. In the fourth line we substituted in the definition of p_{iN} in the grand canonical distribution to get the fourth line. We then identify that $\ln \mathcal{Z}_{gc}$ is constant and so the final term of the fourth line vanishes when we sum over dp_{iN} .

We now write

$$d\bar{N} = \sum_{i,N} N dp_{iN} \quad (3.2.37)$$

which follows from

$$\bar{N} = \sum_{i,N} p_{iN} N \quad (3.2.38)$$

and noticing that N is the index of the sum, so is constant in any given term, we have

$$\frac{\partial \bar{N}}{\partial p_{iN}} = N p_{iN}. \quad (3.2.39)$$

We can then write

$$dS = \lambda_E \sum_{i,N} E_{iN} dp_{iN} + \lambda_N d\bar{N}. \quad (3.2.40)$$

Rearranging this we get

$$d\bar{E} = \frac{1}{\lambda_E} dS - \frac{\lambda_N}{\lambda_E} d\bar{N}. \quad (3.2.41)$$

Comparing this to the most general form of the first and second law,

$$d\bar{E} = T dS + \sum_Y f_Y dX_Y + \mu dN \quad (3.2.42)$$

for a single particle species we can identify

$$\lambda_E = \frac{1}{T}, \quad \text{and} \quad \lambda_N = -\lambda_E \mu = -\frac{\mu}{T}. \quad (3.2.43)$$

We have found the standard form of the grand canonical ensemble distribution:

$$p_{iN} = \frac{1}{\mathcal{Z}_{\text{gc}}} e^{-\beta(E_{iN} - N\mu)}, \quad \text{where} \quad \mathcal{Z}_{\text{gc}} = \sum_{i,N} e^{-\beta(E_{iN} - N\mu)}. \quad (3.2.44)$$

Considering the entropy we have

$$S = \sum_{i,N} p_{iN} \ln p_{iN} \quad (3.2.45)$$

$$= -k_B \sum_{i,N} p_{iN} [-\beta(E_{iN} - N\mu) - \ln \mathcal{Z}_{\text{gc}}] \quad (3.2.46)$$

$$= \frac{1}{T} \sum_{i,N} p_{iN} E_{iN} - \frac{\mu}{T} \sum_{i,N} p_{iN} N - k_B \ln \mathcal{Z}_{\text{gc}} \sum_{i,N} p_{iN} \quad (3.2.47)$$

$$= \frac{\bar{E}}{T} - \mu \bar{N} + k_B \ln \mathcal{Z}_{\text{gc}}. \quad (3.2.48)$$

Rearranging this we have

$$-k_B T \ln \mathcal{Z}_{\text{gc}} = \bar{E} - TS - \mu \bar{N} = \Phi \quad (3.2.49)$$

where Φ is the **grand potential**, which we will discuss more in the next chapter. This is another bridge equation relating microscopic and macroscopic quantities.

Four

Thermodynamic Potentials

4.1 Thermodynamic Potentials

Recall that for a PVT system with a single species of particle the first and second laws of thermodynamics combine to give the central equation

$$d\bar{E} = T dS - P dV + \mu d\bar{N}. \quad (4.1.1)$$

There are three conjugate pairs of variables here, (T, S) , (P, V) , and (μ, \bar{N}) . Of these S , V , and \bar{N} are extensive, and T , P , and μ are intensive. The natural variables for \bar{E} are S , V , and \bar{N} . We can use this to calculate the temperature, pressure, and chemical potential:

$$T = \left(\frac{\partial \bar{E}}{\partial S} \right)_{P, \bar{N}}, \quad P = - \left(\frac{\partial \bar{E}}{\partial V} \right)_{S, \bar{N}}, \quad \text{and} \quad \mu = \left(\frac{\partial \bar{E}}{\partial \bar{N}} \right)_{S, V}. \quad (4.1.2)$$

The **Helmholtz free energy** is defined as

$$F := \bar{E} - TS. \quad (4.1.3)$$

Therefore

$$dF = d(\bar{E} - TS) \quad (4.1.4)$$

$$= d\bar{E} - T dS - S dT \quad (4.1.5)$$

$$= T dS - P dV + \mu d\bar{N} - T dS - S dT \quad (4.1.6)$$

$$= -S dT - P dV + \mu d\bar{N}. \quad (4.1.7)$$

This means that T , V , and \bar{N} are the natural variables for F . This also gives us new ways to calculate the entropy, pressure, and chemical potential:

$$S = - \left(\frac{\partial F}{\partial T} \right)_{V, \bar{N}}, \quad P = - \left(\frac{\partial F}{\partial V} \right)_{T, \bar{N}}, \quad \mu = - \left(\frac{\partial F}{\partial \bar{N}} \right)_{T, V}. \quad (4.1.8)$$

The **grand potential** is defined as

$$\Phi := F - \mu \bar{N}. \quad (4.1.9)$$

Hence,

$$d\Phi = d(F - \mu\bar{N}) \quad (4.1.10)$$

$$= dF - \mu d\bar{N} - \bar{N} d\mu \quad (4.1.11)$$

$$= -S dT - P dV - \mu d\bar{N} - \bar{N} d\mu \quad (4.1.12)$$

$$= -S dT - P dV + \bar{N} d\mu. \quad (4.1.13)$$

This means that T , V , and μ are the natural variables for Φ . We can also use this to calculate the entropy, pressure, and mean number of particles:

$$T = -\left(\frac{\partial\Phi}{\partial T}\right)_{V,\mu}, \quad P = -\left(\frac{\partial\Phi}{\partial V}\right)_{T,\mu}, \quad \text{and} \quad \bar{N} = \left(\frac{\partial\Phi}{\partial\mu}\right)_{T,V}. \quad (4.1.14)$$

At this point we see a pattern forming. We take a potential and then subtract the product of conjugate variables to get the next potential. Since we have three pairs of conjugate variables for a PVT single species system there are $2^3 = 8$ possible potentials, each with three different natural variables. The important ones when it comes to physics are given below.

Symbol	Definition	Name	Natural Variables
\bar{E}		Energy	S, V, \bar{N}
F	$\bar{E} - TS$	Helmholtz free energy	T, V, \bar{N}
H	$\bar{E} + PV$	Enthalpy	S, P, \bar{N}
G	$F + PV$	Gibbs Free Entropy	T, P, \bar{N}
Φ	$F - \mu\bar{N}$	Grand Potential	T, V, μ

4.1.1 Legendre Transforms



For more examples of Legendre transforms see the notes from the Lagrangian dynamics course, in particular the Hamiltonian is the Legendre transform of the Lagrangian.

The similarity in these definitions is no accident. The thermodynamic potentials are related by Legendre transforms, which we can think of as a way to define a new function with different variables related to the previous variables in a non-trivial way such that the new function contains all the information of the first.

Consider some function, f , which is a function of x_i for $i = 1, \dots, k$. Then

$$df = \sum_{i=1}^k \frac{\partial f}{\partial x_i} dx_i = \sum_{i=1}^k u_i dx_i, \quad \text{where} \quad u_i = \frac{\partial f}{\partial x_i}. \quad (4.1.15)$$

We then define a new function,

$$g := f - \sum_{i=r+1}^k u_i x_i. \quad (4.1.16)$$

Which means

$$dg = d \left(f - \sum_{i=r+1}^k u_i x_i \right) \quad (4.1.17)$$

$$= df - \sum_{i=r+1}^k (u_i dx_i + x_i du_i) \quad (4.1.18)$$

$$= \sum_{i=1}^r u_i dx_i - \sum_{i=r+1}^k (u_i dx_i + x_i du_i) \quad (4.1.19)$$

$$= \sum_{i=1}^r u_i dx_i - \sum_{i=r+1}^k x_i du_i. \quad (4.1.20)$$

The function g is then a natural function of the variables $x_1, \dots, x_r, u_{r+1}, \dots, u_k$. We say that g is the **Legendre transform** of f .

The logic behind Legendre transforms can be explained by a one-dimensional example. Let f be a function of the single variable x . This function can be defined by the value $f(x)$ at all possible points x , which is usually done through an equation like $f(x) = \text{something with } x$. We can also specify f , up to some constant, with the values of the derivative, $u(x) = \partial f / \partial x$, at all points. We can think about this as specifying the gradient of the tangent at each point, x . These tangents are lines with a slope u and a y -intercept, which we'll call g . This means that we can express the value at x as $f(x) = g + ux$. We can invert this and get $g(u) = f(x) - ux$. We see that g contains the same information as f and is its Legendre transform.

