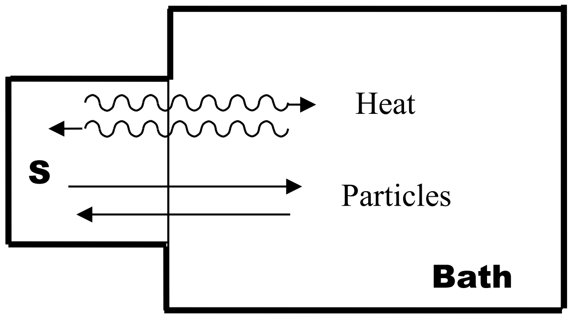
CHAPTER 4: SUMMARY OF THE GRAND CANONICAL PRESCRIPTION AND SOME FURTHER DEVELOPMENTS

4a Summary of the grand ensemble (read 2nd year notes for details)



(0) Systems in diffusive equilibrium have equal chemical potential μ.

(1) The Grand Ensemble is appropriate for a system S (of any size) in both *thermal* and '*chemical'* equilibrium contact with a much larger Bath. (4a.1)

(2) The Bath's inverse temperature β, and either its chemical potential μ – or, equivalently, its *fugacity* $\xi \equiv \exp(\beta\mu)$ – are assumed to be given. (4a.2)

(3) The probability that system S is in microstate (*i,N*) with *N* particles and energy *E*(*i,N*) is

$$\begin{aligned} p_{i,N} &= Q^{-1}\exp(-\beta(E(i,N)-\mu N)) \\ &= Q^{-1}\exp(-\beta E(i,N))\xi^N \end{aligned} \quad \text{(Gibbs distribution)} \tag{4a.3}$$

(4) The *Grand Partition Function* (*Gibbs Sum*) in (4a.3) is

$$Q = \sum_N \sum_i \exp(-\beta E(N,i) - \beta\mu N) = \sum_N \sum_i \exp(-\beta E(N,i))\xi^N \tag{4a.4}$$

(5) The particle number *N* is indeterminate (it fluctuates) but its mean value is

$$\langle N \rangle = \sum_{N,i} p_{i,N} N = \frac{1}{\beta}\frac{\partial}{\partial\mu}\ln Q(\beta,\mu,...) = \xi\frac{\partial}{\partial\xi}Q(\beta,\xi,...) \tag{4a.5}$$

(6) The Internal energy is

$$U = \langle E \rangle = -\frac{\partial}{\partial\beta}\ln(Q(\beta,\xi,...)) \tag{4a.6}$$

(7) The equilibrium expectation value of any property θ with value (or quantum expectation value) θ(*i,N*) for microstate (*i,N*) is

$$\langle\theta\rangle = \sum_{N,i} p_{i,N}\theta_{i,N} \tag{4a.7}$$

4b Application of Grand Ensemble to noninteracting Fermions or Bosons (Summary from 2nd year notes, Ch 6).

The states of this system can be specified by the set of occupation numbers (N_1, N_2,…,N_k, …) of the 1-particle "orbitals" k.

For Bosons, N_k = 0,1,2,3,4,5,....
For Fermions N_k= 0,1.

Even though the particles are non-interacting (do not have forces acting between them), the dynamics of the particles are not truly independent because of the symmetry (for Bosons) or antisymmetry (for Fermions) of the many-body wavefunctions. This means the theory of classical (distinguishable) non-interacting (or weakly interacting) particles does not work, at sufficiently low temperatures / high chemical potentials. In the Grand Ensemble where total number is not conserved, we can however treat each orbital k (rather than each particle) as an independent system. Note that for Bosons, people often use "mode" rather than "orbital" because of the analogy with modes of the electromagnetic field, where photons can be thought of as Bosons occupying a mode.

