Statistical Physics

Tanguy Marsault

Preface

Contents

1   Warm up discussions

1.1   Classical and quantum descriptions

Before diving really into the beauty of statistical mechanics, we recall some notions about classical systems. We mostly follow the discussion of the Appendix A to chapter 3 of [3] In order to achieve this we need to recall a few important points about the formulations of classical and quantum mechanics.

1.1.1   Classical mechanics

We emphasize on the Hamiltonian formulation of classical mechanics (which is equivalent to the Lagrangian formulation) in which the system is described by its *phase space* vector,

$$\mathbf{\Gamma} = (q_i, p_i) \tag{1.1.1}$$

where the p_i are the conjugate momenta associated to the q_i. To be precise, we suppose that i ranges from 1 to N. i can label particles or any kind of degree of freedom of the system. For instance, a system with k particles will be described by their positions in the D-dimensional space considered and their conjugate momenta in each dimension as well, that is to say $N = 2Dk$. For terminology we call $\mathbf{\Gamma}$ a microstate.

$$\boxed{\textbf{Microstate} \quad \longleftrightarrow \quad \mathbf{\Gamma}} \tag{1.1.2}$$

The evolution of the system is ruled by *Hamilton's equations* which read,

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i} \tag{1.1.3}$$

where H is a function from the phase space of the system to the real numbers which describes the energy of the system in a specific phase space configuration. We call it the *Hamiltonian*.

This formulation raises a simple question, why should we introduce conjugate momenta p_i instead of just computing everything in terms of the classical variables q_i and their temporal derivatives ?

This question is easily answered by using the *Liouville's invariant* which states that the elementary volume in the phase space $\mathrm{d}\mathbf{\Gamma}$ is invariant under evolution, which needs not be the case if we were working with the q_i and $\dot{q}_i$ for instance. Before proving this result, let us explain what this means. As we shall see in the next sections, what is under study in statistical physics is the probability density function of the phase space. It is described by a function $\rho(\mathbf{\Gamma})$ such that the quantity

$$\rho(\mathbf{\Gamma})\mathrm{d}\mathbf{\Gamma}$$

represents the probability of being in a volume $\mathrm{d}\mathbf{\Gamma}$ around $\mathbf{\Gamma}$. Hence, in general, to study a physical system one needs to know the full evolution of

$$\rho(\mathbf{\Gamma}(t))\mathrm{d}\mathbf{\Gamma}(t).$$

This would mean that one needs to take into account variation of the volume $\mathrm{d}\mathbf{\Gamma}$ along physical trajectories to really compute the relevant physical quantities. To this regard, Liouville's invariant is a very powerful statement as it ensures that along physical trajectories (that is

to say trajectories that for Hamitlon's equations, or if you want to be even more precise that minimizes the action),

$$d\boldsymbol{\Gamma}(t) = d\boldsymbol{\Gamma}(t_0), \quad \forall t \tag{1.1.4}$$

where t_0 denotes an arbitray fixed time.

We now prove the claim. Under evolution of the system during an infinitesimal time δt, one can write the variables

$$q_i(t + \delta t) = q_i(t) + \dot{q}_i(t)\delta t + o(\delta t)$$
$$p_i(t + \delta t) = p_i(t) + \dot{p}_i(t)\delta t + o(\delta t)$$

Using Hamilton's equations yield,

$$q_i(t + \delta t) = q_i(t) + \frac{\partial H}{\partial p_i}\delta t + o(\delta t)$$

$$p_i(t + \delta t) = p_i(t) - \frac{\partial H}{\partial q_i}\delta t + o(\delta t)$$

Now let us compute the elementary volume product, $dq_i(t + \delta t)dp_i(t + \delta t)$. It is given by,

$$dq_i(t + \delta t)dp_i(t + \delta t) = |J|dq_i(t)dp_i(t) \tag{1.1.5}$$

where J is the change of variable matrix given by,

$$J = \begin{pmatrix} 1 + \frac{\partial^2 H}{\partial q_i \partial p_i}\delta t & \frac{\partial^2 H}{\partial p_i^2}\delta t \\ -\frac{\partial^2 H}{\partial q_i^2}\delta t & 1 - \frac{\partial^2 H}{\partial p_i \partial q_i}\delta t \end{pmatrix}$$

At first order in δt, we get

$$|J| = 1 + \left(\frac{\partial^2 H}{\partial q_i \partial p_i} - \frac{\partial^2 H}{\partial p_i \partial q_i} \right)\delta t = 1 \tag{1.1.6}$$

assuming that H is smooth enough. This proves the result.

One could be more pedantic (some would say precise) and call back to symplectic geometry. In fact, the phase space is a symplectic manifold and the evolution of the system is described by a symplectic transformation. The Liouville's theorem is then a direct consequence of the symplectic nature of the phase space, it reads that the Lie derivative of the volume form along the vector field that describes the evolution of the system is zero,

$$\mathcal{L}_X \Omega = 0$$

where X is the Hamiltonian vector field. We shall not go further into this rather mathematical explanation, but it is important to keep in mind that classical mechanics can be made extremely precise using a purely mathemaical language.

1.1.2 Quantum mechanics

Surprisingly quantum mechanics description of microstates is much easier. In quantum mechanics states are represented by a normalized vector of an Hilbert space $\mathcal{H}^*$. Possible states

---

*Actually the correct description is to take physical states to be represented by rays of the Hilbert space, but we will come back to this issue in due times.

of a system in quantum mechanics are the eigenvectors of some operator under study. Most of the time, we will study the eigenvectors of the Hamiltonian H^*.

In general we will resolve to Dirac's notation of *bras* and *kets* and will denote the quantum states by $|l\rangle$. These states evolve in time according to Schrödinger's equations,

$$i\hbar \frac{\mathrm{d}\,|l\rangle}{\mathrm{d}t} = H\,|l\rangle \qquad (1.1.7)$$

This way, in quantum mecahnics, a microstate is just represented by a ket of the system

$$\boxed{\textbf{Microstate} \quad \longleftrightarrow \quad |l\rangle \in \mathcal{H}} \qquad (1.1.8)$$

Quantum mechanics become tricky when it comes to combining different systems together. Classically, combining systems together just amounts to stacking them up in a vector, that is to say, if $\boldsymbol{\Gamma}_1$ represents the state of the first system and $\boldsymbol{\Gamma}_2$ represents the one of the second system, we build up the combination of these two systems by considering only the vector[†],

$$\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) \qquad (1.1.9)$$

In quantum mechanics we now need to combine states of two Hilbert spaces and the natural way to do that is to consider the tensor product of the Hilbert spaces (with obvious notations),

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \qquad (1.1.10)$$

Contrary to classical mechanics quantum mechanics isn't so nice because in general the physical states won't be all the normalized states of the tensor product Hilbert space, but rather by a subspace which depends on the spin statistics theorem. In fact, depending on the nature of the spin of the combined systems, the obtained states need to be symmetric (boson) or antisymmetric (fermion) with respect to particle exchange. Because of this, one needs to keep in mind that when particles are identical, simply taking the states,

$$|l_1\rangle \otimes |l_2\rangle \quad (l_1, l_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \qquad (1.1.11)$$

as the physical states involve a huge number of overcounting. Namely, for n identical particles, symmetrization or antisymmetrization divide the number of states by $n!$[‡].

The above discussion may seem quite unclear for those not really acquainted with multiparticle quantum physics. The easiest (or laziest) way to carry on is to just admit the result we will encounter later without caring too much about the reason it appears. Otherwise, it's fairly easy to find some materials that explains really clearly the way the Hilbert space needs to be symmetrized or antisymmetrized in the litterature, for instance [1]. Though, it might take a lot of reading to get a full grasp of it if the reader isn't mathematically inclined.

1.1.3 Ergodicity and macrostates

The whole point of statistical physics is to study collective phenomena without having to go through the microstates description of the system. One generally wants to know the macroscopic characteristics of a system, namely its energy, temperature, pressure,... For this reason

---

[*]We use the same notation for the quantum Hamiltonian and the classical one unless we need to be precise, in which case we shall note the quantum Hambiltonian $\hat{H}$.

[†]Up to some rearrangements if we really want to have q_{1i}s next to q_{2i}s...