4.2 Gibbs–Duhem Relation

Consider the energy of the system. This is an extensive value, therefore it should be proportional to the size of the system. Similarly the natural variables for energy, S , X_Y , and \bar{N}_α , are extensive and so also proportional to the size of the system. Consider then what happens if we scale the size of the system by some factor, b . On the one hand, the energy is extensive and so scales by b also, meaning $\bar{E} \rightarrow b\bar{E}$. On the other hand we can also view this as scaling all of the natural variables for \bar{E} by b , so

$$\bar{E}(S, \{X\}, \{\bar{N}\}) \rightarrow \bar{E}(bS, \{bX\}, \{b\bar{N}\}). \quad (4.2.1)$$

We therefore have

$$b\bar{E}(S, \{X\}, \{\bar{N}\}) = \bar{E}(bS, \{bX\}, \{b\bar{N}\}). \quad (4.2.2)$$

Now consider what happens if we differentiate both sides with respect to b , the left hand side is simple:

$$\frac{\partial}{\partial b} b\bar{E}(S, \{X\}, \{\bar{N}\}) = \bar{E}(S, \{X\}, \{\bar{N}\}). \quad (4.2.3)$$

The right hand side is slightly more complex:

$$\frac{\partial}{\partial b} \bar{E}(bS, \{bX\}, \{b\bar{N}\}) = S \frac{\partial}{\partial S} \bar{E}(bS, \{bX\}, \{b\bar{N}\}) \quad (4.2.4)$$

$$+ \sum_Y X_Y \frac{\partial}{\partial X_Y} \bar{E}(bS, \{bX\}, \{b\bar{N}\}) \quad (4.2.5)$$

$$+ \sum_\alpha \bar{N}_\alpha \frac{\partial}{\partial \bar{N}_\alpha} \bar{E}(bS, \{bX\}, \{b\bar{N}\}) \quad (4.2.6)$$

Now evaluating at $b = 1$ the left hand side gives

$$\left. \frac{\partial}{\partial b} b\bar{E}(S, \{X\}, \{\bar{N}\}) \right|_{b=1} = \bar{E}(S, \{X\}, \{\bar{N}\}). \quad (4.2.7)$$

The right hand side gives

$$\left. \frac{\partial}{\partial b} \bar{E}(bS, \{bX\}, \{b\bar{N}\}) \right|_{b=1} = S \frac{\partial \bar{E}}{\partial S} + \sum_Y X_Y \frac{\partial \bar{E}}{\partial X_Y} + \sum_\alpha \bar{N}_\alpha \frac{\partial \bar{E}}{\partial \bar{N}_\alpha}. \quad (4.2.8)$$

Recognising $T = \frac{\partial \bar{E}}{\partial S}$, $f_Y = \frac{\partial \bar{E}}{\partial X_Y}$, and $\mu_\alpha = \frac{\partial \bar{E}}{\partial \bar{N}_\alpha}$ from [Equations \(3.2.16\) and \(4.1.2\)](#) we have

$$\bar{E} = TS \sum_Y f_Y X_Y + \sum_\alpha \mu_\alpha \bar{N}_\alpha. \quad (4.2.9)$$

We can rewrite the two sums using the Gibbs free energy and the grand potential:

$$G = \bar{E} - TS - \sum_Y - \sum_Y f_Y X_Y = \sum_\alpha \mu_\alpha \bar{N}_\alpha \quad (4.2.10)$$

$$\Phi = \bar{E} - TS - \sum_\alpha \mu_\alpha \bar{N}_\alpha = \sum_Y f_Y X_Y. \quad (4.2.11)$$

Hence

$$\bar{E} = TS + G + \Phi. \quad (4.2.12)$$

The important thing is that we can now compute \bar{E} based solely on the assumption that this is the form of \bar{E} . Doing so we find that

$$d\bar{E} = T dS + S dT + \sum_Y (f_Y dX_Y + X_Y df_Y) + \sum_\alpha (\mu_\alpha d\bar{N}_\alpha + \bar{N}_\alpha d\mu_\alpha). \quad (4.2.13)$$

But from the first and second law we know that

$$d\bar{E} = T dS + \sum_Y f_Y dX_Y + \sum_\alpha \mu_\alpha d\bar{N}_\alpha. \quad (4.2.14)$$

Subtracting this from the previous expression for $d\bar{E}$ we get

$$0 = S dT + \sum_Y X_Y df_Y + \sum_\alpha \bar{N}_\alpha d\mu_\alpha. \quad (4.2.15)$$

This is the **Gibbs-Duhem relation**. It shows that the intensive variables T , f_Y , and μ_α , are not independent.

For example, in a PVT system with a single species we have

$$0 = S dT - V dP + \bar{N} d\mu \quad (4.2.16)$$

which means that an increase in, say, the temperature must be balanced by an increase in pressure and/or decrease in the chemical potential.

4.3 Ensembles

At this point we look slightly more at what we mean by an ensemble. So far we have seen that maximising S subject to constraints gives the probability associated with each microstate of the system. Different ensembles correspond to different sets of constraints. This works well with the Bayesian view of probability. In this section we will use the frequentist view.

We can think of an ensemble as a very large number, M , of the same assembly. Using the frequentist definition of probability the probability of being in microstate i is given by

$$p_i = \lim_{M \rightarrow \infty} \frac{m_i}{M} \quad (4.3.1)$$

where m_i is the number of assemblies in microstate i .

We can think of a volume being divided up into smaller pieces. Although each small piece is small relative to the entire ensemble it still has a large number of particles, say on the order of Avogadro's number.

The entire megasystem is isolated from the rest of the universe so all microstates of the megasystem are equally likely, that is the megasystem is a microcanonical ensemble. If we allow the subsystems to exchange energy then each one is a canonical ensemble. If we allow them to exchange particles also then each one is a grand canonical ensemble.

It can be shown that using the Boltzmann distribution to describe the entropy of the whole system we get the Gibbs entropy as the definition of the entropy for each assembly.

Five

Fluctuations

5.1 Energy Fluctuations

Consider a canonical ensemble. The internal energy fluctuates randomly about the mean value, \bar{E} . In the thermodynamic limit we expect these fluctuations to vanish and we get a sharply defined mean energy, which we take to be *the* energy. For a canonical ensemble we can express the mean energy as

$$\bar{E} = \sum_i p_i E_i \quad (5.1.1)$$

$$= \frac{1}{Z_c} \sum_i E_i e^{-\beta E_i} \quad (5.1.2)$$

$$= \frac{1}{Z_c} \sum_i \frac{\partial}{\partial \beta} e^{-\beta E_i} \quad (5.1.3)$$

$$= \frac{1}{Z_c} \frac{\partial}{\partial \beta} \sum_i e^{-\beta E_i} \quad (5.1.4)$$

$$= \frac{1}{Z_c} \frac{\partial Z_c}{\partial \beta} \quad (5.1.5)$$

$$= -\frac{\partial}{\partial \beta} \ln Z_c. \quad (5.1.6)$$

We will soon see that being able to write thermodynamic variables as “logarithmic derivatives” of the partition function, by which we mean derivatives of the log of the partition function, is a useful way to write things.

We wish to estimate the size of the fluctuation. To do this we consider the heat capacity at constant volume, C_V :

$$C_V := \left(\frac{\partial \bar{E}}{\partial T} \right)_V \quad (5.1.7)$$

$$= \frac{d\beta}{dT} \frac{\partial \bar{E}}{\partial \beta} \quad (5.1.8)$$

$$= -\frac{1}{k_B T^2} \frac{\partial}{\partial \beta} \left(-\frac{1}{Z_c} \frac{\partial Z_c}{\partial \beta} \right) \quad (5.1.9)$$

$$= -\frac{1}{k_B T^2} \left[-\frac{1}{Z_c} \frac{\partial^2 Z_c}{\partial \beta^2} + \frac{1}{Z_c^2} \left(\frac{\partial Z_c}{\partial \beta} \right)^2 \right] \quad (5.1.10)$$

$$= -\frac{1}{k_B T^2} \left[-\frac{1}{Z_c} \frac{\partial^2}{\partial \beta^2} \sum_i e^{-\beta E_i} + \left(\frac{1}{Z_c} \frac{\partial Z_c}{\partial \beta} \right)^2 \right] \quad (5.1.11)$$

$$= -\frac{1}{k_B T^2} \left[-\frac{1}{Z_c} \sum_i E_i^2 e^{-\beta E_i} + (-\bar{E})^2 \right] \quad (5.1.12)$$

$$= -\frac{1}{k_B T^2} \left[-\sum_i E_i^2 p_i + \bar{E}^2 \right] \quad (5.1.13)$$

$$= -\frac{1}{k_B T^2} \left[-\bar{E}^2 + \bar{E}^2 \right] \quad (5.1.14)$$

$$= \frac{1}{k_B T^2} \left[\bar{E}^2 - \bar{E}^2 \right]. \quad (5.1.15)$$

Considering the fluctuation of microstate i from the mean, $\Delta E_i := E_i - \bar{E}$, we get that the mean-square fluctuation is

$$\overline{\Delta E^2} := \sum_i \Delta E_i^2 p_i \quad (5.1.16)$$

$$= \sum_i (E_i - \bar{E})^2 p_i \quad (5.1.17)$$

$$= \sum_i (E_i^2 + \bar{E}^2 - 2E_i \bar{E}) p_i \quad (5.1.18)$$

$$= \sum_i E_i^2 p_i + \bar{E}^2 \sum_i p_i - 2\bar{E} \sum_i E_i p_i \quad (5.1.19)$$

$$= \bar{E}^2 + \bar{E}^2 - 2\bar{E}^2 \quad (5.1.20)$$

$$= \bar{E}^2 - \bar{E}^2. \quad (5.1.21)$$

Comparing this to our result for C_V we see that the root-mean-square fluctuation is

$$\Delta E_{\text{rms}} := \sqrt{\overline{\Delta E^2}} = \sqrt{k_B T^2 C_V}. \quad (5.1.22)$$

This quantity is extensive and scales with the total energy of the system. It is more useful to consider a measure of *relative* fluctuation size. Normalising by the mean energy we have

$$\frac{\Delta E_{\text{rms}}}{\bar{E}} = \frac{\sqrt{k_B T^2 C_V}}{\bar{E}} \sim \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}}. \quad (5.1.23)$$

Hence as the size of the system increases the size of the relative fluctuations vanishes. This ratio, of the standard deviation divided by the mean, is called the **coefficient of variation**. For a system of one mole we have $N \approx 10^{24}$ and so the coefficient of variation is approximately 10^{-12} . This fluctuation size is far lower than can be detected by experiments on this scale and so for all intents and purposes the energy is fixed at \bar{E} . If the total energy isn't fluctuating then we expect the canonical ensemble to behave exactly like the microcanonical ensemble in this limit.

5.2 Magnetic Fluctuations

As a second example consider a system in an external magnetic field, H . The energy of a microstate is then given by the normal energy without the field, minus the de-

crease in energy due to the systems magnetisation aligning with the external field. If the magnetisation of system in microstate i is \mathbf{M}_i then this contribution is $\mu_0 \mathbf{M}_i \cdot \mathbf{H}$ and so

$$E_i(\mathbf{H}) = E_i(\mathbf{H} = \mathbf{0}) - \mu_0 \mathbf{M}_i \cdot \mathbf{H}. \quad (5.2.1)$$

For simplicity we will now only consider the one dimensional case, so $\mathbf{M}_i \cdot \mathbf{H}$ becomes $M_i H$.