We treat each orbital k as a system in contact with a heat-and-particle bath consisting of all the other particles with chemical potential μ. From (4a.3) we can obtain the Grand Partition Function of orbital k as

$$Q_k = \sum_{N=0}^{\infty} \exp(-\beta N E_k + \beta\mu N) = \frac{1}{1-\exp(-\beta(E_k-\mu))} \quad \text{for Bosons} \qquad (4b.1)$$

$$Q_k = \sum_{N=0}^{1} \exp(-\beta N E_k + \beta\mu N) = 1+\exp(-\beta(E_k-\mu)) \quad \text{for Fermions} \qquad (4b.2)$$

Then from (4a.5) we can differentiate with respect to μ to give the mean number of particles $\bar{N}_k$ on an orbital k:

$$\bar{N}_k = \frac{1}{\exp(\beta(E_k-\mu)) \pm 1} \quad \text{with + for Fermions, – for Bosons} \qquad (4b.3)$$

Discussion: How is this expression different from the calculation of the mean number of photons in a single-mode thermal state? Why?

Later we will use this expression to understand some features of Bose-Einstein condensation. For now we add just one thing that was not covered in the second year Notes, namely the entropy of a collection of noninteracting Fermions or Bosons.

4c Shannon entropy of independent Bosons or Fermions from the Grand distribution

Recall from paragraph (3.21) that for independent systems, the entropy is additive. Thus we can work out the entropy of each orbital, and then add them. In each orbital, we have probabilities $p_N = Q_k^{-1}\exp(-\gamma_k N)$ where $\gamma_k = \beta(E_k - \mu)$, and $Q_k = (1 \pm e^{-\gamma_k})^{\pm 1}$ for Fermions and Bosons respectively (i.e. top and bottom). Thus the entropy for orbital k is

$$\begin{aligned}
S_k &= -k_B \sum\nolimits_N p_N \ln p_N \\
&= -k_B \sum\nolimits_N Q_k^{-1} \exp(-\gamma_k N)[-\ln Q_k - \gamma_k N] \\
&= k_B\left(\ln Q_k + \gamma_k \bar{N}_k\right) = k_B\left[\ln(1 \pm e^{-\gamma_k})^{\pm 1} + \gamma_k (e^{\gamma_k} \pm 1)^{-1}\right] \\
&= k_B\left[\pm \ln(1 \pm e^{-\gamma_k}) + \gamma_k (e^{\gamma_k} \pm 1)^{-1}\right]
\end{aligned}$$

where the final expression involves only $\gamma_k = \beta(E_k - \mu)$. Alternatively, using (4b.3), we have $\gamma_k = \ln(\bar{N}_k^{-1} \mp 1)$ for Fermions/Boson, and

$$S = k_B\left(\pm \ln(1 \pm (\bar{N}_k^{-1} \mp 1)^{-1}) + \bar{N}_k \ln(\bar{N}_k^{-1} \mp 1)\right)$$

Exercise: Show that this can be rearranged as $S = -k_{\mathrm{B}}(\bar{N}_k \ln \bar{N}_k \pm (1 \mp \bar{N}_k)\ln(1 \mp \bar{N}_k))$

Thus, summing over all orbitals/modes, we have the total entropy of

$$\boxed{S = -k_{\mathrm{B}} \sum_k (\bar{N}_k \ln \bar{N}_k \pm (1 \mp \bar{N}_k)\ln(1 \mp \bar{N}_k))} \qquad \begin{matrix}\text{Fermions}\\ \text{Bosons}\end{matrix} \qquad (4c.8)$$

Exercise: show that S > 0 for both Fermions and Bosons.

Ch. 5 INFORMATION THEORETIC APPROACH TO STATI STICAL MECHANICS

C. E. Shannon, Bell System Technical Journal **27**, 379 (1948)
E. T. Jaynes, Phys. Rev. **106**, 171b (1957)

How should we define "*information*"? Shannon sought to express the "uncertainty" or "**lack** of information" $\mathcal{S}$ associated with a probability distribution f_i ($0 \le f_i \le 1$ and $\sum_i f_i = 1$) by the following postulates:

(i) $\boxed{\mathcal{S}(\{f_i\}) \text{ is a differentiable, non-negative, single-valued function of the } f_i \qquad (5.1)}$

