First Principles of Statistical Mechanics

M.Simon

July 21, 2015

Abstract

The theory of thermodynamics aims at understanding how two types of energy - mechanical and thermal - can be converted into each other. At the end of the XIXth century, the Austrian physicist Ludwig Boltzmann stated that the laws of thermodynamics should be derived from Newton's mechanical first principles, on the basis of the atomistic theory of matter. A gas, for instance, can be represented as a collection of atoms - or point particles - moving under Newton's laws.

Mechanical systems are mainly characterized by geometric quantities, such as the positions and the velocities of its mass points. However, this description completely fails for gases, which have properties that may interfere with their mechanical characteristics, in particular their temperature. The kinetic theory of gases has turned into what we know as statistical mechanics through the work of the American mathematical physicist J. Willard Gibbs.

In this context, finding the equilibrium value of macroscopic variables (like temperature) amounts to computing a probability distribution. Such a probability law depends on a finite number of parameters, and describes the possible states of the system composed of a huge amount of particles. This introductory course will take a mathematical look at the first principles of statistical mechanics, and will aim at:

- defining the notion of a thermodynamic system of particles; understanding the natural probability distributions that characterize such systems, namely the micro-canonical ensemble and the Boltzmann-Gibbs distribution;
- proving (at least in simpler cases) the theorem of equivalence of ensembles, which justifies the use of the Boltzmann-Gibbs distribution to compute thermodynamics quantities;
- asking some mathematical questions (some of them still unsolved) that naturally arise in this area.

1 Introduction

Classical physics mainly relies on the following assumption: if the state of the system is precisely known in one instance, the laws of physics determine its future states completely. Therefore, why and how to introduce probabilities into classical physics? Here are listed a few examples of physical statements whose precise formulation requires probabilistic concepts:

1. Water freezers at 0 degree Celsius.
2. When two bodies are in contact with one another, the heat flows from the warmer body to the colder one.
3. When the heat of a container full of gas increases, the pressure on the walls of the contained increases too.

The formulation of these statements should involve probability, for the following reason: one can indeed construct special scenarios that are consistent with the laws of physics where each of the above statements is false. These scenarios are physically possible but extremely unlikely!

The most famous example of such a physically possible but highly improbable scenario was given by Maxwell: let A be an insulated contained that is divided into two boxes, B and C. Both boxes are filled with gas, and separated by a sliding door. Assume that the temperature of B is higher than the temperature of C. When the door is open, molecules may flow from one compartment to the other; when it is closed, B and C are completely insulated. Assume now that the door is intelligent enough to select the most improbable situations:

- when the door sees a very "fast" molecule (with high temperature) approaching from C toward the door, it quickly opens and lets the molecule pass from C to B;
- when the door sees a very "slow" molecule (with low temperature) approaching from B toward the door, it quickly opens and lets the molecule pass from B to C.

This will cause a heat transfer from the colder body to the warmer one! Now, take a second container that is an exact replica of the first. This time, the door cannot influence the passage of molecules any more, but instead there is a random device that opens the door at irregular intervals. It may happen (very unlikely) that this device opens the door exactly at the same times as before! And the same conclusion follows: a heat flow is created from C to B. This second case is consistent with the laws of physics.

1.1 Thermodynamics

The purpose of *thermodynamics* is to describe the properties of various macroscopic systems at (or near) equilibrium. This is done with the help of a few *macroscopic variables* such as the internal energy E, the volume V, the pressure P, the number of particles N, the temperature T, the entropy S, and others. These variables are not all independent, and thermodynamics provides general relations among them. Some of these relations are not specific to particular systems, but are valid in a very general setting.

However, because thermodynamics makes no link with the microscopic physics that is involved at the level of particles, it cannot give a complete picture. This is the role of *statistical mechanics*. In fact, thermodynamics could stand very well on its own. Statistical mechanics aims at extending and complementing its framework. We recall here the most famous laws of thermodynamics:

1. Thermodynamics variables can be divided into two groups, depending on their behaviour under a scaling of the system: *extensive* variables scale with the system size (like E, S, V, and N), while *intensive* variables stay invariant (like T, P).
2. An isolated system is in *thermal* equilibrium if T is uniform throughout the system; it is in *mechanical* equilibrium if P is uniform.
3. The *first law of thermodynamics* states that the internal energy E of a system can be increased by either adding heat or doing work. Mathematically,

$$\mathrm{dE} = \bar{d}\mathrm{Q} + \bar{d}\mathrm{W},$$

where $\bar{d}\mathrm{Q}$ and $\bar{d}\mathrm{W}$ are *inexact* differentials, meaning that the integrals

$$\int_{\mathrm{A}}^{\mathrm{B}} \bar{d}\mathrm{Q} \quad \text{and} \quad \int_{\mathrm{A}}^{\mathrm{B}} \bar{d}\mathrm{W}$$

depend not only on the end points, but also on the path taken to go from state A to state B. However, we have

$$\int_{\mathrm{A}}^{\mathrm{B}} \mathrm{dE} = \mathrm{E(B)} - \mathrm{E(A)},$$

independently of the path.

4. The *second law of thermodynamics* states that there exists a function S called the entropy, such that, for *thermally isolated systems* (such that $\bar{d}\mathrm{Q} = 0$), it satisfies $\mathrm{dS} \geqslant 0$.

1.2 A few notions of probability theory

The concept of probability is one of the foundations of general statistics. The likelihood of a particular outcome among the set of possible outcomes is expressed by a number from 0 to 1 (0 representing the impossibility and 1 representing the absolute certainty). Dealing with basic probability as a discrete counting process is satisfactory if we work with reasonably small numbers, like throwing dice or picking cards: in these cases the probability for a given event can be thought of as the ratio "the number of ways that event can happen" divided by "the number of ways that any possible outcome could happen". If the number of events is very large, as in the distribution of energy among the molecules of a gas, then the probability has to be approximated by a continuous variable.

In the discrete case we denote by $\mathbb{P}(x_i)$ the probability that x_i is observed. The *normalization* constraint reads as

$$\sum_i \mathbb{P}(x_i) = 1,$$

and the mean value of x is

$$\langle x \rangle = \sum_i x_i \mathbb{P}(x_i).$$

If the outcome x takes a continuous range of real values, then the probability $\mathbb{P}(x)$ takes a different character, and the sum above takes the form of an integral: the probability of finding outcomes x between a and b can be expressed as

$$\int_a^b f(x)\mathrm{d}x = \int_a^b \frac{d\mathbb{P}}{dx}\mathrm{d}x.$$

The function f is called the *probability density* of the random variable. The normalization condition is then

$$\int_{\mathbb{R}} f(x)\mathrm{d}x = 1,$$

and the mean value equals

$$\langle x \rangle = \int_{\mathbb{R}} x f(x)\mathrm{d}x.$$

In other words, if $\mathrm{d}x$ is an infinitely small number, the probability that x belongs to $(x_0, x_0 + \mathrm{d}x)$ is equal to $f(x_0)\mathrm{d}x$:

$$\mathbb{P}[x \in (x_0, x_0 + \mathrm{d}x)] = f(x_0)\mathrm{d}x.$$

EXAMPLE 1.1 (Probability distributions). We give here the main probability densities that are used in statistical mechanics.

1. Taking a number x *uniformly* at random in the interval (a, b) means that its probability density is

$$\mathrm{U}(x) = \begin{cases} 1/(b-a) & \text{if } x \in [a, b], \\ 0 & \text{otherwise.} \end{cases}$$

2. The *normal distribution* is widely used in various domains, and its density is defined on $\mathbb{R}$ as

$$\mathrm{N}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(\frac{(x-\mu)^2}{2\sigma^2}\Big), \qquad \mu \in \mathbb{R}, \sigma > 0. \tag{1}$$

The normal distribution is remarkably useful because of the *Central Limit Theorem*[1]. Physical quantities that are expected to be the sum of many independent processes (such as measurement errors) often have distributions that are nearly normal. When $\mu = 0$ and $\sigma = 1$, N is called the *standard normal distribution*.

3. Perhaps the most widely used distribution function in classical physics is the *Boltzmann-Gibbs distribution* function, which describes the probability of finding particles with an amount of energy $E \geqslant 0$ at a given temperature T:

$$f(E) = \frac{1}{Z} e^{-E/(k_B T)}, \tag{2}$$

where k_B is the Boltzmann constant, and Z is the normalization constant which has to satisfy:

$$\frac{1}{Z} \int_0^{+\infty} e^{-E/(k_B T)} dE = 1.$$

In other words, the probability that any one molecule will be found with energy E decreases exponentially with the energy.