Following similar logic to the energy case we have

$$\bar{M} := \sum_i p_i M_i \quad (5.2.2)$$

$$= \frac{1}{Z_c} \sum_i M_i e^{-\beta E_i(H)} \quad (5.2.3)$$

$$= \frac{1}{Z_c} \sum_i M_i \exp[-\beta(E_i(H=0) - \mu_0 M_i H)] \quad (5.2.4)$$

$$= \frac{1}{Z_c} \sum_i \frac{1}{\beta \mu_0} \frac{\partial}{\partial H} \exp[-\beta(E_i(H=0) - \mu_0 M_i H)] \quad (5.2.5)$$

$$= \frac{1}{Z_c} \frac{1}{\beta \mu_0} \frac{\partial}{\partial H} \sum_i \exp[-\beta(E_i(H=0) - \mu_0 M_i H)] \quad (5.2.6)$$

$$= \frac{1}{Z_c} \frac{1}{\beta \mu_0} \frac{\partial Z_c}{\partial H} \quad (5.2.7)$$

$$= \frac{1}{\beta \mu_0} \frac{\partial}{\partial H} \ln Z_c. \quad (5.2.8)$$

It can then be shown that

$$\overline{\Delta M^2} = \frac{k_B T}{\mu_0} \chi \quad (5.2.9)$$

where χ is the isothermal magnetic susceptibility, defined as

$$\chi := \left(\frac{\partial M}{\partial H} \right)_{T,V}. \quad (5.2.10)$$

5.3 Density Fluctuations

Now consider grand canonical ensemble and consider the fluctuation of the number of particles. First notice that we can write \bar{N} as

$$\bar{N} := \sum_{i,N} p_{iN} N \quad (5.3.1)$$

$$= \frac{1}{Z_{gc}} \sum_{i,N} N e^{-\beta(E_{iN} - \mu N)} \quad (5.3.2)$$

$$= \frac{1}{Z_{gc}} \sum_{i,N} \frac{1}{\beta} \frac{\partial}{\partial \mu} e^{-\beta(E_{iN} - \mu N)} \quad (5.3.3)$$

$$= \frac{1}{\beta} \frac{1}{Z_{gc}} \frac{\partial}{\partial \mu} \sum_{i,N} e^{-\beta(E_{iN} - \mu N)} \quad (5.3.4)$$

$$= \frac{1}{\beta} \frac{1}{Z_{gc}} \frac{\partial Z_{gc}}{\partial \mu} = \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln Z_{gc}. \quad (5.3.5)$$

Recall that in the grand canonical ensemble we have $\Phi = -k_B T \ln \mathcal{Z}_{\text{gc}}$. This allows us to write

$$\bar{N} = - \left(\frac{\partial \Phi}{\partial \mu} \right)_{T,V}. \quad (5.3.6)$$

We can also write

$$\left(\frac{\partial \bar{N}}{\partial \mu} \right)_{T,V} = \frac{1}{\beta} \frac{\partial^2}{\partial \mu^2} \ln \mathcal{Z}_{\text{gc}} \quad (5.3.7)$$

$$= \frac{1}{\beta} \left[\frac{1}{\mathcal{Z}_{\text{gc}}} \frac{\partial^2 \mathcal{Z}_{\text{gc}}}{\partial \mu^2} - \frac{1}{\mathcal{Z}_{\text{gc}}^2} \left(\frac{\partial \mathcal{Z}_{\text{gc}}}{\partial \mu} \right)^2 \right] \quad (5.3.8)$$

$$= \beta [\overline{N^2} - \bar{N}^2]. \quad (5.3.9)$$

Hence,

$$\overline{\Delta N^2} = k_B T \left(\frac{\partial \bar{N}}{\partial \mu} \right)_{T,V}. \quad (5.3.10)$$

This is extensive and so it follows that

$$\frac{\Delta N_{\text{rms}}}{\bar{N}} = \frac{\sqrt{\overline{\Delta N^2}}}{\bar{N}} \sim \frac{1}{\sqrt{\bar{N}}} \quad (5.3.11)$$

and so the fluctuations again vanish in the thermodynamic limit. Therefore in this limit we expect the grand canonical ensemble to behave like the canonical ensemble, which we have already shown behaves like a microcanonical ensemble in this limit.

It is possible to rewrite this result in terms of the isothermal compressibility, κ_T , which is somewhat analogous to C_V , and is defined as

$$\kappa_T := -\frac{1}{v} \left(\frac{\partial v}{\partial P} \right)_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_{T,\bar{N}} \quad (5.3.12)$$

where $v = V/\bar{N} = 1/\rho$ is the volume per particle. It can then be shown that

$$\left(\frac{\partial \bar{N}}{\partial \mu} \right)_{T,V} = \rho \kappa_T \bar{N}. \quad (5.3.13)$$

From which it again follows that

$$\frac{\Delta N_{\text{rms}}}{\bar{N}} = \sqrt{\frac{k_B T \rho \kappa_T}{\bar{N}}} \sim \frac{1}{\sqrt{\bar{N}}}. \quad (5.3.14)$$

5.4 General Theory

We can now formulate our observations above into a general procedure for some observable, A , with conjugate field f , which play the roles of the magnetisation and magnetic field respectively. If in microstate i the observable takes the value A_i then the energy can be written as

$$E_i(f) = E_i(f=0) - f A_i. \quad (5.4.1)$$

In the canonical ensemble we then have

$$\beta \bar{A} = \frac{\partial}{\partial f} \ln Z_c, \quad (5.4.2)$$

which follows exactly the same derivation as used for \bar{M} . We can then define the generalised susceptibility, which we call the **response function**:

$$\chi_{AA} = \frac{\partial \bar{A}}{\partial f} \quad (5.4.3)$$

and it can be shown that

$$\overline{\Delta A^2} = \frac{\chi_{AA}}{\beta}. \quad (5.4.4)$$

The double subscript A on χ_{AA} denotes that this is the response of A in a change to the conjugate field to A . This notation suggests that we may also consider the response of A due to a change in the conjugate field to some other observable, B , and then define the response function χ_{AB} , but we won't do so here.

In general χ_{AA} will be extensive and therefore

$$\frac{\Delta A_{\text{rms}}}{\bar{N}} = \frac{\sqrt{\overline{\Delta A^2}}}{\bar{N}} \sim \frac{1}{\sqrt{N}}. \quad (5.4.5)$$

We therefore expect fluctuations in A to vanish in the large \bar{N} limit.

The only time that there may be issues with this is if the coefficient of proportionality in $\chi_{AA} \propto N$ diverges at some point. In this case we get large fluctuations. This typically happens near phase transitions.

It is possible to derive a similar general theory for the grand canonical ensemble.

Six

Weakly Interacting Systems

A system is **weakly interacting** if no energy is stored in any interaction potential. This means the energy of the system is entirely due to the kinetic energy of the particles and their interactions with some external potential. The only interactions between particles are exchanging kinetic energy and momentum. Weakly interacting is essentially the closest we can get to no interactions without just having particles that behave like single particles.

6.1 Localised Particles

Consider a system of N particles fixed on a lattice. For example we could consider a crystal with atoms fixed at the lattice points. Each particle can be distinguished from the others by its location. We say that the particles are **distinguishable**.

Each particle in this case will have its own energy spectrum. As with a single particle we can label the states by an integer, but we now need another subscript to denote the particle. So the m th particle may be in state j_m , in which it has energy ε_{j_m} . The microstate is then a set of labels, $i = \{j_1, j_2, \dots, j_N\}$. For weakly interacting particles (we will show later that) the energy of a given microstate is the sum of the energies of the particles in this microstate:

$$E_i = \varepsilon_{j_1} + \varepsilon_{j_2} + \dots + \varepsilon_{j_N} = \sum_{n=1}^N \varepsilon_{j_n}. \quad (6.1.1)$$

The partition function in the canonical ensemble is then

$$Z_c = \sum_i e^{-\beta E_i} \quad (6.1.2)$$

$$= \sum_{j_1} \sum_{j_2} \dots \sum_{j_N} \exp \left[-\beta \sum_{n=1}^N \varepsilon_{j_n} \right] \quad (6.1.3)$$

$$= \sum_{j_1} \sum_{j_2} \dots \sum_{j_N} \prod_{n=1}^N e^{-\beta \varepsilon_{j_n}} \quad (6.1.4)$$

$$= \left(\sum_{j_1} e^{-\beta \varepsilon_{j_1}} \right) \left(\sum_{j_2} e^{-\beta \varepsilon_{j_2}} \right) \dots \left(\sum_{j_N} e^{-\beta \varepsilon_{j_N}} \right) \quad (6.1.5)$$

$$= [Z(1)]^N \quad (6.1.6)$$

where

$$Z(1) = \sum_j e^{-\beta \epsilon_j} \quad (6.1.7)$$

is the single particle partition function and this sum is over single particle states.

The thermodynamics of a weakly interacting system can then be derived using the bridge equation

$$F = -k_B T \ln Z_c = -N k_B T \ln(Z(1)). \quad (6.1.8)$$

The probability of finding a particular particle in state j is given by

$$p_j = \frac{1}{Z(1)} e^{-\beta \epsilon_j}. \quad (6.1.9)$$

This is the canonical, or Boltzmann, distribution for a single particle.

6.2 Non-Localised Particles

If the particles aren't localised then there is no way to keep track of them and so we can't distinguish them. We must therefore get the same state if we permute the particles since the final result cannot depend on what we call the particles.

The microstate, i , of the assembly is specified by stating how many particles are in each single particle state, j . Therefore i is specified by $\{n_j\}$, where n_j is the number of particles in the single particle state j .