(ii) When all the probabilities are equal, i.e. $f_i = 1/N,\ \ i = 1,\ldots,N$ the entropy increases with N:

$\boxed{\text{The function } A(N) \equiv \mathcal{S}(\{1/N, 1/N, 1/N, \ldots\}) \text{ is monotonically increasing.} \qquad (5.2)}$

(iii) If there is only one possible state, the entropy vanishes: $\boxed{\mathcal{S}(\{1,0,0,\ldots\})=0 \qquad (5.3)}$

(iv) If the states are grouped into subsets, then the total entropy is the average of the entropies within the subsets, plus the entropy associated with the distribution over subsets

$$\underbrace{i = 1,2,\ldots,n_1;}_{\substack{\text{subset } s_1\\ \text{(sates up}\\ \text{to \# } n_1)}}\ \ \underbrace{n_1+1, n_1+2\ldots,n_2;}_{\substack{\text{subset } s_2\\ \text{(further}\\ \text{states up}\\ \text{to \# } n_2)}}\ \ \bullet\bullet\bullet\bullet\bullet\ \ \underbrace{;\, n_{k-1}, n_{k-1}+1,\ldots,n_k}_{\substack{\text{subset } s_k\\ \text{(remaining}\\ \text{states up to}\\ \text{\# } n_k = N)}}\ \ ,\ \ \text{where } n_k \equiv N.$$

The probability of a particular event being in subset s_j is $w_j = \sum_{i \in s_j} f_i = \sum_{i=n_{j-1}+1}^{n_j} f_i$

Before we find out which subset the system state lies, there is an entropy associated, with that uncertainty, $\mathcal{S}(\{w_j\})$. Once we find out that the system state is in some particular subset j, we still have some uncertainty: there is a *conditional* probability of finding the system in state i, given it is in subset j, equal to f_i/w_j for $i \in s_j$. The conditional entropy is equal to $\mathcal{S}(\{w_j/f_j\}|i \in s_j)$. Thus the total uncertainty is postulated to be the uncertainty associated with choice of subset, plus a weighted sum of conditional uncertainties associated with events inside the subsets:

$$\boxed{\mathcal{S}(\{f_i\}) = \mathcal{S}(\{w_j\}) + \sum_{j=1}^{k} w_j\ \ \mathcal{S}(\{f_i/w_j\}|i \in s_j) \qquad (5.4)}$$

The above postulates (5.1- 4) turn out to define the function $\mathcal{S}$ completely, and it will turn out to agree with the formula for the entropy of the previous chapters.

The following is a hand-wavy derivation of the above claim. Consider the case of $N = mn$ equally likely events, $f_i = 1/(mn)$, with uncertainty $A(mn)$ (in the notation of the second postulate). We divide these events into m subsets, each containing n events. Then the probability of a subset is $w_j = 1/m$: $j=1,...,m$, and the conditional probability of an event inside j is $(1/mn)/(1/m) = 1/n$. The uncertainty inside each subset is $A(n)$. Then the postulate (5.4) becomes

$$A(mn) = A(m) + m.(1/m)A(n) = A(m) + A(n) \tag{5.5}$$

By the first postulate we may differentiate this relationship with respect to n (if we ignore the fact that the arguments are integers!). Remembering $mn = N$, or $n = N/m$,

$$m\frac{dA(N)}{dN} = \frac{dA(n)}{dn}$$

and multiplying both sides by n,

$$N\frac{dA(N)}{dN} = n\frac{dA(n)}{dn}.$$

Thus

$$x\frac{dA(x)}{dx} = C = \text{const}, \qquad \frac{dA(x)}{dx} = C\frac{1}{x}.$$

$$\therefore A(x) = C\ln x + d$$

The constant of integration d must be zero from (5.3). Thus $A(x) = C\ln(x)$ where C is a positive constant, by postulate (i).