[‡]That's actually not quite true, but we'll come back to this later on.

we define and study macrostates.

Let $\mathcal{A}$ be a quantity that characterizes the macroscopic state of the system. The value of $\mathcal{A}$ depends on the vector $\boldsymbol{\Gamma}$ or the state $|l\rangle$, depending on the description of the system we choose. There are two ways to consider the average of $\mathcal{A}$.

One can calculate its statistical average

$$A^* = \sum_l P_l \left\langle \widehat{\mathcal{A}} \right\rangle_l \tag{1.1.12}$$

$$= \int \mathrm{d}^3 \boldsymbol{\Gamma} \rho(\boldsymbol{\Gamma}) \mathcal{A}(\boldsymbol{\Gamma}) \tag{1.1.13}$$

(the first equality is considered in the quantum case and the second in the classical case). P_l denotes the probability of being in the state l, and $\rho(\boldsymbol{\Gamma})$ is a sort of probability density function on the phase space.

Note that defined this way ρ has a unit, which is not really satisfying as it is defined as a probability density function. We come back to this matter later on, including the correct units. For now only the general form needs to be understood, as the actual correct form needs much more work to be derived.

Alternatively, one can calculate the time average of $\mathcal{A}$:

$$\bar{A}_{\mathcal{T}} = \frac{1}{\mathcal{T}} \int_t^{t+\mathcal{T}} \mathrm{d}t' \mathcal{A}(\boldsymbol{\Gamma}(t')) \tag{1.1.14}$$

In practice, and this can only be demonstrated in simple cases, we make the ergodic hypothesis, which assumes that the processes are so chaotic that over an interval of time of size $\mathcal{T}$ tending to infinity, the system will pass through all the states $\boldsymbol{\Gamma}$ and will remain in each state for a duration $\tau_{\boldsymbol{\Gamma}} = \mathcal{T}\rho(\boldsymbol{\Gamma})$. The ergodic hypothesis thus leads us to write:

$$\lim_{\mathcal{T}\to\infty} \bar{A}_{\mathcal{T}} \simeq A^* \tag{1.1.15}$$

This is why we will only study A^*. Therefore, for any quantity $\mathcal{A}$ that we want to study, we need to know $\rho(\boldsymbol{\Gamma})$ (or P_l), which then defines what we will call a macrostate:

$$\boxed{\textbf{Macrostate} \quad \longleftrightarrow \quad \rho(\boldsymbol{\Gamma})} \tag{1.1.16}$$

As for microstates, in quantum mechanics we will have rather:

$$\boxed{\textbf{Macrostate} \quad \longleftrightarrow \quad \{P_l, \ l \in \mathcal{H}\}} \tag{1.1.17}$$

From now on we will no longer distinguish between the classical and quantum descriptions (except when explicitly stated), and we will simply denote by l a microstate.

1.1.4 Discrete states and continuous states

In quantum mechanics we generally have a very important number of states that are *close* to each other. To emphasize this statement let us take the exampe of a free particle in a D-dimensional box of size L^D. Its wave function is given by,

$$\psi(\mathbf{x}) \propto e^{i\mathbf{k}\cdot\mathbf{x}} \tag{1.1.18}$$

with the boundary conditions,

$$\psi(\ldots, x_i = 0, \ldots) = \psi(\ldots, x_i = L, \ldots) \tag{1.1.19}$$

which in turn enforces,

$$k_i = n_i \frac{2\pi}{L}, \quad n_i \in \mathbb{Z} \tag{1.1.20}$$

Thus a quantum state of the particle (ignoring the spin) is given by a D-dimensional multiplet $(n_1, \ldots, n_D)$ and so the sum over states should be written as

$$\sum_{n_1, \ldots n_D}$$

In practice, if L is large compared to the characteritics wavelength of the particle, then the state are extremely close to each other and we can approximate the sum using an integral. One just needs to understand the density of states in the wavevectors space. We count a single state in any square of volume $\frac{(2\pi)^D}{L^D}$ and then we get,

$$\sum_{n_1, \ldots n_D} = \frac{L^D}{(2\pi)^D} \int d^D \mathbf{k} \tag{1.1.21}$$

That we can recast more generally as,

$$\sum_{n_1, \ldots n_D} = \frac{1}{(2\pi)^D} \int d^D \mathbf{r} d^D \mathbf{k} = \frac{1}{h^D} \int d^D \mathbf{r} d^D \mathbf{p} \tag{1.1.22}$$

That's what we call the semiclassical approximation.

1.2 Microstates counting

1.2.1 General Idea

In everything that will follow, and as can already be noticed from the shape of the Hamilton's and Schrödinger's equations, the energy plays a very special and central role in physics. And so in general we will be interested in computing quantities that depend on the energy. To be more precise, let us denote a physical quantity f that depends on the microstate l^* only through its energy E_l. We can be interested in its mean value, that would be defined by,

$$\bar{f} = \sum_l P_l f(E_l) \tag{1.2.1}$$

where P_l denotes the probability of the system to be in state l, as per usual. However, we pointed out in the last paragraph that in general the spectrum of states can be quite dense and it seems reasonable to replace the sum by an integral. As we are focusing particularly on the energy, we write

$$\bar{f} = \sum_l P_l f(E_l) = \int dE \rho(E) g(E) f(E) \tag{1.2.2}$$

where $g(E)$ is the density of states. Roughly speaking, $g(E)dE$ is the number of states between E and $E + dE$. It can be seen as the continuous extension of the degeneracy. In the

---

*We focus from now on on a quantum approach to statistical mechanics, as in the end, it is the one true law!

previous equation, $\rho(E)$ replaces P_l as the probability density of finding a particle in a state with energy between E and $E + dE$. Note that this example assumed that the probability density of states only depend on their energy, which may not always be the case.

It's quite easy to find the correct form of g,

$$g(E) = \sum_l \delta(E - E_l) = \sum_{E_l} g_l \delta(E - E_l) \tag{1.2.3}$$

where we introduced the degeneracy g_l of a state l in the second equation. It is not always convenient to work with Dirac's delta functions, and so it might be easier to work with the integrated density of state,

$$\Phi(E) = \int^E dE' g(E') \tag{1.2.4}$$

It corresponds to the number of states with energy smaller than E. This equation can be equivalently written as,

$$g(E) = \frac{d\Phi(E)}{dE} \tag{1.2.5}$$

which shows that it is completely equivalent to work with g or Φ. Then using g's expression, one quickly finds,

$$\Phi(E) = \sum_l \Theta(E - E_l) \tag{1.2.6}$$

where Θ denotes Heaviside's function.

1.2.2   Particle in a box

Let us now go back to the semiclassical approximation introduced in the previous section and let us first study a free particle in a D-dimensional box. As a free particle its energy is given by,

$$E_{\mathbf{k}} = \frac{\hbar^2 \mathbf{k}^2}{2m} \tag{1.2.7}$$

Our goal is to use the semiclassical approximation and to find its density and integrated density of states. Formally it is straightforward to write,

$$\Phi(E) = \sum_{\mathbf{k}} \Theta\left(E - \frac{\hbar^2 \mathbf{k}^2}{2m}\right) \tag{1.2.8}$$

where as explained above the sum over $\mathbf{k}$ is really a sum over D integers ranging over $\mathbb{Z}$. Using the semi classical approximation, the computation of $\Phi(E)$ is fairly straightforward,

$$\Phi(E) = \sum_{\mathbf{k}} \Theta\left(E - \frac{\hbar^2 \mathbf{k}^2}{2m}\right) = \frac{1}{(2\pi)^D} \int d^D \mathbf{r} d^D \mathbf{k} \Theta\left(E - \frac{\hbar^2 \mathbf{k}^2}{2m}\right) \tag{1.2.9}$$

This can be rewritten as,

$$\Phi(E) = \frac{V}{(2\pi)^D} \int_{\|\mathbf{k}\| < \sqrt{2mE}/\hbar} d^D \mathbf{k} \tag{1.2.10}$$

The volume of the D-dimensional sphere of radius r is given by

$$V_D(r) = \frac{\pi^{D/2}}{\Gamma(\frac{D}{2} + 1)} r^D \tag{1.2.11}$$

Thus the integrated density of states is given by,

$$\Phi(E) = \frac{V}{\Gamma(\frac{D}{2}+1)} \left(\frac{mE}{2\pi\hbar}\right)^{D/2} \tag{1.2.12}$$

And we can now differentiate this equation to get the density of states g,

$$g(E) = \frac{d\Phi}{dE} = \frac{V}{\Gamma(\frac{D}{2}+1)} \left(\frac{m}{2\pi\hbar}\right)^{D/2} E^{D/2-1} \tag{1.2.13}$$

Let us briefly expand this formula for the low dimensional cases.