A sum over microstates then becomes a sum over allowed occupation numbers. We can write the energy of a microstate as

$$E_{iN} = \sum_j n_j \epsilon_j. \quad (6.2.1)$$

The total number of particles is simply

$$N = \sum_j n_j. \quad (6.2.2)$$

If we work in the canonical ensemble then we are constrained to microstates with a fixed value of N . This makes the sums difficult since we would have to be cautious about partitioning the particles into states. Instead we work in the grand canonical ensemble and we can relax this limit to just having the average number of particles fixed. Doing so we find that

$$\mathcal{Z}_{gc} = \sum_{n_1} \sum_{n_2} \dots \sum_{n_N} \exp \left[-\beta \left(\sum_j n_j \epsilon_j - \mu \sum_j n_j \right) \right] = \prod_j \mathcal{Z}_j. \quad (6.2.3)$$

We arrive at this in the same way as we did in the localised canonical case. Again

$$\mathcal{Z}_j = \sum_{n_j} e^{-\beta n_j (\epsilon_j - \mu)} \quad (6.2.4)$$

is the single state partition function for state j . Notice that the factorisation is over states here, rather than particles as with the localised canonical case.

The probability of finding the system in the microstate $\{n_j\}$ is

$$p_{\{n_j\}} = \frac{1}{\mathcal{Z}_{\text{gc}}} \exp \left[\beta \left(\mu \sum_j n_j - \beta \sum_j n_j \epsilon_j \right) \right] = \prod_j p_{n_j}. \quad (6.2.5)$$

The last term is a product over states and p_{n_j} is the probability of finding exactly n_j particles in state j . This is given by

$$p_{n_j} = \frac{1}{\mathcal{Z}_j} e^{\beta n_j (\mu - \epsilon_j)}. \quad (6.2.6)$$

The mean number of particles in state j is then

$$\bar{n}_j = \sum_{n_k} n_j p_{\{n_k\}} \quad (6.2.7)$$

$$= \sum_{n_j} n_j p_{n_j} \quad (6.2.8)$$

$$= \sum_{n_j} n_j \frac{1}{\mathcal{Z}_j} e^{\beta n_j (\mu - \epsilon_j)} \quad (6.2.9)$$

$$= \sum_{n_j} \frac{1}{\mathcal{Z}_j} \frac{1}{\beta} \frac{\partial}{\partial \mu} e^{\beta n_j (\mu - \epsilon_j)} \quad (6.2.10)$$

$$= \frac{1}{\mathcal{Z}_j} \frac{1}{\beta} \frac{\partial}{\partial \mu} \sum_{n_j} e^{\beta n_j (\mu - \epsilon_j)} \quad (6.2.11)$$

$$= \frac{1}{\mathcal{Z}_j} \frac{1}{\beta} \frac{\partial}{\partial \mu} \mathcal{Z}_j \quad (6.2.12)$$

$$= \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln \mathcal{Z}_j \quad (6.2.13)$$

$$= k_B T \frac{\partial}{\partial \mu} \ln \mathcal{Z}_j. \quad (6.2.14)$$

6.3 Fermions and Bosons

In order to perform the sum over allowed occupation numbers we need to distinguish between fermions and bosons. Recall that **fermions** have half integers spin, $n/2$ for odd n , and the Pauli exclusion principle applies meaning that no two fermions can be in the same state. On the other hand, **bosons** have integer spin and there is no Pauli exclusion principle.

6.3.1 Fermions

For fermions the Pauli exclusion principle means that the only allowed occupation numbers are 0 and 1 since if $n_j = 2$ then there are two particles in state j , which isn't allowed. Hence

$$\mathcal{Z}_j = \sum_{n_j} e^{\beta n_j (\mu - \epsilon_j)} = e^0 + e^{\beta (\mu - \epsilon_j)} = 1 + e^{\beta (\mu - \epsilon_j)}. \quad (6.3.1)$$

6.3.2 Bosons

${}^1\mathbb{N} = \{0, 1, 2, \dots\}$.

For bosons the occupation number can be any natural number¹. Hence

$$\mathcal{Z}_j = \sum_{n_j=0}^{\infty} e^{\beta n_j (\mu - \varepsilon_j)} \quad (6.3.2)$$

$$= \sum_{n_j=0}^{\infty} (e^{\beta(\mu - \varepsilon_j)})^{n_j} \quad (6.3.3)$$

$$= \frac{1}{1 - e^{\beta(\mu - \varepsilon_j)}}. \quad (6.3.4)$$

Here we have identified a geometric series, $\sum_{n=0}^{\infty} x^n = 1/(1-x)$, which converges uniformly for $|x| < 1$, which is indeed the case here since μ is negative and $-\varepsilon_j$ is negative and β is positive so we have x being e raised to some negative power, which will always give something in the range $(0, 1)$.

6.3.3 Both

It is possible to write both of these in one by taking $+$ for the case of fermions and $-$ for bosons in the following:

$$\mathcal{Z}_j = [1 \pm e^{\beta(\mu - \varepsilon_j)}]^{\pm 1}. \quad (6.3.5)$$

We can then compute the mean number of particles in state n_j :

$$\bar{n}_j = \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln \mathcal{Z}_j \quad (6.3.6)$$

$$= \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln [1 \pm e^{\beta(\mu - \varepsilon_j)}]^{\pm 1} \quad (6.3.7)$$

$$= \pm \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln [1 \pm e^{\beta(\mu - \varepsilon_j)}] \quad (6.3.8)$$

$$= \frac{e^{\beta(\mu - \varepsilon_j)}}{1 \pm e^{\beta(\mu - \varepsilon_j)}} \quad (6.3.9)$$

$$= \frac{1}{e^{\beta(\varepsilon_j - \mu)} \pm 1} \quad (6.3.10)$$

$$= f_{\pm}(\varepsilon_j, \mu). \quad (6.3.11)$$

Taking $+$ for fermions and $-$ for Bosons we get the **Fermi–Dirac distribution**,

$$\bar{n} = f_+(\varepsilon, \mu) = \frac{1}{e^{\beta(\varepsilon - \mu)} + 1}, \quad (6.3.12)$$

and **Bose–Einstein distribution**,

$$\bar{n} = f_-(\varepsilon, \mu) = \frac{1}{e^{\beta(\varepsilon - \mu)} - 1}, \quad (6.3.13)$$

respectively. To remember the sign we can use the fact that $+$ represents addition and this goes with the Fermi–Dirac distribution and $-$ represents subtraction and this goes with the Bose–Einstein distribution.

Seven

Dilute Limit

7.1 Dilute Limit

At high temperatures and low particle densities, this is known as the **dilute limit**. In this limit the de Broglie wavelength is much less than the mean interparticle separation, and so particles are sufficiently separated that they behave as distinct particles and are essentially distinguishable. The chance of any one state being occupied in this limit is low and so we don't need to worry about the fermion/boson distinction since the probability of two particles trying to occupy the same state is negligible. We therefore expect the Fermi–Dirac and Bose–Einstein distributions to be the same in this limit.

The way we achieve this limit mathematically is by taking $e^{\beta\mu} \ll 1$. We have to be careful since μ is a typically increasing function of T and $\beta \propto 1/T$.

Without loss of generality we can take the lowest energy state to define zero energy, $\varepsilon_0 = 0$, so for all other states, i , we have $\varepsilon_i \geq 0$. In this case we have

$$e^{\beta(\mu-\varepsilon_i)} \ll 1 \implies e^{\beta(\varepsilon_i-\mu)} \gg 1. \quad (7.1.1)$$

This means that adding or subtracting 1 from this quantity is negligible so we have

$$\bar{n}_i = f_{\pm}(\varepsilon_i, \mu) = \frac{1}{e^{\beta(\varepsilon_i-\mu)} \pm 1} \approx \frac{1}{e^{\beta(\varepsilon_i-\mu)}} = e^{\beta(\mu-\varepsilon_i)}. \quad (7.1.2)$$

So we get something similar to the Boltzmann distribution, $e^{-\beta\varepsilon_i}$, and this result is the same for both fermions and bosons.

If we had classically distinguishable particles then we would have

$$\bar{n}_i = \bar{N} p_i = \bar{N} \frac{e^{-\beta\varepsilon_i}}{Z(1)}. \quad (7.1.3)$$

Therefore in the dilute limit we have

$$\bar{n}_i \approx e^{\beta(\mu-\varepsilon_i)} \approx \frac{\bar{N}}{Z(1)} e^{-\beta\varepsilon_i} \implies e^{\beta\mu} \approx \frac{\bar{N}}{Z(1)}. \quad (7.1.4)$$

7.2 Semi-Classical Approximation

Consider $\ln \mathcal{Z}_j$ in the dilute limit, using [Equation \(6.3.5\)](#) we have

$$\ln \mathcal{Z}_j = \pm \ln[1 \pm e^{\beta(\mu-\varepsilon_j)}] \approx e^{\beta(\mu-\varepsilon_j)}. \quad (7.2.1)$$

Where we've used the Taylor expansion $\ln(1 \pm x) \approx \pm x$ for small x .

Now consider the grand potential, Φ :

$$\Phi = -k_B T \ln \mathcal{Z}_{gc} \quad (7.2.2)$$

$$= -k_B T \sum_j \ln \mathcal{Z}_j \quad (7.2.3)$$

$$\approx -k_B T \sum_j e^{\beta(\mu - \epsilon_j)} \quad (7.2.4)$$

$$= -k_B T e^{\beta\mu} \sum_j e^{-\beta\epsilon_j} \quad (7.2.5)$$

$$\approx -k_B T e^{\beta\mu} Z(1) \quad (7.2.6)$$

$$\approx -\bar{N} k_B T. \quad (7.2.7)$$

Here we have used the approximation $e^{\beta\mu} Z(1) = \bar{N}$ in the dilute limit.

