

Willoughby Seago

Theoretical Physics

Statistical Physics

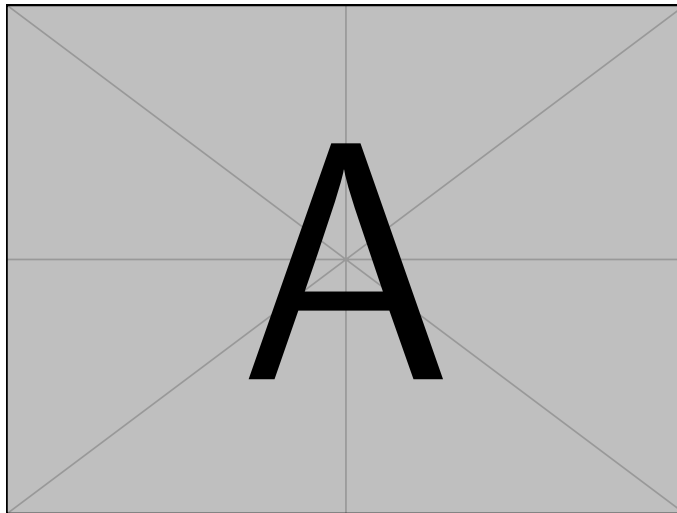
COURSE NOTES

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 19:21 on January 29, 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
Index	24

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
Index	24

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.

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.

Index

B

Bayesian, 1
bridge equation, 17

C

canonical ensemble, 10
canonical partition function, 11
chemical potential, 14
conjugate field, 13

E

entropy
 Boltzmann, 2
 Gibbs, 2
extensive, 14

F

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

P

partition function
 canonical, 11
 grand canonical, 12

T

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

W

weight, 2