- $D = 1$ the density of states reads,

$$g_{1D}(E) = \frac{2L}{\sqrt{\pi}} \sqrt{\frac{m}{2\pi\hbar}} \frac{1}{\sqrt{E}} \tag{1.2.14}$$

- $D = 2$ the density of states reads,

$$g_{2D}(E) = S\frac{m}{2\pi\hbar} \tag{1.2.15}$$

  The density of state is a constant!

- $D = 3$ the density of states reads,

$$g_{3D}(E) = \frac{4V}{3\sqrt{\pi}} \left(\frac{m}{2\pi\hbar}\right)^{3/2} \sqrt{E} \tag{1.2.16}$$

An important remark is that we worked with a particle of mass m without taking its spin into account. For a particle of spin s, $g(E)$ should also be multiplied by a factor $2s + 1$ to take account of the spin degenracy.

1.2.3 Particle in a potential

Now consider the particle not te be free anymore but confined in some potential. In this way we get a very genreral Hamiltonian $H(\mathbf{r}, \mathbf{p})$. We then assume without proof that the space filled by a state in phase space is still h^d, and so one can compute the integrated density of states as,

$$\Phi(E) = \frac{1}{h^d} \int d\mathbf{r}d\mathbf{p}\Theta(E - H(\mathbf{r}, \mathbf{p})) \tag{1.2.17}$$

1.2.4 Identical particles

When there are several particles, in general a potential will emerge because of the interactions between the particles. In general one can then write the Hamiltonian of the system as,

$$H(\mathbf{r}_i, \mathbf{i}_i) = \sum_i \frac{\mathbf{p}_i^2}{2m} + U(\mathbf{q}_i) \tag{1.2.18}$$

But we now need to be careful as we make our states continuous. In fact, before we've been using the sum over quantum states without being precise but as we study identical particles there is a new subtelty because of quantum mechanics. As explained above, as we get many particles the representative states form a subspace of the tensor product of

the different Hilbert spaces $\otimes_i \mathcal{H}_i$ of each particle. This subspace depends on the nature of particles, namely if these are bosons or fermions. In both cases, a representation of a physical state will be a combination of $N!$ states of the tensor product space so that a state can be iether totally antisymmetric or symmetric. In this way, the sum

$$\sum_l$$

should have an index l ranging over a very complicated set of multiplets.

Fortunately we are not doomed and we can work around that quite easily. Let's be naive for a second and take $l_1, \ldots, l_{N!}$ to be vectors of the tensor product of Hilbert spaces $\otimes_i \mathcal{H}_i$, which can all be obtained as permutation of single particle states in the tensor product. Denote l the actual physical states that already takes into account the symmetric nature of the state. Let's now take a function f that depends on the state of the particles.

f is physical and so doesn't care about the symmetry of the state or wave function under interchange of particles. This means that,

$$\sum_{i=1}^{N!} f(l_i) = \sum_{i=1}^{N!} f(l) = N! f(l) \tag{1.2.19}$$

And so we see that instead of summing over l we can just as well sum over the multiplets of single particle states (that I will call *naive states* from now on) as long as we remember to divide by $N!$ at the end of the computation.

If you are convinced by this *proof*, let me announce you gently that you got scammed. In fact, there is something I hid under the rug in the previous derivation, namely the fact that two particles may be in the same state. To make this clear, let us consider two particles with two states denoted by 0 and 1. The possible *naive states* are $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. As explained above, $|10\rangle$ and $|01\rangle$ really are the same state from a quantum point of view and we should count it only once, hence the divison by $2! = 2$. However, for the two other states, $|00\rangle$ and $|11\rangle$, we are not overcounting and we should not divide by 2 in this case.

Let us now suppose that we have d possible states per particle. There are d^2 *naive states*, d of them, namely the $|00\rangle, \ldots, |dd\rangle$ are not overcounted when we sum independently over states of the first and second particle. However, the $d^2 - d$ other states really are overcounted $2!$ times and their contribution should be divide by $2!$. Hence we see that the number of overcounted states increases much more dramatically than the number of states that we are not overcounting. Let us now see what happens when it comes to N particles.

Consider an ensemble of N identical particles with d possible states. From the discussion just above, it is straightforward to convince ourselves that we a *naive state* is not overcounted $N!$ times as soon as two particles are in the same state $|i\rangle$. Let us now count how many of these states there are.

Let us then consider a naive state, $|i_1, \ldots, i_N\rangle$. There are d^N of these states. To count the number of states that have 2 or more particles in the same state we'd rather count the number of states for which all particles are in different states. This is really easy to count and is given by,

$$d(d-1)\ldots(d-N+1) = N!\binom{d}{N} \tag{1.2.20}$$

The number of naive states with two or more particles in the same state is then given by,

$$d^N - N!\binom{d}{N} \tag{1.2.21}$$

as the total number of naive states is d^N. Let us now evaluate what fraction these states represent as we increase the number of states. This reads,

$$\frac{\text{\#2 or more particles in same state}}{\text{\#Total number of states}} = \frac{d^N - N!\binom{d}{N}}{d^N} = 1 - \frac{N!\binom{d}{N}}{d^N} \tag{1.2.22}$$

Let us focus on the second term, it can be rewritten as

$$\frac{N!\binom{d}{N}}{d^N} = \frac{1}{d^N}\frac{d!}{(d-N)!} \tag{1.2.23}$$

Let us now use Stirling's asymptotic series for $d!$,

$$n! = n^n e^{-n}\sqrt{2\pi n}(1 + \frac{1}{12n} + o(n^{-1})) \quad \text{when } n \to \infty \tag{1.2.24}$$

This allows us to simplify the fraction of factorials,

$$
\begin{aligned}
\frac{d!}{(d-N)!} &= \frac{d^d e^{-d}\sqrt{2\pi d}}{(d-N)^{d-N}e^{-d+N}\sqrt{2\pi(d-N)}}\frac{1 + \frac{1}{12d} + o(d^{-1})}{1 + \frac{1}{12(d-N)} + o(d^{-1})} \\
&= e^{-N}\frac{d^{1/2}}{(d-N)^{-N+1/2}}\left(\frac{d}{d-N}\right)^d \frac{1 + \frac{1}{12d} + o(d^{-1})}{1 + \frac{1}{12} + o(d^{-1})} \\
&= e^{-N}\frac{d^{1/2}}{(d-N)^{-N+1/2}}\left(\frac{1}{1-\frac{N}{d}}\right)^d \left(1 + \frac{1}{12d} + o(d^{-1})\right)\left(1 - \frac{1}{12d} + o(d^{-1})\right) \\
&= e^{-N}\frac{d^{1/2}}{(d-N)^{-N+1/2}}e^{-d\log\left(1-\frac{N}{d}\right)}(1 + o(d^{-1}))
\end{aligned}
$$

In the limit where d goes to infinity we can derive a first order asymptotic expansion for the exponential as,

$$e^{-d\log\left(1-\frac{N}{d}\right)} = e^{-d(-N/d - N^2/2d^2 + o(d^{-2}))} = e^N\left(1 + \frac{N^2}{2d} + o(d^{-1})\right) \tag{1.2.25}$$

Injecting this expression back in yields,

$$
\begin{aligned}
\frac{d!}{(d-N)!} &= \frac{d^{1/2}}{(d-N)^{-N+1/2}}(1 + o(d^{-1}))\left(1 + \frac{N^2}{2d} + o(d^{-1})\right) \\
&= \frac{d^{1/2}}{(d-N)^{-N+1/2}}\left(1 + \frac{N^2}{2d} + o(d^{-1})\right)
\end{aligned}
$$