Consider instead the Helmholtz free energy, using $e^{\beta\mu} Z(1) = \bar{N}$ again, now rearranged to give $\mu = \ln(\bar{N}/Z(1))/\beta$, we get

$$F = \Phi + \mu \bar{N} \quad (7.2.8)$$

$$\approx -k_B T \bar{N} + k_B T \bar{N} \ln \left[\frac{\bar{N}}{Z(1)} \right] \quad (7.2.9)$$

$$= k_B T (\bar{N} \ln \bar{N} - \bar{N} - \bar{N} \ln [Z(1)]) \quad (7.2.10)$$

$$\approx k_B T (\ln(\bar{N}!) - \ln(Z(1)^{\bar{N}})) \quad (7.2.11)$$

$$= -k_B T \bar{N} \ln \left[\frac{Z(1)^{\bar{N}}}{\bar{N}!} \right]. \quad (7.2.12)$$

Here we have used **Stirling's approximation**

$$\ln(N!) \sim N \ln N - N \quad (7.2.13)$$

for large N . Using

$$F = -k_B T \ln Z_c \quad (7.2.14)$$

we can identify that in the dilute limit

$$Z_c = \frac{Z(1)^{\bar{N}}}{\bar{N}!}. \quad (7.2.15)$$

This is the **semi-classical approximation**.



Recall that in the statistical mechanics part of the thermal physics course we arrived at this result by a not particularly convincing argument of fixing the overcounting of states which occurs due to failing to account for the indistinguishability of particles.

Eight

Density of States

To calculate thermodynamic properties we need to perform sums over states. This is not that easy to do. Instead we can often approximate the sums as integrals, which can then be computed more easily, or numerically, or just looked up. The goal of this section will be to find the **density of states** in energy-space, g , which is a function such that $g(\varepsilon) d\varepsilon$ is the number of states with energies in the interval $[\varepsilon, \varepsilon + d\varepsilon]$. This function is such that

$$\bar{N} = \sum_j \bar{n}_j = \sum_j f(\varepsilon_j, \mu) \approx \int_0^\infty f(\varepsilon, \mu) g(\varepsilon) d\varepsilon. \quad (8.0.1)$$

This approximation is generally valid when there are many states with similar energies, so that the discrete steps between states in the sum are small, and the sum can be approximated by the continuous integral. Generally we get states with close energy levels at high temperatures. So we expect the density of states to be a good approximation when the temperature of the system is high enough that the higher energy states dominate.

8.1 Density of States for a Particle in a Box

Consider a particle in a box, which we take to be a cube for simplicity. Mathematically the states are given by the solutions to the time independent Schrödinger equation for a free particle,

$$-\frac{\hbar^2}{2m} \nabla^2 \psi = \varepsilon \psi, \quad (8.1.1)$$

over the region $[0, L]^3$, with the boundary condition that $\psi = 0$ when any of x , y , or z are either 0 or L . This is simply an eigenvalue problem.

The eigenfunctions of the Laplacian are sin and cos, for our particular boundary conditions we take

$$\psi = A \sin(k_x x) \sin(k_y y) \sin(k_z z). \quad (8.1.2)$$

If we take $k_i = n_i \pi / L$ for $n_i = 1, 2, \dots$ then we satisfy the boundary conditions. We can easily show that

$$\nabla^2 \psi = -(k_x^2 + k_y^2 + k_z^2) \psi = -k^2 \psi \quad (8.1.3)$$

$$^1\mathbb{Z}_{>0} = \{1, 2, \dots\} = \mathbb{N} \setminus \{0\}$$

where $k = |\mathbf{k}|$ with $\mathbf{k} = (k_x, k_y, k_z)$ being a vector in k -space. We can also write¹ $\mathbf{k} = \pi \mathbf{n}/L$ where $\mathbf{n} \in \mathbb{Z}_{>0}^3$ is a vector in n -space. We can then identify

$$\varepsilon = \frac{\hbar^2}{2m} k^2. \quad (8.1.4)$$

8.1.1 Two Dimensions

The states are spaced π/L apart in k -space. In two dimensions this means that the area per state is $(\pi/L)^2$. We are only interested in states with $k_x, k_y \geq 0$. This corresponds to only considering the top right quadrant, which in turn we can do by considering all four quadrants and including factors of $1/4$ as necessary.

The area of an annulus of inner radius k and thickness dk in the top right quadrant is

$$A = \frac{1}{4} [\pi(k + dk)^2 - \pi k^2] = \frac{1}{4} 2\pi k dk + O(dk^2) \approx \frac{\pi}{2} k dk. \quad (8.1.5)$$

This means that this quarter annulus contains

$$\frac{\pi k dk/2}{(\pi/L)^2} = \frac{L^2}{2\pi} k dk \quad (8.1.6)$$

states.

8.1.2 Three Dimensions

In three dimensions the volume per state is $(\pi/L)^3$. Again we are only interested in states where $k_x, k_y, k_z \geq 0$. This now means we only care about one out of eight octants, so we include a factor of $1/8$. The volume of a spherical shell of inner radius k and thickness dk in this octant is

$$\frac{1}{8} \left[\frac{4}{3} \pi (k + dk)^3 - \frac{4}{3} \pi k^3 \right] = \frac{1}{8} 4\pi k^2 dk + O(dk^2) \approx \frac{\pi}{2} k^2 dk. \quad (8.1.7)$$

The number of states in this eighth of a spherical shell is then

$$\frac{\pi k^2 dk/2}{(\pi/L)^3} = \frac{L^3}{2\pi^2} k^2 dk. \quad (8.1.8)$$

Now define Γ to be the density of states in k -space, so that $\Gamma(k)dk$ is the number of states with wave vectors in the interval $[k, k+dk]$. We can relate this to the density of states in energy-space by a simple change of variables. First we notice that since $\varepsilon = \hbar^2 k^2 / (2m)$

$$\frac{d\varepsilon}{dk} = \frac{\hbar^2}{m} k = \sqrt{\frac{2\hbar^2}{m}} \sqrt{\varepsilon}. \quad (8.1.9)$$

Hence

$$g(\varepsilon) d\varepsilon = \Gamma(k) dk = \frac{L^3}{2\pi^2} k^2 dk = \frac{V}{2\pi^2} k^2 dk \quad (8.1.10)$$

where $V = L^3$ is the volume of the cube. Putting in

$$dk = \sqrt{\frac{2\hbar^2}{m}} \sqrt{\varepsilon} d\varepsilon \quad (8.1.11)$$

and $k^2 = 2m\varepsilon/\hbar^2$ we get

$$g(\varepsilon) = \left(\frac{2m}{\hbar}\right)^{3/2} \frac{V}{4\pi^2} \sqrt{\varepsilon}. \quad (8.1.12)$$

The important thing here is that $g(\varepsilon)$ scales linearly with the volume and as $\sqrt{\varepsilon}$.

Nine

Statistical Mechanics and Quantum Mechanics

9.1 Many Particle Schrödinger Equation

In statistical mechanics we are interested in many particle systems. To fully treat these systems we need to solve the time independent many particle Schrödinger equation:

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (9.1.1)$$

Here \mathbf{r}_i is the position of the i th particle, Ψ is the wave function for the state, E is the energy of the state and \hat{H} is the Hamiltonian.

The most general Hamiltonian is

$$\hat{H} = - \sum_{k=1}^N \frac{\hbar^2}{2m_k} \nabla_k^2 + U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (9.1.2)$$

Here m_k is the mass of the k th particle, from now on we will assume all particles have the same mass. ∇_k^2 is the Laplacian operator acting only on \mathbf{r}_k , that is if $\mathbf{r}_k = (x_k, y_k, z_k)$ then

$$\nabla_k^2 := \frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2}. \quad (9.1.3)$$

U is the interaction potential, which in general depends on the positions of all particles.

Often we can write U as a sum of interactions between pairs of particles, such as the case where the only interactions are electrostatic, we then have

$$U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{l < k} \sum_{k=1}^N \varphi(|\mathbf{r}_k - \mathbf{r}_l|) \quad (9.1.4)$$

where $\varphi(r) = q^2/(r\pi\epsilon_0 r^2)$ where q is the charge carried by each particle.

For the case of weakly interacting particles the potential depends only on the external potential, V , felt by each particle and so

$$U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{k=1}^N V(\mathbf{r}_k). \quad (9.1.5)$$

In this case we can write the Hamiltonian as

$$\hat{H}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{k=1}^N \hat{h}_k(\mathbf{r}_k). \quad (9.1.6)$$

Here

$$\hat{h}_k(\mathbf{r}_k) := -\frac{\hbar^2}{2m} \nabla_k^2 + V(\mathbf{r}_k). \quad (9.1.7)$$

The single particle Hamiltonians \hat{h}_k have wave functions ψ_i as solutions and corresponding energies ε_i . These solutions are common between all particles. The many body wave function is then factorised in terms of these single-particles solutions so

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \prod_{k=1}^N \psi_{\alpha_k}(\mathbf{r}_k) \quad (9.1.8)$$

and

$$E = \sum_{k=1}^N \varepsilon_{\alpha_k} \quad (9.1.9)$$

where α_k are integers labelling the states of each individual particle.

It should be noted that this form is for localised or otherwise distinguishable particles. For indistinguishable particles this form over counts and instead we take suitable symmetrised combinations of the single-particle eigenstates to get the multi-particle state. What the “suitable” symmetrised combination is depends on whether we are dealing with fermions or bosons.