Now keep $f_i = 1/N$ but allow breaking them up k *arbitrary* (possible unequal) subsets s_j, $j = 1,..,k$, containing numbers of events given by $\alpha_1,\alpha_2,....,\alpha_k$ and with probabilities $w_j = \alpha_j/N$. Returning to the fourth postulate we find

$$A(N) = \mathcal{S}(\{w_j\}) + \sum_{j=1}^{k} w_j A(w_j N)$$

$$C\ln(N) = \mathcal{S}(\{w_j\}) + \sum_{j=1}^{k} w_j C\ln(w_j N) = \mathcal{S}(\{w_j\}) + C\sum_{j=1}^{k} w_j \ln(w_j) + C\left(\sum_{j=1}^{k} w_j\right)\ln(N)$$

$$= \mathcal{S}(\{w_j\}) + C\sum_{j=1}^{k} w_j \ln(w_j) + C\ln(N)$$

Thus

$$\boxed{\mathcal{S}(\{w_j\}) = -C\sum_{j=1}^{k} w_j \ln(w_j) \qquad (5.6)}$$

Shannon uncertainty ("ignorance", lack of information)

This is **identical to the generalised entropy** (eq. (3.15)) derived from canonical considerations in a previous Chapter, provided that we choose the constant to be $\boxed{C = k_B}$.
Shannon's quantitative ideas on information have also been applied in fields from all areas of knowledge, including economics, genetics, linguistics, computing, and communication.
We can derive all the standard equilibrium ensembles by maximising this ignorance, subject to suitable constraints. First we outline the theory of Lagrange multipliers, typically used to handle maximum/minimum problems with constraints.

AN ASIDE: LAGRANGE MULTIPLIERS

The problem: Find where the function $F(\mathbf{r}) = F(x_1,x_2,...,x_N)$ is max or min (stationary) ,
SUBJECT TO THE CONSTRAINTS
$G_1(\mathbf{r}) = c_1,\quad G_2(\mathbf{r}) = c_2, \ldots\ldots\ldots\ldots \quad G_M(\mathbf{r}) = c_M$

- Each contraint condition specifies a **constraint surface** in the vector space R^N.
- The points **r** satisfying all the constraints constitute the **line of constraint** (it is the intersection of all constraint surfaces)
- We only require that F be stationary under displacements **dx'** **lying along the line of constraint.**
- Thus we require that

$$0 = \delta F^{(\text{constrained})} = \vec{\nabla} F \bullet d\vec{x}' \qquad (5.7)$$

which means that ∇F needs to be perpendicular to the line of constraint.

- Since the line of constraint lies in all constraint surfaces, it is perpendicular to all the **constraint–normal vectors** ∇G_m, m=1,...,M. In particular **dx'** is perpendicular to all constraint normals,

$$d\vec{x}' \bullet \vec{\nabla} G_m = 0 \qquad (5.8)$$

- **The solution:** To achieve (5.7) it is therefore sufficient that ∇F be an arbitrary linear combination of the constraint-normal vectors,

$$\vec{\nabla} F = \lambda_1 \vec{\nabla} G_1 + \lambda_2 \vec{\nabla} G_2 + \ldots + \lambda_M \vec{\nabla} G_M, \qquad (5.9)$$