This way, the asymptotic series for the number of states with two or more particles in the

same state is given by,

$$\frac{N!\binom{d}{N}}{d^N} = \frac{1}{d^N}\frac{d!}{(d-N)!}$$

$$= \frac{1}{d^N}\frac{d^{1/2}}{(d-N)^{-N+1/2}}\left(1 + \frac{N^2}{2d} + o(d^{-1})\right)$$

$$= \left(1 + \frac{N^2}{2d} + o(d^{-1})\right)\left(\frac{d-N}{d}\right)^{N-1/2}$$

$$= \left(1 + \frac{N^2}{2d} + o(d^{-1})\right)\left(1 - \frac{N}{d}\right)^{N-1/2}$$

$$= \left(1 + \frac{N^2}{2d} + o(d^{-1})\right)\left(1 - \frac{N(N-1/2)}{d} + o(d^{-1})\right)$$

$$= 1 - \frac{N(N-1)}{2d} + o(d^{-1})$$

Finally the fraction of states is given by,

$$\frac{\text{\#2 or more particles in same state}}{\text{\#Total number of states}} = \frac{N(N-1)}{2d} + o(d^{-1}) \tag{1.2.26}$$

This means that as the number of states d increases, the fraction of state we don't over-count $N!$ times by summing over the naive states goes to zero as $\frac{1}{d}$. This way as d increases, the number of states to which we apply a wrong overcounting rule goes to 0.

Thinking about it, $\frac{1}{d}$ can seem to be a slow convergence but the situation is much less dramatic than the way we described it. In fact, for all the states that have exaclty 2 particles in the same state, eventhough we don't overcount them $N!$ times, we still overcount them a lot, namely, $(N-1)!$ times. Only the states where all particles are in the same state are actually not overcounted at all.

If we now go back to our function f depending on the states and suppose that we actually have a continuum of states described for instance by some wavevectors $(k_1, \ldots, k_N)$. Assuming that f is well behaved enough, when we go to the continuum limit the $N!$ prescription becomes exact as the measure of the set of states with two or more particles with the same wavevector in the space of states is 0.

However, let us be clear that, as far as we know, quantum mechanics is right and everything should be quantized, hence we never have a continuum of states. The $N!$ prescription and the classical approximation remain true as long as f really is well behaved. In general, f will be the probability density of states and it will be up to us to know if we are allowed to use the classical approximation. We will tackle these issues again in due times but for now we shall just use the classical approximation with the $N!$ prescription as described above. We will just remember that,

$$\Phi(E) = \frac{1}{N!h^{dN}}\int\prod_i \mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{p}_i\Theta(E - H(\mathbf{r}_i, \mathbf{p}_i)) \tag{1.2.27}$$

This is still not quite general as $\mathbf{r}$ and $\mathbf{p}$ need not be the coordinates appearing explicitly in the Hamiltonian. The generalization is extremely easy as q_i and p_i, the generalized coordinates appearing in the Lagrangian are really quantized in the same way as $\mathbf{r}$ and $\mathbf{p}^*$. In

---

*Please refer to [4] for a full description of quantum mechanical quantization of classical variables

particular, they also verify Heisenberg's uncertainty principle and so states (p_i, q_i) should also fill a space h in phase space.

1.3 Entropy

After all these formalities, that have nothing specific to do with statistical physics, we now turn to the actual topic at stake. Let us introduce the main character of our discussion that is *entropy*. Entropy will be a measure of disorder in the systems we study. Introduced in this way, it's not entirely clear to what extent this entropy should be identified with the classical entropy used in thermodynamics. Some theoretical arguments could be made that these two are the same, but I don't find them quite convincing and I'm actually not sure I fully grasp them. The best argument that I could give to state that these two entropies are in fact equal is that physics only boils down to one thing at the end of the day, *does it work in real life*. This identification actually works, and this will be our starting point.

1.3.1 Shannon's entropy

A measure of disorder is not quite precise enough to get an explicit formula for the entropy. Let us think for a while and undertand what we'd like the entropy to verify. Suppose we have a system that can be in a number of states we denote by l. The system is in state l with probability P_l. Now a measure of the disorder should quantify what we know about the system, hence it should be a function of the P_ls. Let us denote this real function

$$S(\{P_l\}_l) \tag{1.3.1}$$

First of all, as the probabilities P_ls vary smoothly, we would like the entropy to vary smoothly as well, S should be continuous in the P_ls. Let us now think about some extreme cases of this function. What we would call a system with minimum disorder would be a system that is almost surely in some state l_0, that is to say,

$$P_{l_0} = 1 \quad \text{and} \quad P_l = 0, \ \forall l \neq l_0 \tag{1.3.2}$$

In this case we would like the entropy to be at the minimum, that is to say $S = 0$.

The other extreme case is when we know nothing about the system, that is to say every state is as likely to happen as any other,

$$P_l = \frac{1}{n} \tag{1.3.3}$$

when there are n accessible states. In this case the entropy should be at its maximum. Furthermore, as n increases there should be even more states and so the disorder should be even greater. That's why S should also be a strictly increasing function of n in this case.

These two conditions are already quite constraining. However, we need to think about more involved cases if we want to really find the only possibility for the entropy. To be clear, let us start with an example. Suppose we have a system characterized by two parameters, its color and shape. Either it is green or red, and either it is a triangle or a square. In total we have 4 possibilities. However, we see these characteristics can be split into two different categories we can study independently. The entropy of this system of four states, should be the entropy that the system is either red or green to which we add the entropy

that the system is either a square or a triangle assuming that is green and multiplied by the probability it is green plus the entropy that the system is either a square or a triangle assuming that is red and multiplied by the probability it is red. It is probably easier to understand in a formula,

$$
\begin{aligned}
&S\left(P(G,S), P(G,T), P(R,S), P(R,T)\right) \\
&= S\left(P(G), P(R)\right) + P(G)S\left(P(S|G), P(T|G)\right) + P(R)S\left(P(S|R), P(T|R)\right)
\end{aligned}
$$

We know have all the charecteristics we need to derive the entropy, first derived by Shannon in [2]. Let us sum up the important bits that S must satisfy.

(i) S is a continuous function of the P_l

(ii) When $P_l = 1/n$ for all l, S should reach its maximum and be a strictly increasing function of n.

(ii) Denote $\mathcal{H}$ the possible states of the system. Suppose we can write $\mathcal{H}$ as a disjoint union,

$$
\mathcal{H} = \biguplus_{i=1}^{K} \mathcal{H}_i \tag{1.3.4}
$$

Then

$$
S(\{P_l\}) = S(\{P_{\mathcal{H}_i}\}) + \sum_{i=1}^{K} P_{\mathcal{H}_i} S(\{P_{l^i}\}) \tag{1.3.5}
$$

The condition (iii) can appear quite involved but it is just the mathematical translation of what we discussed above with very childish words.

We now prove Shannon's theorem that states that S must be of the form,

$$
S(\{P_l\}) = -K \sum_l P_l \log P_l \tag{1.3.6}
$$

First note that the previous definition appears as ill-defined when some $P_l = 0$, but it can actually be extended continuously at any point of $\mathbb{R}^n$ where some P_l are zero.

Let us now suppose we have a function H^* satisfying the three conditions above. Define $A(n)$ as,

$$
A(n) = H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \tag{1.3.7}
$$

We first show that $A(st) = A(s) + A(t)$ for s, t integers. That's easy to do using the composition rule. Having st states, we split them into s packs of t states. This yields,

$$
H\left(\frac{1}{st}, \ldots, \frac{1}{st}\right) = H\left(\frac{1}{s}, \ldots, \frac{1}{s}\right) + \sum_{i=1}^{s} \frac{1}{s} H\left(\frac{1}{t}, \ldots, \frac{1}{t}\right) \tag{1.3.8}
$$

which is just $A(st) = A(s) + A(t)$. This in turn implies that $A(r) = -K \log r$.

Now suppose we have $P_l \in \mathbb{Q}$. We can always put the vector of probabilities in the form

$$
(P_1, \ldots, P_N) = (\frac{r_1}{T}, \ldots \frac{r_N}{T}) \tag{1.3.9}
$$

---

*This H has nothing to do with the Hamiltonian. We're just following Shannon's notations.

with $T \in \mathbb{N}$. (In fact one just needs ot find the least common multiple of the denominators of the P_i.)