For bosons the wave function should be symmetric under exchanging two particles. In this case the most general wave function is a sum over all permutations:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \propto \sum_{\sigma \in S_N} \psi_{\alpha_{\sigma(1)}}(\mathbf{r}_1) \psi_{\alpha_{\sigma(2)}}(\mathbf{r}_2) \cdots \psi_{\alpha_{\sigma(N)}}(\mathbf{r}_N). \quad (9.1.10)$$

Here S_n is the permutation group on n objects and σ is a permutation. For the case of $N = 2$ this means that

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{\sqrt{2}}{2} [\psi_{\alpha_1}(\mathbf{r}_1) \psi_{\alpha_2}(\mathbf{r}_2) + \psi_{\alpha_2}(\mathbf{r}_1) \psi_{\alpha_1}(\mathbf{r}_2)]. \quad (9.1.11)$$

For fermions the wave function should be antisymmetric under exchanging two particles. Therefore the most general wave function comes from summing over all permutations and the totally antisymmetric Levi-Civita symbol, $\varepsilon(\sigma) = \varepsilon_{\sigma(1)\sigma(2)\dots\sigma(N)}$, which is defined to be 1 if σ is an even permutation and -1 if σ is an odd permutation:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \propto \sum_{\sigma \in S_N} \varepsilon(\sigma) \psi_{\alpha_{\sigma(1)}}(\mathbf{r}_1) \psi_{\alpha_{\sigma(2)}}(\mathbf{r}_2) \cdots \psi_{\alpha_{\sigma(N)}}(\mathbf{r}_N). \quad (9.1.12)$$

For the case of $N = 2$ this means that

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{\sqrt{2}}{2} [\psi_{\alpha_1}(\mathbf{r}_1) \psi_{\alpha_2}(\mathbf{r}_2) - \psi_{\alpha_2}(\mathbf{r}_1) \psi_{\alpha_1}(\mathbf{r}_2)]. \quad (9.1.13)$$

9.2 A Tale of Two Probabilities

¹or introduce statistical mechanics to quantum mechanics. See the principles of quantum mechanics notes for that perspective

Things get a bit confusing when we start to introduce quantum mechanics to statistical mechanics¹ due to the fact that we have two notions of probability. In statistical mechanics we have been considering classical probabilities, such as the probability of finding a system in a given state, $p_i = e^{-\beta E_i} / Z_c$ for a canonical ensemble.

On the other hand in quantum mechanics we usually assume that the state is known and that the inherent probability aspect to quantum mechanics occurs when making measurements which give specific results with specific probabilities given by projecting the state onto the relevant eigenspace of the measurement operator.

We now have to deal with both of these and we will do so by coming up with a mathematical notion unifying them. But first, a recap of quantum mechanics.

9.3 Quantum Mechanics Recap



For more details on quantum mechanics see the notes from principles of quantum mechanics.

In quantum mechanics we typically use bracket notation. The state of a particle is a vector, $|\psi\rangle$, in some Hilbert space. The energy eigenstates are the eigenvectors of the Hamiltonian operator, \hat{H} , and the energy is the associated eigenvalue. That is $\hat{H}|i\rangle = E_i|i\rangle$, where \hat{H} is the Hamiltonian operator, $\{|i\rangle\}$ are the energy eigenvalues and E_i is the energy of a particle in state $|i\rangle$.

Since \hat{H} is Hermitian $\{|i\rangle\}$ are orthogonal, and we take them to be normalised so the energy eigenstates are orthonormal and so

$$\langle i|j\rangle = \delta_{ij}. \quad (9.3.1)$$

The identity can then be written as

$$I = \sum_i |i\rangle\langle i|. \quad (9.3.2)$$

We can write a general wave function using projection operators as

$$|\psi\rangle = I|\psi\rangle = \sum_i |i\rangle\langle i|\psi\rangle = \sum_i c_i |i\rangle \quad (9.3.3)$$

where $c_i = \langle i|\psi\rangle \in \mathbb{C}$.

Given an observable, A , there is an associated operator, \hat{A} . The expected value of a measurement of A is

$$\langle A \rangle = \langle \psi | \hat{A} | \psi \rangle = \left(\sum_i c_i^* \langle i | \right) \hat{A} \left(\sum_j | j \rangle \right) = \sum_i \sum_j c_i^* c_j \langle i | \hat{A} | j \rangle = \sum_i \sum_j c_i^* c_j A_{ij}$$

where $A_{ij} := \langle i | \hat{A} | j \rangle$ are the matrix elements of \hat{A} in the energy eigenbasis.

For example, in the energy eigenbasis $\langle i | \hat{H} | j \rangle = E_i \delta_{ij}$ and so

$$\langle E \rangle = \sum_i \sum_j c_i^* c_j E_i \delta_{ij} = \sum_i |c_i|^2 E_i. \quad (9.3.4)$$

The problem that we face when it comes to including statistical mechanics is that in statistical mechanics the mean value of an observable A is

$$\bar{A} = \sum_i p_i A_i. \quad (9.3.5)$$

Identifying $A_{ii} = A_i$ we therefore want the mean in statistical mechanics to be

$$\bar{A} = \sum_i p_i A_{ii} = \sum_i p_i \langle i | \hat{A} | i \rangle. \quad (9.3.6)$$

Unfortunately this can't be written as a simple superposition. Suppose we could write it as

$$|\psi\rangle = \sum_i \sqrt{p_i} |i\rangle, \quad (9.3.7)$$

then

$$\langle A \rangle = \sum_i \sum_j \sqrt{p_i p_j} A_{ij}. \quad (9.3.8)$$

Identifying $A_i = A_{ii}$ we see that this is almost what we wanted but we have off diagonal terms which don't necessarily vanish.

9.4 Density Matrix

The resolution to this problem is to introduce the **density matrix**, ρ . First we define the density matrix as the diagonal matrix with the classical probabilities as its diagonal, that is $\rho_{ij} := p_i \delta_{ij}$. Then

$$\text{tr}(\rho \hat{A}) = \sum_i (\rho \hat{A})_{ii} = \sum_i \sum_j \rho_{ij} A_{ji} = \sum_i \sum_j p_i \delta_{ij} A_{ji} = \sum_i p_i A_{ii} = \bar{A}. \quad (9.4.1)$$

So we see that we recover the statistical mechanics mean in this case by taking $\langle A \rangle := \text{tr}(\rho \hat{A})$.

We can do this same calculation in bracket notation as

$$\rho := \sum_i p_i |i\rangle \langle i|. \quad (9.4.2)$$

Then

$$\begin{aligned} \text{tr}(\rho \hat{A}) &= \sum_j \langle j | \rho \hat{A} | j \rangle = \sum_j \sum_i \langle j | p_i | i \rangle \langle i | \hat{A} | j \rangle \\ &= \sum_i \sum_j p_i \delta_{ij} \langle j | \hat{A} | j \rangle = \sum_i p_i \langle i | \hat{A} | i \rangle. \end{aligned} \quad (9.4.3)$$

Which was what we wanted the statistical mechanics mean to be in bracket notation.

Recall that we can define the exponential of an operator through a power series:

$$e^{\hat{A}} := \sum_{n=0}^{\infty} \frac{1}{n!} \hat{A}^n = \hat{1} + \hat{A} + \frac{1}{2} \hat{A}^2 + \frac{1}{3!} \hat{A}^3 + \dots \quad (9.4.4)$$

We therefore have

$$e^{-\beta \hat{H}} |i\rangle = \sum_{n=0}^{\infty} \frac{1}{n!} (-\beta)^n \hat{H}^n |i\rangle = \sum_{n=0}^{\infty} \frac{1}{n!} (-\beta)^n E_i^n |i\rangle = e^{-\beta E_i} |i\rangle. \quad (9.4.5)$$

Hence we have

$$\begin{aligned}\rho &:= \sum_i p_i |i\rangle\langle i| = \frac{1}{Z_c} \sum_i e^{-\beta E_i} |i\rangle\langle i| = \frac{1}{Z_c} \sum_i e^{-\beta \hat{H}} |i\rangle\langle i| \\ &= \frac{1}{Z_c} e^{-\beta \hat{H}} \sum_i |i\rangle\langle i| = \frac{1}{Z_c} e^{-\beta \hat{H}}.\end{aligned}\quad (9.4.6)$$

Notice the resemblance between $p_i = e^{-\beta E_i}/Z_c$ and $\rho = e^{-\beta \hat{H}}/Z_c$.

We can write the canonical partition function as the trace of $e^{-\beta \hat{H}}$ since

$$\text{tr}(e^{-\beta \hat{H}}) = \sum_{i=0}^{\infty} \langle i | e^{-\beta \hat{H}} | i \rangle = \sum_{i=0}^{\infty} e^{-\beta E_i} \langle i | i \rangle = \sum_{i=0}^{\infty} e^{-\beta E_i} = Z_c. \quad (9.4.7)$$

Now consider what happens if instead $\rho := |\psi\rangle\langle\psi|$. Then

$$\begin{aligned}\langle A \rangle &= \text{tr}(\rho \hat{A}) = \sum_j \langle j | \psi \rangle \langle \psi | \hat{A} | j \rangle = \sum_j c_j \langle \psi | \hat{A} | j \rangle \\ &= \sum_j \sum_i c_j c_i^* \langle j | \hat{A} | i \rangle = \sum_j \sum_i c_i^* c_j A_{ij}.\end{aligned}\quad (9.4.8)$$

So we recover the expectation value in quantum mechanics. Now we just need to find a way to make these two definitions work together so we can do quantum mechanics and statistical mechanics at the same time.

What we have been discussing so far is the density matrix for a canonical ensemble. The most general form of density matrix arises when we have double probability. First we have the inherent quantum probability, but suppose we also don't know what state the system is in, just that it is in a state $|\psi_i\rangle$ with probability p_i . We then can define the density operator

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (9.4.9)$$

This is called a **mixed state**, as opposed to when we know for sure the state of the system is $|\psi\rangle$, which we call a **pure state**. In this case the expectation value for an operator is still given as

$$\langle A \rangle = \overline{A} = \text{tr}(\rho \hat{A}) \quad (9.4.10)$$

and this correctly accounts for both the statistical mechanics and quantum mechanics.