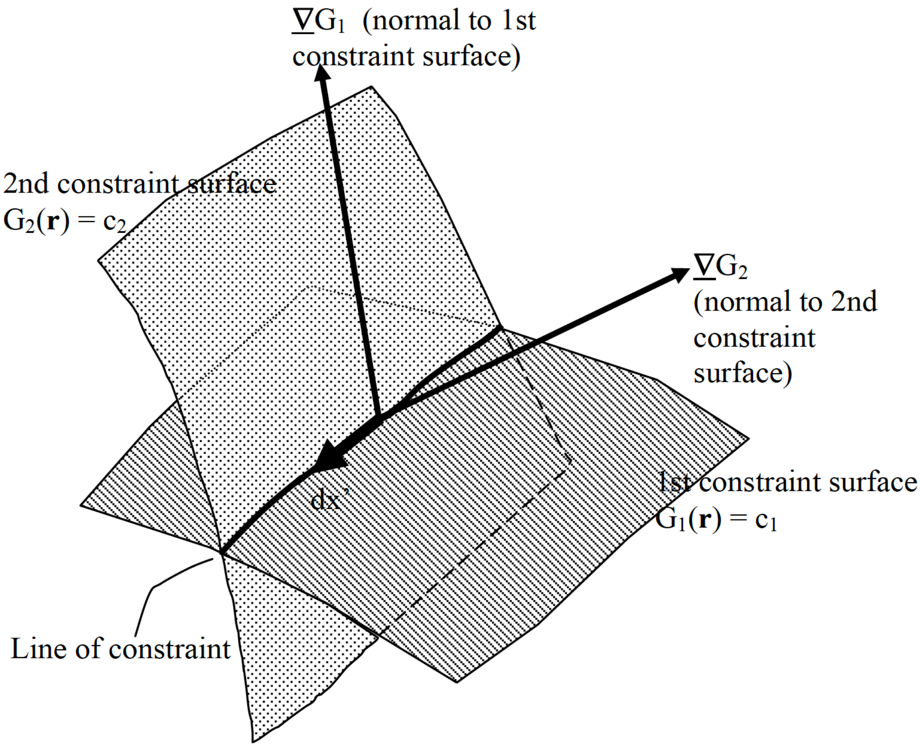
(Check:Using (5.9) and (5.8) we find (5.7) is satisfied).

- The **Langrange multipliers** $\{\lambda_m;\ m=1,...,M\}$ are constants determined by insisting that the point **r** satisfying (5.9) actually lies on the **chosen** line of constraint (i.e. the that the G functions take the **particular** values $\{c_m\}$ specified at the outset, and not some other constant values.)
- The condition (5.9) is more conveniently written

$$\vec{0} = \vec{\nabla}\Big(F - \sum_{m=1}^{M} \lambda_m G_m\Big)$$

- Thus the problem looks like UNCONSTRAINED minimization of the MODIFIED FUNCTION

$$\tilde{F}(\vec{r}) = F(\vec{r}) - \sum_{m=1}^{M} \lambda_m G_m(\vec{r}) \qquad (5.10)$$



Derivation of microcanonical ensemble from Information Theory

The general information-theoretic approach to statistical mechanics is that we assume maximum ignorance, subject to what we do know (typically the mean values of macroscopic variables). That is, we are looking for a solution in the vector space of $\vec{f} \equiv \{f_i\}$ which corresponds to maximum entropy $\mathbb{S}(\{f_i\})$, subject to certain constraints.

In the microcanonical situation we have an isolated system able to access a set of g_N possible states, all of energy E (= U). The only constraint on the probabilities is normalisation:

$$G_1(\vec{f}) \equiv \sum_{i=1}^{g_N} f_i = 1 \quad \text{(1st constraint condition)} \tag{5.11}$$

We maximise the function $F(\vec{f}) \equiv \mathbb{S}(\{f_i\})$ subject to the above constraint.

From (5.10) this is equivalent to maximising $\tilde{F} = F - \lambda G$ with no constraint on the $\{f_i\}$, where λ is the Lagrange multiplier:

$$0 = \delta\tilde{F} = \delta\Big(\sum_{i=1}^{g_N} f_i \ln(f_i) - \lambda f_i\Big) = \sum_{i=1}^{g_N} \delta f_i\Big[\ln(f_i) + f_i\frac{1}{f_i} - \lambda\Big]$$

Since the $\{\delta f_i\}$ are arbitrary, the square bracket must vanish, so

$\ln(f_i) = \lambda - 1$ = a constant (independent of i) ➔ f_i **is also independent of** $\boldsymbol{i}$.