Then splitting correctly the states, one finds,

$$H(P_1, \ldots, P_N) = A(T) + \sum_{i=1}^{N} \frac{r_i}{T} A(r_i) \tag{1.3.10}$$

As $\sum_i P_i = 1$, $\sum_i r_i = K$ and so,

$$H(P_1, \ldots, P_N) = -K \log T - K \sum_{i=1}^{N} \log r_i = -K \sum_{i=1}^{N} \frac{r_i}{T} \log \frac{r_i}{T} \tag{1.3.11}$$

which then becomes,

$$H(P_1, \ldots, P_N) = -K \sum_i P_i \log P_i, \quad \forall P_i \in \mathbb{Q} \tag{1.3.12}$$

Then as Q is dense in $\mathbb{R}$,

$$H(P_1, \ldots, P_N) = -K \sum_i P_i \log P_i, \quad \forall P_i \in \mathbb{R} \tag{1.3.13}$$

It's easy to verify that the obtained function satisfies the required condition, and so we showed Shannon's theorem.

In statistical physics, K is called the Boltzman's constant and denoted by k_B*.

1.3.2   Properties of entropy

Let us now check that entropy behaves like we would want it to. We mention a few properties of the entropy below.

- *Minimum.* The entropy in minimum when one of the states l_0 is alsmost sure, that is to say $P_l = \delta_{ll_0}$. It simply means that we know almost surely the state of the system, there is no disorder.

- *Maximum.* The entropy is maximum when all the states are equiprobable, $P_l = 1/n$, where n is the number of possible states. In this case, we have no information about the system.

- *Monotony.* Entropy increases with the number of states in the equiprobable case.

- *Concavity* Let us denote two probability ditributions on the states l as $\{P_l^0\}$ and $\{P_l^1\}$. For $\lambda \in [0, 1]$,
  $$P_l^\lambda = \lambda P_l^0 + (1 - \lambda) P_l^1 \tag{1.3.14}$$
  is a probability distribution on the states l. It can be shown that the entropy is concave, that is to say,
  $$S(P_l^\lambda) \geq \lambda S(P_l^0) + (1 - \lambda) S(P_l^1) \tag{1.3.15}$$
  It means that if we mix up many probability distributions, the disorder increases and so does the entropy.

---

*This is actually chosen to match with observation and entropy definition in thermodynamics.

2   Equilibrium and Ensembles

In this section we really dive into statistical physics. We will show how our main character, the entropy, really rules statistical physics through the *maximum entropy principle*. Using this we will derive our first ensemble to study a system, namely the *microcanonical ensemble*. From this ensemble we'll climb our way up to the other ensembles that are the *canonical* one and the *grand canonical* one.

Let us consider a system $\mathcal{S}$, knowing its accessible states l, the only thing we are missing to describe it are its probabilities P_l. How should we determine them ?

The point of view adopted in statistical physics is to admit our ignorance and choose the P_l to minimize the information on the system, that is to say maximize the entropy. This will be our fundamental principle in statistical physics,

A system at thermodynamical equilibrium will be described by the probability distribution on its states that maximizes its entropy and satisfies the constraints applied on the system.

2.1   Thermodynamical equilibrium

The first thing to note in the *fundamental principle* mentionned above is that the system should be at *thermodynamical equilibrium*. Let us first describe what we mean by thermodynamical equilibrium. A system will be said to be at themordynamical equilibrium

2.2   Microcanonical ensemble

The simplest system we can consider is an isolated one with fixed energy E. First, consider a physical system $\mathcal{S}$ whose energy is fixed at E. We assume that there are a finite number of possible states $\Omega(E)$.

The first thing to do is then to determine the probability distribution of the system using our fundamental principle, namely the *maximum of entropy*. There are actually two ways of doing so. An extremely easy one[*], and a more general method that can be used in various cases. We start by explaining how the latter goes.

Our goal is to maximize the entropy, which is given by the formula,

$$S = k_B \sum_l P_l \log P_l \tag{2.2.1}$$

This formula doesn't encapsulate the full story as it doesn't explain how the P_l should behave. In fact, to take into account the fact that the P_ls are probabilities we need to maximize under the constraint that the sum of the P_ls must be equal to one. There actually exist a mathematical theory on how to do that, it goes by the name of *Lagrange multipliers*. Without diving into the theory, it basically tells us that to minimize a function $f(x_i)$ under a constraint $g(x_i) = 0$, one needs to minimize the Lagrange function,

$$\mathcal{L}(x_i, \lambda) = f(x_i) + \lambda g(x_i) \tag{2.2.2}$$

with respect to x_i and λ.

---

[*]One could almost call it luck

Let us use it in our special case when $x_i = P_l$, f is the entropy and the constraint g is the fact that the P_ls should add up to 1. This yields,

$$\mathcal{L}(P_l, \lambda) = -k_B \sum_l P_l \log P_l + \lambda \left(\sum_l P_l - 1 \right) \tag{2.2.3}$$

Differentiating this function with respect to P_l and λ and setting the result to zero yields,

$$-k_b \left(\log P_l + P_l \cdot \frac{1}{P_l} \right) + \lambda = 0 \quad \text{and} \quad \sum_l P_l = 1 \tag{2.2.4}$$

We recover the constraint on the P_ls [*] and we solve for P_l using the first equation. This gives a constant value for the P_ls, that we don't even

need to express. Then using the constraint the summ of the P_ls must be equal to one, so that

$$P_l = \frac{1}{\Omega(E)} \tag{2.2.5}$$

we now show that this could have been derived directly by showing that the uniform distribution does indeed maximize the entropy. [†]

Let u be the uniform distribution and P be an arbitrary distribution on the microstates l.

$$
\begin{aligned}
S(p) - S(u) &= -k_B \sum_x P_l \ln P_l + \ln u \\
&= -k_B \sum_x P_l \ln P_l + k_B \sum_x P_l \ln u_l \\
&= k_B \sum_x P_l \ln \frac{u_l}{P_l} \\
&\leq k_B \sum_x P_l \left(1 - \frac{u_l}{P_l} \right) \\
&= 0
\end{aligned}
$$

where we used $u_l = u = \frac{1}{\Omega}$, $S(u) = -\sum_x u_l \ln u_l = -\ln u$, and the concavity of the logarithm. This in turn implies that the uniform distribution does indeed maximize the entropy.

We've reached ou first key checkpoint in studying our physical system from a macroscopic point of view as we can now determine its entropy,[‡]

$$S = k_b \ln \Omega(E) \tag{2.2.6}$$

Now that we have the entropy, still assuming that we can identify the thermodynamical entropy with the statistical one, we can define various quantities, that we hope can be identified with their well known thermodynamical counterpart. The way to do that is to use the famous thermodynamic identities, in the special case where the energy is constant

The first macroscopic characteristic we are interested in is the temperature as it is usually defined in thermodynamics by simply recalling the classical thermodynamic identity,

$$\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V + \mu\mathrm{d}N + x\mathrm{d}X \tag{2.2.7}$$

---

[*]That's actually the whole point of the Lagrange multipliers method.

[†]One could then raise the question of guessing that the uniform distribution would be the correct one. Hating myself for it, I would simply say that the uniform distribution is the one where we have as *little information* as we possibly can.

[‡]The proof is straightforward and not detailed here.

where x, X refers to some conjugate variables so that the energy of the system can be decrived as a function of S, V, X.