9.4.1 Von Neumann Entropy

We can further generalise the entropy of the system to the **von Neumann entropy**, which is defined to be

$$S := -k \text{tr}(\rho \ln \rho). \quad (9.4.11)$$

The logarithm is defined through its power series:

$$\begin{aligned}\ln \rho &= \ln[1 + (\rho - 1)] = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (\rho - 1)^n \\ &= (\rho - 1) - \frac{1}{2}(\rho - 1)^2 + \frac{1}{3}(\rho - 1)^3 + \dots\end{aligned}\quad (9.4.12)$$

Taking $k = k_B$ in this gives a definition that coincides with the Gibbs entropy for a canonical ensemble.

Ten

Vibrations in a Solid

In this chapter we will look at vibrations in a solid as an example of a many body problem. In a general many body problem we want to solve the time independent Schrödinger equation

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (10.0.1)$$

where

$$\hat{H} = - \sum_{k=1}^N \frac{\hbar^2}{2m} \nabla_k^2 + U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (10.0.2)$$

The problem is if U is not a sum of single particle potentials we can't, except in very special cases, solve the many body Schrödinger equation.

If U is a sum of single particle potentials, V , then the Hamiltonian can be written as a sum of single particle Hamiltonians:

$$\hat{H}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{k=1}^N \hat{h}_k(\mathbf{r}_k), \quad \text{where} \quad \hat{h}_k(\mathbf{r}_k) = -\frac{\hbar^2}{2m} \nabla_k^2 + V(\mathbf{r}_k). \quad (10.0.3)$$

This then has a factorised solution

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \prod_{k=1}^N \psi_{\alpha_k}(\mathbf{r}_k) \quad (10.0.4)$$

which has energy

$$E = \sum_{k=1}^N \varepsilon_{\alpha_k}. \quad (10.0.5)$$

Here ψ_{α_k} is the wave function of the k th particle in state α_k and ε_{α_k} is the energy of this particle. We call this the diagonal form of the Hamiltonian since the particles aren't coupled by off diagonal terms. A lot of the work of many body physics involves making appropriate transformations and approximations to get the Hamiltonian into a diagonal form.

10.1 Einstein Model

Consider a crystalline solid of N atoms arranged in a regular lattice. Since the atoms are localised they are distinguishable. Let \mathbf{r}_k be the displacement of the k th atom from its equilibrium position.

The basis of the Einstein model is to replace the complicated potential that each atom experiences with a single particle potential. The obvious choice being the harmonic potential, either because we can imagine expanding the potential to second order around a minimum and choosing for the constant term to be zero or just because physicists love harmonic oscillators. If we do this then homogeneity and isotropy of the crystal means that all atoms should have the same spring constant and hence the same angular momentum. The form of the Hamiltonian is then

$$\hat{H} = \sum_{k=1}^N \left[-\frac{\hbar^2}{2m} \nabla_k^2 + \frac{1}{2} m \omega^2 |\mathbf{r}_k|^2 \right]. \quad (10.1.1)$$

The angular frequency, ω , is to be determined from experimental measurements.

The Hamiltonian is then a sum of N three-dimensional quantum harmonic oscillators. This further simplifies since it is easy to show via separation of variables that a three-dimensional quantum harmonic oscillator behaves as three independent one-dimensional quantum harmonic oscillators. Therefore the system can be considered as $3N$ one-dimensional quantum harmonic oscillators, which we label with k . The energy of a quantum harmonic oscillator is $\hbar\omega(n + 1/2)$ where $n \in \mathbb{N}$ ¹. The total energy is then

$$^1\mathbb{N} = \{0, 1, 2, \dots\}$$

$$E = \sum_{k=1}^{3N} \varepsilon_k \quad \text{where} \quad \varepsilon_k = \hbar\omega \left(n_k + \frac{1}{2} \right) \quad (10.1.2)$$

for non-negative integer n_k .

As the number of atoms is fixed we are working in the canonical ensemble and since the particles are weakly interacting the partition function is

$$Z_c = [Z(1)]^{3N} \quad (10.1.3)$$

with the single particle partition function given by

$$Z(1) = \sum_{n=0}^{\infty} \exp \left[-\beta \hbar \omega \left(n + \frac{1}{2} \right) \right] \quad (10.1.4)$$

$$= e^{-\beta \hbar \omega / 2} \sum_{n=0}^{\infty} e^{-\beta \hbar \omega n} \quad (10.1.5)$$

$$= \frac{e^{-\beta \hbar \omega / 2}}{1 - e^{-\beta \hbar \omega}}. \quad (10.1.6)$$

Here we have used the fact that this is a geometric series, and $|e^{-\beta \hbar \omega}| < 1$ so it converges. Notice that up to a factor of $e^{-\beta \hbar \omega / 2}$ this is the same as the single particle partition function for bosons and so we can think of the quanta of energy in the harmonic oscillator as a boson.

We can calculate the average energy:

$$\bar{E} = -\frac{\partial}{\partial \beta} \ln Z_c \quad (10.1.7)$$

$$= -\frac{\partial}{\partial \beta} \ln \left[\left(\frac{e^{-\beta \hbar \omega / 2}}{1 - e^{-\beta \hbar \omega}} \right)^{3N} \right] \quad (10.1.8)$$

$$= -3N \frac{\partial}{\partial \beta} \left[-\frac{1}{2} \beta \hbar \omega - \ln(1 - e^{-\beta \hbar \omega}) \right] \quad (10.1.9)$$

$$= \frac{3}{2} N \hbar \omega + 3N \hbar \omega \frac{e^{-\beta \hbar \omega}}{1 - e^{-\beta \hbar \omega}} \quad (10.1.10)$$

$$= \frac{3}{2} N \hbar \omega + \frac{3N \hbar \omega}{e^{\beta \hbar \omega} - 1} \quad (10.1.11)$$

The heat capacity is then given by

$$C_V := \left(\frac{\partial \bar{E}}{\partial T} \right)_V \quad (10.1.12)$$

$$= \frac{\partial \beta}{\partial T} \frac{\partial \bar{E}}{\partial \beta} \quad (10.1.13)$$

$$= -\frac{1}{k_B T^2} \frac{\partial}{\partial \beta} \frac{3N \hbar \omega}{e^{\beta \hbar \omega} - 1} \quad (10.1.14)$$

$$= \frac{1}{k_B T^2} 3N (\hbar \omega)^2 \frac{e^{\beta \hbar \omega}}{(1 - e^{-\beta \hbar \omega})^2} \quad (10.1.15)$$

$$= 3N k_B (\hbar \omega \beta)^2 \frac{e^{\beta \hbar \omega}}{(e^{\beta \hbar \omega} - 1)^2}. \quad (10.1.16)$$

At high temperatures β is small and so we can expand about $\beta = 0$, doing so we have

$$e^{\beta \hbar \omega} \approx 1, \quad (10.1.17)$$

and

$$(e^{\beta \hbar \omega} - 1)^2 \approx (1 + \beta \hbar \omega - 1)^2 = (\beta \hbar \omega)^2, \quad (10.1.18)$$

hence,

$$C_V \approx 3N k_B. \quad (10.1.19)$$

This is the same result that we get by applying the equipartition theorem to the $3N$ harmonic oscillators, each of which has two quadratic degrees of freedom, the kinetic energy and potential energy, each quadratic degree of freedom contributing $k_B T / 2$ to the specific heat capacity.

At low temperatures β is large and so $e^{\beta \hbar \omega} - 1 \approx e^{\beta \hbar \omega}$ meaning

$$C_V \approx 3N k_B (\hbar \omega \beta)^2 e^{-\beta \hbar \omega}. \quad (10.1.20)$$

This tends to zero as $T \rightarrow 0$, as it should. However, the Einstein model predicts that the heat capacity goes to zero exponentially, when careful measurements show that the heat capacity follows a cubic law and so a better model is needed. The problem is that the Einstein model doesn't account for interactions between atoms, the next model does.

10.2 Debye Model

Keeping with the notation of \mathbf{r}_k as the displacement of the k th atom from its equilibrium we introduce x_k^i as the i th Cartesian coordinate of this displacement. The general potential, U , could be a very complicated function of the displacements of the atoms, however, as long as these displacements are small we can always Taylor expand. This is slightly complicated because U is a function of $3N$ variables, x_k^i , but not that much more complicated than the one dimensional case, we just need sums:

$$U = U_0 + \sum_k \sum_i \left. \frac{\partial U}{\partial x_k^i} \right|_0 x_k^i + \frac{1}{2} \sum_k \sum_i \sum_l \sum_j \left. \frac{\partial^2 U}{\partial x_k^i \partial x_l^j} \right|_0 x_k^i x_l^j + \dots \quad (10.2.1)$$

We are free to choose $U_0 = 0$ and the first derivative vanishes since we are expanding about the equilibrium, which is a minimum. Therefore the lowest order term is the quadratic term. We don't know the second derivative, since we don't know U , but we can just hide all of that away by defining

$$A_{kl}^{ij} := \frac{\partial^2 U}{\partial x_k^i \partial x_l^j} \quad (10.2.2)$$

and so the Hamiltonian takes the form

$$\hat{H} = -\frac{\hbar^2}{2m} \sum_{k=1}^N \nabla_k^2 + \frac{1}{2} \sum_{k,l,i,j} A_{kl}^{ij} x_k^i x_l^j. \quad (10.2.3)$$

This leaves us with a pairwise potential which is quadratic in the displacements.

By choosing an appropriate set of coordinates it is possible to diagonalise the Hamiltonian. We know this is possible since this is the Hamiltonian of $3N$ coupled harmonic oscillators. Recall that coupled oscillators have normal modes in which all displacements oscillate with the same frequency. The actual motion of the system is then a superposition of these normal modes and the energy is the sum of the energies of each normal mode. We define some transformation matrix, L_r , such that we can define **normal coordinates**, q_r which diagonalise the Hamiltonian:

$$q_r := \sum_{k,i} L_{r,ki} x_k^i \quad (10.2.4)$$

for $r = 1, \dots, 3N$. Notice that these normal coordinates in general depend on all $3N$ physical coordinates and describe oscillations of the whole system. Since there are $3N$ coordinates there are $3N$ normal modes. After diagonalising the Hamiltonian becomes

$$\hat{H} = \sum_{r=1}^{3N} \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial q_r^2} + \frac{1}{2} m \omega_r^2 q_r^2 \right]. \quad (10.2.5)$$

From here the process is very similar to the Einstein model so we won't go into too much detail.