The normalisation condition (5.11) is achieved when the Lagrange multiplier takes a value such that $\exp(\lambda - 1) = f_i = 1/g_N$

Thus the distribution maximizing uncertainty is

$$\boxed{f_i = 1/g_N} \tag{5.12}$$

This is indeed the microcanonical distribution in which all accessible microstates (those of the fixed energy E) are equally likely, as demanded by the postulate of Equal a Priori Proibability, in our previous way of approaching Ensemble Theory. See (2.5b). The Shannon entropy is then (see also item (iv), p.7)

$$\mathbb{S} = -k_{\mathrm{B}}\sum_{i=1}^{g_N}\frac{1}{g_N}\ln(1/g_N) = +k_{\mathrm{B}}\ln(g_N)$$

To make contact with thermodynamics we equate the Shannon entropy $\mathbb{S}$ with the thermodynamic entropy S and use the thermodynamic relation

$$\mathrm{d}U = \mathrm{d}Q = T\mathrm{d}S$$

so

$$\frac{1}{T} = \frac{\partial S}{\partial U} = \frac{\partial}{\partial U}(k_{\mathrm{B}}\ln(g_N))$$

$$\frac{1}{k_{\mathrm{B}}T} = \frac{\partial(\ln g_N)}{\partial U}$$

This was the basic formula of the microcanonical ensemble in our previous approach (p. 7).

Derivation of the Canonical Ensemble from Information theory

Here we have a system in contact with a heat bath so the energy is not fixed. Instead we specify a certain *average* energy (the internal energy), giving a second constraint.

$$G_1(\vec{f}) = \sum_i f_i = 1 \qquad \text{(normalisation)}$$

$$G_2(\vec{f}) = \sum_i f_i E_i = U \qquad \text{(internal energy specified)}$$

Discussion point: why is reasonable to say that, when energy is allowed to flow into and out of a system from a heat bath, we know only the average energy of the system. Why not also the average of the energy squared? After all, in an isolated system the energy squared is also a conserved quantity.

To maximise the uncertainty $F(\vec{f}) \equiv \mathbb{S}(\{f_i\})$ subject to these two constraints we maximise the related function

$$\tilde{F} = F - \lambda_1 G_1 - \lambda_2 G_2$$

with no constraints on the variations (see 5.10). Thus

$$0 = \delta\tilde{F} = \delta(-k_\mathrm{B}\sum_i f_i \ln f_i - \lambda_1 \sum_i f_i - \lambda_2 \sum_i f_i E_i) = \sum_i \delta f_i[-k_\mathrm{B}\{\ln f_i + \frac{1}{f_i} f_i) - \lambda_1 - \lambda_2 E_i]$$

Since the $\{\delta f_i\}$ are arbitrary the square bracket must vanish:

$$k_\mathrm{B} \ln(f_i) = -k_B - \lambda_1 - \lambda_2 E_i$$

$$f_i = Z^{-1}\exp(-\beta E_i), \quad Z^{-1} = \exp(-1 - \lambda_1 / k_\mathrm{B}), \quad \beta = \lambda_2 / k_\mathrm{B}$$

This is of course the usual canonical (Boltzmann) distribution.

Exercise: show that the above constrained maximisation is equivalent to the minimisation of the Helmholtz free energy $F = U\text{-}T\mathbb{S} = \sum_i f_i(E_i + k_B T \ln f_i)$ considered as a function $F(\vec{f})$ subject to the constraint of normalization.

Grand ensemble from Information Theory

This ensemble corresponds to a system in contact with a heat and particle bath. As a result, neither E nor N is fixed and we have a distribution $f_{i,N}$ over both particle number N and microstate i given N. Instead of specifying fixed values of E or N we therefore specify the average values

$$\sum_{i,N} f_{i,N} E_i = U, \quad \sum_{i,N} f_{i,N} N = \bar{N}$$

Exercise: taking these, plus normalisation, as constraints, maximise the Shannon uncertainty to obtain the Grand canonical distribution $f_{i,N} = Q^{-1}\exp[\beta(-E_i + \mu N)]$

Exercise: Show that the Grand canonical distribution equivalently is the minimization of $U\text{-}T\mathbb{S} - \mu\bar{N}$, subject to normalization of the distribution.