The statistical temperature is then defined by,

$$\frac{1}{T} \equiv \frac{\partial S}{\partial E} \tag{2.2.8}$$

We also define the parameter β that we will use later by,

$$\beta \equiv \frac{1}{k_b T} \tag{2.2.9}$$

Depending on the problem, by analogy with thermodynamics, if we denote V as the volume and N as the number of particles, we can also define pressure and chemical potential:

Définition 2.2.1.

$$p \equiv T \frac{\partial S}{\partial V} \tag{2.2.10}$$

$$\mu \equiv -T \frac{\partial S}{\partial N} \tag{2.2.11}$$

$$\tag{2.2.12}$$

All these relationships have been defined with respect to classical thermodynamic identities, for example:

$$dE = TdS - PdV + \mu dN \tag{2.2.13}$$

Now let's consider several subsystems $\mathcal{S}_i$. For example, let's study the case with two subsystems and the complete system $\mathcal{S} = \mathcal{S}_1 \otimes \mathcal{S}_2$. If the two systems interact very little, the total energy of the system is

$$E_{\mathcal{S}} = E_1 + E_2 + E_{1 \leftrightarrow 2} \simeq E_1 + E_2$$

And the number of states of the system is

$$\Omega(E = E_1 + E_2) = \Omega(E_1)\Omega(E_2)$$

This implies the additivity of entropy:

$$\begin{aligned} S &= k_b \ln \Omega(E = E_1 + E_2) \\ &= k_b \ln \left(\Omega(E_1)\Omega(E_2) \right) \\ &= k_b \left(\ln \Omega(E_1) + \ln \Omega(E_2) \right) \\ S &= S_1 + S_2 \end{aligned}$$

2.3   Canonical ensemble

2.4   Grand canonical ensemble

2.5   Generalized grand canonical ensemble

3 Quantum statistics

Untiln now, everything we've discussed remained rather classical. In fact,

3.1 Varieties of statistics

3.2 Bosons

3.3 Fermions

3.4 Application: The CMB

4   Out of equilibrium pehnomena

5   Phase transitions

6   Introduction to statistical fields

7 Statistical physics, from individual behaviour to collective phenomena

"It is the mark of a truly intelligent person to be moved by statistics"

-George Bernard Shaw

7.1 Foundations

7.2 Microcanonical Ensemble

First, consider a physical system $\mathcal{S}$ whose energy is fixed at E. We assume that there are a finite number of possible states $\Omega(E)$. Using the principle of maximum entropy (i.e., taking the distribution that maximizes entropy subject to the given constraints), we obtain:

Propriété 7.2.1 (Microcanonical Probabilities). In the microcanonical case where the energy of the system is fixed at E, all states l with energy E are equally probable:

$$P_l = \frac{1}{\Omega(E)} \tag{7.2.1}$$

Proof. Let u be the uniform distribution and p be an arbitrary distribution. For the proof, we take $k_b = 1$.

$$
\begin{aligned}
S(p) - S(u) &= -\sum_x p(x) \ln p(x) + \ln u \\
&= -\sum_x p(x) \ln p(x) + \sum_x p(x) \ln u(x) \\
&= \sum_x p(x) \ln \frac{u(x)}{p(x)} \\
&\leq \sum_x p(x) \left(1 - \frac{u(x)}{p(x)}\right) \\
&= 0
\end{aligned}
$$

where we used $u(x) = u = \frac{1}{\Omega}$, $S(u) = -\sum_x u(x) \ln u(x) = -\ln u$, and the concavity of the logarithm. $\square$

This result is generalized to a uniform distribution in the classical case. The associated entropy is then:

Propriété 7.2.2 (Microcanonical Entropy).

$$S = k_b \ln \Omega(E) \tag{7.2.2}$$

Proof. Substitute P_l into the definition of entropy. $\square$

he statistical temperature is defined by analogy with thermodynamics, which in most cases corresponds to the real temperature:

Définition 7.2.3 (Statistical temperature and parameter β). The statistical temperature T is defined by:

$$\frac{1}{T} \equiv \frac{\partial S}{\partial E} \tag{7.2.3}$$

and the parameter β that we will continue to use is defined by:

$$\beta \equiv \frac{1}{k_b T} \tag{7.2.4}$$

Depending on the problem, by analogy with thermodynamics, if we denote V as the volume and N as the number of particles, we can also define pressure and chemical potential:

Définition 7.2.4.

$$p \equiv T\frac{\partial S}{\partial V} \tag{7.2.5}$$

$$\mu \equiv -T\frac{\partial S}{\partial N} \tag{7.2.6}$$

$$\tag{7.2.7}$$

All these relationships have been defined with respect to classical thermodynamic identities, for example:

$$dE = TdS - PdV + \mu dN \tag{7.2.8}$$

Now let's consider several subsystems $\mathcal{S}_i$. For example, let's study the case with two subsystems and the complete system $\mathcal{S} = \mathcal{S}_1 \otimes \mathcal{S}_2$. If the two systems interact very little, the total energy of the system is

$$E_{\mathcal{S}} = E_1 + E_2 + E_{1\leftrightarrow 2} \simeq E_1 + E_2$$

And the number of states of the system is

$$\Omega(E = E_1 + E_2) = \Omega(E_1)\Omega(E_2)$$

This implies the additivity of entropy:

$$\begin{aligned} S &= k_b \ln \Omega(E = E_1 + E_2) \\ &= k_b \ln\left(\Omega(E_1)\Omega(E_2)\right) \\ &= k_b \left(\ln \Omega(E_1) + \ln \Omega(E_2)\right) \\ S &= S_1 + S_2 \end{aligned}$$

7.3   Canonical ensemble

We now consider a system $\mathcal{S} = \mathcal{S}_1 \otimes \mathcal{S}_2$. The goal is to study $\mathcal{S}_1$ for which the fixed energy constraint of the microcanonical study has been relaxed. The fixed total energy is then written as

$$E = E_1 + E_2$$

And let us consider the subsystem $\mathcal{S}_2$ as an energy reservoir such that $E_1 \ll E_2 \simeq E$. The system $\mathcal{S}$, considered as isolated, can be treated using microcanonical methods but not the system $\mathcal{S}_1$. Let $l_1 \otimes l_2$ be the microstates of $\mathcal{S}$, then we have:

$$P_{l_1 \otimes l_2} = \frac{1}{\Omega(E)}$$

We want to determine $P_{l_1}(E_1)$, the probability that system 1 is in microstate l_1 given that its energy is E_1. The law of total probability gives:

$$P_{l_1}(E_1) = \sum_{\substack{l_2 \\ E_2 = E - E_1}} P(l_1 \cap l_2) = \sum_{\substack{l_2 \\ E_2 = E - E_1}} \frac{1}{\Omega(E)} = \frac{\Omega_2(E_2 = E - E_1)}{\Omega(E)}$$

Thus,

$$P_{l_1}(E_1) = \frac{\Omega_2(E_2 = E - E_1)}{\Omega(E)} \tag{7.3.1}$$

Since $E_1 \ll E$, we can expand $\ln \Omega_2(E - E_1)$ in a Taylor series:

$$\ln \Omega_2(E - E_1) = \ln \Omega_2(E) - \left. \frac{\partial \ln \Omega_2}{\partial E_2} \right)_E E_1 + \dots$$

Then, using the definition of β from the previous section (in fact, we show that the equality of temperatures between the two systems is a necessary equilibrium condition, which allows us to make this definition):

$$P_{l_1}(E_1) = \frac{e^{\ln \Omega_2(E-E_1)}}{\Omega(E)} \propto e^{-\beta E_1}$$

Since $E_{l_1} = E_1$, we have:

$$P_{l_1} \propto e^{-\beta E_{l_1}} \tag{7.3.2}$$

We then define the canonical partition function:

Définition 7.3.1 (Canonical partition function)**.**

$$Z = \sum_l e^{-\beta E_l} \tag{7.3.3}$$

which is called the canonical partition function, and since the sum of probabilities is 1, we have:

$$P_l = \frac{e^{-\beta E_l}}{Z} \tag{7.3.4}$$

The probability of system 1 being in microstate l.

Remarque 7.3.2. It is very important to understand the summation index in the partition function. The sum is taken over the microstates l and not over the energies E_l. To sum over the energies E_l, one must take into account the degeneracy of each energy, which is denoted by g_E. Thus, we have

$$\sum_l \longleftrightarrow \sum_E g_E \tag{7.3.5}$$

It turns out that the partition function contains all the information necessary to obtain all the characteristics of the system.