The normal modes are quantum oscillators. We call the quanta of energy **phonons**. The total energy is

$$E = \sum_{r=1}^{3N} \hbar \omega_r \left(n_r + \frac{1}{2} \right) \quad (10.2.6)$$

where n_r is the number of phonons in mode r . The canonical partition function is then

$$Z_c = \prod_{r=1}^{3N} Z_r \quad (10.2.7)$$

where Z_r is the partition function for mode r and is given by

$$Z_r = \frac{e^{\beta\hbar\omega_r/2}}{1 - e^{-\beta\hbar\omega_r}}. \quad (10.2.8)$$

The average energy is then

$$\bar{E} = -\frac{\partial}{\partial\beta} \ln Z_c = \text{constant} + \sum_r \frac{\hbar\omega_r}{e^{\beta\hbar\omega_r} - 1}. \quad (10.2.9)$$

The constant term isn't important since we are interested in the heat capacity, which is defined as the temperature derivative of this.

The second term is the same as an ideal Bose gas at $\mu = 0$, so phonons behave as if they were bosons. The factor $1/(\exp(\beta\hbar\omega) - 1)$ can then be interpreted as the mean number of phonons in mode r and $\hbar\omega_r = E_r$ as the energy of a phonon in mode r . We can treat the system as an ideal Bose gas of free phonons, free since $\mu = 0$. The number of phonons is not conserved, which is why $\mu = 0$, essentially it takes no energy to create or remove phonons, so long as the net energy remains constant the number of phonons is free to change.

If we assume that the density of modes, g , is known as a function of ω then we can write the sum as an integral over the density of modes:

$$\bar{E} \approx \text{constant} + \int \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} g(\omega) d\omega. \quad (10.2.10)$$

The problem is that we cannot in general accurately calculate g , since for any reasonable number of particles the calculation becomes quickly infeasible. Instead we need an approximation.

10.2.1 Debye Theory

Debye theory is an approximation of the density of modes. Recall that for a gas of free particles, which we can treat the phonons as, the density of states in energy-space is

$$g(\varepsilon) = \frac{\Gamma(k)}{d\varepsilon/dk} \quad \text{where} \quad \Gamma(k) = g_s \frac{V}{2\pi^2} k^2 \quad (10.2.11)$$

where we have added a spin degeneracy factor, g_s .

In Debye theory we regard phonons as sound waves and we can then use the same density of modes that follows from the relation

$$k = \frac{\omega}{c_s} \quad (10.2.12)$$

where c_s is the speed of sound in the material, which we assume to be constant. There are two transverse and one longitudinal wave modes and so $g_s = 3$. Changing variables to $\omega = \varepsilon/\hbar$ we have

$$g(\omega) d\omega = \Gamma(k) dk \quad (10.2.13)$$

which gives

$$g(\omega) d\omega = 3 \frac{V}{2\pi^2} \left(\frac{\omega}{c_s} \right)^2 \frac{1}{c_s} d\omega \quad (10.2.14)$$

and so

$$g(\omega) = 3 \frac{V}{2\pi^2} \frac{\omega^2}{c_s^3} =: AV\omega^2 \quad (10.2.15)$$

where we define the constant $A := 3/(2\pi^2 c_s^3)$ for notational simplicity. Debye theory simply posits that this is the correct form for the density of modes.

Recall that there are $3N$ normal modes. We therefore need to introduce an cut-off frequency, ω_{\max} so that

$$\int_0^{\omega_{\max}} g(\omega) d\omega = 3N. \quad (10.2.16)$$

This gives

$$\omega_{\max} = \left(\frac{9}{A} \frac{N}{V} \right)^{1/3} = \left(6\pi^2 \frac{N}{V} \right)^{1/3} c_s. \quad (10.2.17)$$

Notice that this depends only on the number density, N/V , of the phonons. We can define a characteristic temperature, Θ_D , called the **Debye temperature**:

$$k_B \Theta_D = \hbar \omega_{\max}. \quad (10.2.18)$$

We can then take Θ_D as a free material dependent parameter to be determined from experiments.

The mean energy is then

$$\bar{E} = \int_0^{\omega_{\max}} \hbar \omega g(\omega) \overline{n(\omega)} d\omega \quad (10.2.19)$$

$$= AV\hbar \int_0^{\omega_{\max}} \frac{\omega^3}{e^{\beta\hbar\omega} - 1} d\omega. \quad (10.2.20)$$

We first consider the high temperature limit, $T \gg \Theta_D$. This implies $\hbar\omega/(k_B T) \ll 1$ and we can expand $e^{\beta\hbar\omega} \approx 1 + \beta\hbar\omega$. We then get

$$\bar{E} \approx AV\hbar \int_0^{\omega_{\max}} \frac{\omega^3}{\beta\hbar\omega} d\omega = 3Nk_B T. \quad (10.2.21)$$

This is, as expected, the same result as we get from equipartition and the Einstein model, which we know work well in the high temperature regions.

In the lower temperature limit, $T \ll \Theta_D$ a change of variables to $x = \beta\hbar\omega$ gives the mean energy as a constant times a dimensionless integral:

$$\bar{E} = \frac{AV\hbar}{(\beta\hbar)^4} \int_0^{\Theta_D/T} \frac{x^3}{e^x - 1} dx. \quad (10.2.22)$$

For $T \ll \Theta_D$ we can replace the upper limit with ∞ . This integral can be done analytically or numerically estimated, either way its value is $\pi^4/15$. The important thing is it is a finite, nonzero, constant. At low temperatures we can see that $\bar{E} \propto T^4$ and so $C_V \propto T^3$, which agrees well with experiments.

At low temperatures only low frequency (and so low energy) modes are excited and it is these modes which are modelled well by the Debye theory as sound waves. This allows Debye theory to work at low and high temperatures, but in between it is approximate.

10.2.2 The Integral

While the exact value of the integral isn't important it is my favourite integral so it would be amiss not to demonstrate how one can come to an analytic solution. We do so for a more general integral,

$$I = \int_0^{\infty} \frac{x^{p-1}}{e^x - 1} dx. \quad (10.2.23)$$

First we need a few facts from complex analysis, first the Riemann zeta function is defined for $s \in \mathbb{C}$ with $\text{Re}(s) \geq 1$ as

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}. \quad (10.2.24)$$

The gamma function is defined for $z \in \mathbb{C}$ with $\text{Re}(z) > 0$ by the integral

$$\Gamma(z) := \int_0^{\infty} t^{z-1} e^{-t} dt. \quad (10.2.25)$$

We therefore have

$$I = \int_0^{\infty} \frac{x^{p-1}}{e^x - 1} dx \quad (10.2.26)$$

$$= \int_0^{\infty} x^{p-1} e^{-x} \frac{1}{1 - e^{-x}} dx \quad (10.2.27)$$

$$= \int_0^{\infty} x^{p-1} e^{-x} \sum_{n=0}^{\infty} e^{-nx} dx. \quad (10.2.28)$$

Here we have identified the geometric series which converges uniformly for $|e^{-x}| < 1$, which is always the case for $x \in \mathbb{R}$ and so we can exchange the sum and the integral giving

$$I = \sum_{n=0}^{\infty} \int_0^{\infty} x^{p-1} e^{-x} e^{-nx} dx \quad (10.2.29)$$

$$= \sum_{n=0}^{\infty} \int_0^{\infty} x^{p-1} e^{-x(n+1)} dx. \quad (10.2.30)$$

Reindexing the sum to start at $n = 1$ this becomes

$$I = \sum_{n=1}^{\infty} \int_0^{\infty} x^{p-1} e^{-xn} dx. \quad (10.2.31)$$

Now let $y = nx$ and so $dy = n dx$ and we have

$$I = \sum_{n=1}^{\infty} \int_0^{\infty} \frac{y^{p-1}}{n^{p-1}} e^{-y} \frac{dy}{n} \quad (10.2.32)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^p} \int_0^{\infty} y^{p-1} e^{-y} dy \quad (10.2.33)$$

$$= \zeta(p) \Gamma(p). \quad (10.2.34)$$

The Riemann zeta function and gamma function can be evaluated to arbitrary precision and for some values can be evaluated exactly. $p = 4$ is one of these values giving $\zeta(4) = \pi^4/90$ and $\Gamma(4) = 3! = 6$, hence

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \zeta(4)\Gamma(4) = \frac{\pi^4}{15}. \quad (10.2.35)$$

Index

B

Bayesian, 1
Bose–Einstein distribution, 32
boson, 31
bridge equation, 17

C

canonical ensemble, 10
canonical partition function, 11
chemical potential, 14
coefficient of variation, 25
conjugate field, 13

D

density matrix, 41
density of states, 35
dilute limit, 33
distinguishable particles, 29

E

entropy
 Boltzmann, 2
 Gibbs, 2
 Shannon, 6
 von Neumann, 42
extensive, 14

F

Fermi–Dirac distribution, 32
fermion, 31
frequentist, 1

G

generalised force, 13
Gibbs–Duhem relation, 22
grand canonical ensemble, 11
grand canonical partition function, 12
grand potential, 18, 19

H

Helmholtz free energy, 17, 19

I

instantaneous pressure, 16
intensive, 14

K

k_B , Boltzmann’s constant, 2

L

Legendre transform, 21

M

microcanonical ensemble, 10
mixed state, 42

P

partition function
 canonical, 11
 grand canonical, 12
pure state, 42

R

response function, 28

S

semi-classical approximation, 34
Stirling’s approximation, 34

T

thermodynamic displacement, *see* conjugate field
thermodynamic force, *see* generalised force

W

weakly interacting, 29
weight, 2