Propriété 7.3.3 (Average Energy)**.**

$$\bar{E} = -\frac{\partial \ln Z}{\partial \beta} \tag{7.3.6}$$

Proof.

$$\bar{E} = \sum_l P_l E_l$$

$$= \frac{1}{Z} \sum_l e^{-\beta E_l} E_l$$

$$= -\frac{1}{Z} \frac{\partial Z}{\partial \beta}$$

$$= -\frac{\partial \ln Z}{\partial \beta}$$

$\square$

It is essential to be able to do this calculation without hesitation, as it is the most important calculation in statistical physics. We can also define the canonical entropy by the calculation:

$$S = -k_b \sum_l P_l \ln(P_l)$$

$$= -k_b \left(\sum_l -P_l \beta E_l - \sum_l P_l \ln Z \right)$$

$$= k_b \ln Z + \frac{\bar{E}}{T}$$

Propriété 7.3.4 (Canonical Entropy).

$$S = k_b \ln Z + \frac{\bar{E}}{T} \tag{7.3.7}$$

This relation echoes the free energy $F = E - TS$, from which we have:

Définition 7.3.5 (Free Energy).
$$F = -k_b T \ln Z \tag{7.3.8}$$

Then, as in thermodynamics, we have:

Propriété 7.3.6 (Chemical Potential).

$$\mu = \frac{\partial F}{\partial N} = -k_b T \frac{\partial \ln Z}{\partial N} \tag{7.3.9}$$

One last important property is:

Propriété 7.3.7 (Heat Capacity).

$$C_v = \frac{\partial E}{\partial T} = \frac{(\sigma_E)^2}{k_b T^2} \tag{7.3.10}$$

Proof. The proof is a good exercise and allows us to manipulate the equality

$$T \frac{\partial}{\partial T} = -\beta \frac{\partial}{\partial \beta}$$

(One should be able to retrieve this operator equality by explicating β; it simplifies many calculations greatly) $\qquad \square$

We can now consider independent systems. We consider N independent subsystems, meaning that for a microstate l of the global system, it is possible to write $l = (\lambda_1, \ldots, \lambda_N)$ with λ_i characterizing the state of subsystem i. The total energy is then written as

$$E = \sum_i E_{\lambda_i}$$

and we have

$$\sum_l \longrightarrow \sum_{\lambda_1} \sum_{\lambda_2} \cdots \sum_{\lambda_N} \tag{7.3.11}$$

Then we have:

Propriété 7.3.8. For N independent subsystems, the global partition function is written as

$$Z = \prod_i z_i \tag{7.3.12}$$

where z_i is the partition function of subsystem i.

Proof.

$$
\begin{aligned}
Z &= \sum_l e^{-\beta E_l} \\
&= \sum_{\lambda_1}\sum_{\lambda_2}\cdots\sum_{\lambda_N} e^{-\beta \sum_i E_{\lambda_i}} \\
&= \left(\sum_{\lambda_1} e^{-\beta E_{\lambda_1}}\right)\left(\sum_{\lambda_2} e^{-\beta E_{\lambda_2}}\right)\cdots\left(\sum_{\lambda_N} e^{-\beta E_{\lambda_N}}\right) \\
Z &= \prod_i z_i
\end{aligned}
$$

$\square$

When the subsystems are identical, this equation reduces to $Z = z^N$. However, when the subsystems are identical and indistinguishable (indiscernible case), we cannot differentiate the ways to number the subsystems as $1, 2, \ldots, N$. Thus, we must divide by $N!$:

Propriété 7.3.9 (Identical indistinguishable subsystems)**.** For N identical indistinguishable subsystems:

$$Z = \frac{z^N}{N!} \tag{7.3.13}$$

We will study the case of the ideal gas where the particles are indistinguishable and the case of a crystal where the particles are identical but distinguishable as they are attached to fixed nodes of the crystal.

Classical approximation and ideal gas Let's place a particle in a box of volume $V = L_x L_y L_z$. Let's choose periodic boundary conditions, i.e., $\psi(\ldots, x_i + L_{x_i}, \ldots) = \psi(\ldots, x_i, \ldots)$. Quantum mechanics tells us that the particle can take wave vectors of the form

$$\vec{k} = 2\pi\left(\frac{n_x}{L_x}, \frac{n_y}{L_y}, \frac{n_z}{L_z}\right), \tag{7.3.14}$$

where n_x, n_y, and n_z are integers. When the box dimensions become large, the different states in wave vector space become increasingly close together (separation of $1/L^3$) and a continuum of states appears in wave vector space that can be expressed in terms of a density of states $\rho(\vec{k})$. Multiplying this density by $d^3\vec{k}$ gives the number of states between $\vec{k}$ and $\vec{k} + d^3\vec{k}$.

The form of the wave vectors shows that each wave vector occupies an elementary volume $\frac{(2\pi)^3}{V}$ in wave vector space. Hence, by cross-multiplication, $\rho(\vec{k}) = \frac{V}{(2\pi)^3}d\vec{k}$. This yields

$$\sum_l \longrightarrow \int d^3\vec{k}\,\frac{V}{(2\pi)^3} = \int d^3\vec{r}\int d^3\vec{p}\,\frac{1}{h^3}, \tag{7.3.15}$$

where we have used $\vec{p} = \hbar\vec{k}$.

A classical application of this result is the calculation of the partition function of an ideal gas of N particles in a volume V. Since there are N indistinguishable particles, we can use the result

$$Z = \frac{z^N}{N!},$$

where z is the partition function of a single particle in the gas. The energy of a particle is $\frac{\vec{p}^2}{2m}$ (ideal gas, so no interactions), so that

$$z = \int d^3\vec{r} \int d^3\vec{p} \frac{1}{h^3} e^{-\beta \frac{\vec{p}^2}{2m}}$$

$$= \frac{V}{h^3} \int dp_x dp_y dp_z e^{-\beta \frac{p_x^2 + p_y^2 + p_z^2}{2m}}$$

$$= \frac{V}{h^3} \left(\int dp e^{-\beta \frac{p^2}{2m}} \right)^3$$

$$z = \frac{V}{\Lambda_T^3}$$

where we have defined $\Lambda_T = \frac{h}{\sqrt{2\pi m k_b T}}$, which is called the de Broglie wavelength.

We will determine more characteristics of the ideal gas in applications.

Propriété 7.3.10. (Boltzmann's equipartition theorem) At thermal equilibrium, each independent squared degree of freedom in the energy is associated with an average energy quantum of $\frac{1}{2} k_B T$.

Proof. In general, we can write the energy in the form

$$E = H(p_i, q_j)$$

We focus on the particular case where the energy takes the following form,

$$H(p_i, q_j) = \alpha p_k^2 + \mathcal{H}(p_{i \neq k}, q_j)$$

Where the second term of the *rhs* of this equation doesn't not depend on p_k.
Let us calculate $\langle \alpha p_k^2 \rangle$:

$$\langle \alpha p_k^2 \rangle = \frac{\int \frac{dp_1}{2\pi\hbar} \ldots dq_1 \ldots \alpha p_k^2 e^{-\beta E}}{\int \frac{dp_1}{2\pi\hbar} \ldots dq_1 \ldots e^{-\beta E}}$$

$$= \frac{\left(\int dq_k \prod_{i \neq k} dp_i dq_i e^{-\beta \mathcal{H}(p_{i \neq k}, q_j)} \right) \left(\int dp_k \alpha p_k^2 e^{-\beta \alpha p_k^2} \right)}{\left(\int dq_k \prod_{i \neq k} dp_i dq_i e^{-\beta \mathcal{H}(p_{i \neq k}, q_j)} \right) \left(\int dp_k e^{-\beta \alpha p_k^2} \right)}$$

$$= -\frac{1}{\int dp_k e^{-\beta \alpha p_k^2}} \frac{\partial}{\partial \beta} \left(\int dp_k e^{-\beta \alpha p_k^2} \right)$$

$$= -\sqrt{\frac{\beta\alpha}{\pi}} \frac{\partial}{\partial \beta} \left(\sqrt{\frac{\pi}{\beta\alpha}} \right) = -\sqrt{\beta} \frac{\partial \beta^{-1/2}}{\partial \beta}$$

$$= \frac{1}{2} \sqrt{\beta} \beta^{-3/2} = \frac{1}{2\beta} = \frac{1}{2} k_b T$$

where we used the important relation :

$$\int_{-\infty}^{+\infty} dx \, e^{-ax^2} = \sqrt{\frac{\pi}{a}}$$

$\square$

7.4   Applications

7.4.1   Spin orientation

Let us consider a crystal with N atoms. The i-th atom carries a magnetic moment $\vec{m}_i = \gamma \widehat{\vec{S}^{(i)}}$. To simplify, we assume that each atom behaves like an electron with a fixed total spin (s) and a spin projection (m_s) on the z axis $\left(\widehat{S}_z^{(i)}\right)$ that can take the values $\pm\hbar/2$. Thus, the i-th spin can be in the state $|\uparrow\rangle$ or $|\downarrow\rangle$.

Any microstate l of the system can then be written as

$$|\uparrow\uparrow\downarrow \cdots \downarrow\rangle$$

Let's place the system in a magnetic field $\vec{B} = B\hat{e}_z$ and assume that the spins do not interact with each other. Consider the Hamiltonian (the energy) of the system which then becomes:

$$\widehat{H} = -\vec{B} \cdot \vec{M} = -B\gamma \sum_i \widehat{S}_z^{(i)} \tag{7.4.1}$$

where $\vec{M} = \sum_i \vec{m}i$ is the total magnetization of the system. Each spin can therefore take 2 energy values: $\pm B\gamma\frac{\hbar}{2} = \pm\epsilon_B$. Let's assume that the system is isolated and therefore its energy is fixed at E (this is really the mantra to remember: if the energy is fixed, the approach is microcanonical; if the temperature is fixed, the approach is canonical). Thus, the number of atoms in the state $|\uparrow\rangle$, which we denote as $n+$, and the number of atoms in the state $|\downarrow\rangle$, which we denote as n_-, are fixed. We then have

$$N = n_+ + n_- \tag{7.4.2}$$
$$E = (-n_+ + n_-)\epsilon_B \tag{7.4.3}$$

These relationships can be inverted and we find:

$$n_\pm = \frac{1}{2}\left(N \mp E/\epsilon_B\right) \tag{7.4.4}$$

The number of microstates that are accessible to achieve the energy E is easily counted (we must select n_+ spins up out of N spins) and is given by

$$\Omega(E) = \binom{N}{n_+} = \frac{N!}{n_+!n_-!} \tag{7.4.5}$$

We can then calculate $S(E, N) = k_b \ln \Omega(E, N)$ using the Stirling's approximation $\ln(N!) \simeq N(\ln N - 1)$, leading to the formula:

$$
\begin{aligned}
S(E, N) &= k_b \ln \Omega(E, N) \\
&= k_b \left[\ln(N!) - \ln(n_+!) - \ln(n_-!)\right] \\
&\simeq k_b \left[N \ln N - N - n_+ \ln n_+ + n_+ - n_- \ln n_- + n_-\right] \\
&= k_b \left[N \ln N - n_+ \ln n_+ - n_- \ln n_-\right] \\
&= Nk_b \left[\ln 2 - \frac{(1 - \xi)\ln(1 - \xi) + (1 + \xi)\ln(1 + \xi)}{2}\right]
\end{aligned}
$$

where we have defined $\xi = E/N\epsilon_B \in [-1, +1]$.

7.4.2 Magnetization

We can also study the previous problem in another way by assuming the temperature to be T and relaxing the constraint on energy E. We now work in the canonical ensemble. There are N identical atoms that are distinguishable because they occupy fixed sites. The canonical partition function is then given by:

$$Z = z^N \tag{7.4.6}$$

where z is the partition function associated with a single atom. We can explicitly calculate z since there are only two possible states per atom with energies $\pm\epsilon_B$:

$$z = e^{-\beta\epsilon_B} + e^{\beta\epsilon_B}$$
$$= 2\cosh(\beta\epsilon_B)$$

We can now calculate the interesting characteristics of the system, starting with the energy:

$$\bar{E} = -\frac{\partial \ln Z}{\partial \beta} \tag{7.4.7}$$

$$= -N\frac{\partial \ln z}{\partial \beta} \tag{7.4.8}$$

$$= -N\epsilon_B \frac{\sinh(\beta\epsilon_B)}{\cosh(\beta\epsilon_B)} \tag{7.4.9}$$

$$\bar{E} = -N\epsilon_B \tanh(\beta\epsilon_B) \tag{7.4.10}$$

We are happy to find an extensive result that shows that the average energy of the system is the sum of the average energies per atom (which reassures us because we assumed that the atoms did not interact with each other). We could have anticipated this result by noting that all the quantities of interest depend on $\ln Z = N \ln z$, so we could calculate each quantity on a per-atom basis and sum them. Since the calculation with Z is not much more complicated, this is the approach we will keep. We can also calculate the entropy:

$$S(N) = k_b \ln Z + \frac{\bar{E}}{T}$$
$$= k_b N \left[\ln(2\cosh(\beta\epsilon_B)) - \beta\epsilon_B \tanh(\beta\epsilon_B)\right]$$

The last quantity of interest that we will calculate is the average magnetization $\bar{M}_z$. It was said earlier that

$$M_z = \sum_i m_z^{(i)}$$

Hence,

$$\bar{M}_z = \sum_i \bar{m}_z^{(i)} = N\bar{m}_z$$

Let's calculate $\bar{m}_z$. There are two possible states for m_z. If the considered spin is up, then $m_z = \gamma\hbar/2$, if it is down $m_z = -\gamma\hbar/2$. The probability of spin up is $e^{\beta\epsilon_B}/z$ (the $+\epsilon_B$ comes from $e^{-\beta(-\epsilon_B)}$), the probability of spin down is $e^{-\beta\epsilon_B}/z$, then we have

$$\bar{m}_z = \frac{1}{z}\left(\frac{\gamma\hbar}{2}e^{\beta\epsilon_B} - \frac{\gamma\hbar}{2}e^{-\beta\epsilon_B}\right)$$

$$= \frac{1}{z}\gamma\frac{\hbar}{2\epsilon_B}\frac{\partial}{\partial\beta}(e^{\beta\epsilon_B} + e^{-\beta\epsilon_B})$$

$$= \frac{1}{B}\frac{\partial \ln z}{\partial\beta}$$

$$= -\frac{\bar{\epsilon}}{B}$$

where $\bar{\epsilon}$ is the average energy per atom.

7.4.3 Properties of the monatomic ideal gas

To conclude this section on statistical physics, let's go back to the model of the ideal gas and try to determine its other characteristics, in particular, to find the ideal gas equation. We recall the result

$$z = \frac{V}{\Lambda_T^3} \text{ with } \Lambda_T = \frac{h}{\sqrt{2\pi m k_b T}} \tag{7.4.11}$$

and $Z = \frac{z^N}{N!}$.

Let's first calculate the average energy:

$$\bar{E} = -\frac{\partial \ln Z}{\partial \beta}$$

$$= -N\frac{\partial \ln z}{\partial \beta}$$

$$= -N\frac{\partial \ln(\beta^{-3/2})}{\partial \beta}$$

$$= \frac{3}{2\beta}N$$

$$\bar{E} = \frac{3}{2}Nk_b T$$

It then naturally follows that

$$C_v = \frac{\partial E}{\partial T} = \frac{3}{2}Nk_b \tag{7.4.12}$$

which corresponds to the classical thermodynamics result for a monoatomic ideal gas (for a diatomic gas, the energy would need to include degrees of freedom associated with rotation). Next, let's calculate the pressure using the thermodynamic identity

$$dF = -SdT - pdV \tag{7.4.13}$$

which tells us that

$$p = -\frac{\partial F}{\partial V}$$

Remembering that $F = -k_b T \ln Z$, we have

$$F = -Nk_b T \ln(V) + \text{const.}(N, T, \ldots) \tag{7.4.14}$$

where the second term is independent of V. Thus, we have

$$p = -\frac{\partial F}{\partial V}$$

$$= Nk_b T\frac{\partial \ln V}{\partial V}$$

$$= Nk_b T\frac{1}{V}$$

$$= nk_b T$$

which is indeed the ideal gas law. We could calculate the other quantities, but as you may have gathered by now, it is not particularly exciting and will be intentionally omitted here.

Conclusion

References

[1]   Michael Astwood. *Lecture notes of Advanced Quantum Theory by Tobias Osborne*. 2021.

[2]   Claude E.Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, 1964.

[3]   Christophe Texier and Guillaume Roux. "Physique statistique, Des processus élémentaires aux phénomènes collectifs". In: (2020).

[4]   Steven Weinberg. *Lecture on Quantum Mechanics, Cambridge*. Cambridge University Press, 2013. ISBN: 978-1-107-02872-2.