1.3 Foundations of statistical mechanics

In what follows, a *thermodynamic system* should be thought as a gas enclosed in a container of a given (but possibly variable) volume V. This container provides the mean to couple the system to an external mechanical system. A general definition will be give in Section 3.

Definition 1.1. 1. *The* macro-state *of a thermodynamic system depends on a relatively small number of thermodynamic coordinates (for instance: pressure and temperature).*

2. *A* micro-state *corresponds to a microscopic description (of the degrees of freedom that compose the system, for instance: the position and velocity of each molecule of gas). The description of each micro-state requires an enormous amount of information, and the corresponding time evolution is usually quite complicated.*

3. *Rather than following the evolution of an individual (pure) micro-state,* statistical mechanics *examines an ensemble of micro-states corresponding to a given (mixed) macro-state.*

In other words, macro-states can be characterized by a probability distribution of possible states across a certain statistical ensemble of all micro-states. Each thermodynamic state, say (V, T, N) for instance, corresponds precisely to one probability distribution $f_{(V,T,N)}$ on the micro-states space.

Example 1.2 (Tossed coin). A tossed coin can land in two positions: head up, or tail up. Considering the coin as a particle, one can say that this particle has two states: 0 or 1. If N coins are tossed, this can be considered as a system of N particles with two states each. The micro-states of the system are specified by the states occupied by each point: they are elements of $\{0, 1\}^N$, there are 2^N possible micro-states.

The macro-states of this system are defined by the number of particles in each state, N_0 and N_1, which have to satisfy the condition $N_0 + N_1 = N$. The number of micro-states ω associated to one macro-state (N_0, N_1) is given by

$$\omega = \binom{N}{N_0} = \frac{N!}{N_0!(N - N_0)!}.$$

The main postulate of statistical physics is that the *equilibrium macro-state* is the one realized by the greatest number of micro-states, i.e. the most probable macro-state. For instance, a macro-state of an ideal gas with all molecules in one half of the container is much less probable than the macro-state with

---

[1] Roughly speaking, it states that large averages of independent random variables are normally distributed.

the molecules equally distributed over the whole container: this exactly corresponds to $N_0 = N_1 = N/2$ in the example above.

More precisely, if the initial state is all molecules in one half of the container, then in the course of evolution the system will come to the most probably state with the molecules equally distributed over the whole container, and will stay in this state forever.

One of the law of thermodynamics states that: if an isolated system is initially in a non-equilibrium state, then it will evolve to the equilibrium state characterized by the *maximal entropy*. More precisely, the entropy S and the number ω of micro-states associated to a macro-state should be related, one being a monotonic function of the other. If one chooses[2]

$$S = k_B \log \omega,$$

then the entropy has the advantage to be additive: if a system consists of two subsystems that weakly interact with each other then $S = S_1 + S_2$ while $\omega = \omega_1 \omega_2$.

1.4 Maximisation of entropy and Boltzmann distribution

We consider now a system of N particles with p states each (as a generalization of Example 1.2). Assume that, if a particle is in the state $i \in \{1, \dots, p\}$, then it has an energy $\varepsilon_i > 0$. Let $(N_1, \dots, N_p)$ be a macro-state, namely: N_i is the number of particles in state i. The number of possible micro-states is given by

$$\omega = \frac{N!}{N_1! \cdots N_p!}.$$

Suppose that there is no energy exchange with the system: in particular, not only the total number of particles N is fixed, but also the total energy E of the system.

We will define the notion of *temperature* by looking for the *equilibrium macro-state*. In that case, the equilibrium is achieved for the macro-state that contains the maximum number of micro-states. Therefore, the task is to find the maximum of ω[3] with respect to all N_i that satisfy

$$\sum_{i=1}^{p} N_i = N, \qquad \sum_{i=1}^{p} N_i \varepsilon_i = NE, \tag{3}$$

where ε_i is the energy of the particle in state i. Practically it is more convenient to maximise $\log(\omega)$ than ω itself. Consequently our problem reads

$$\max_{N_1,\dots,N_p} \left\{ \log\left(\frac{N!}{N_1! \cdots N_p!}\right) ; \text{ such that } \sum_{i=1}^{p} N_i = N, \ \sum_{i=1}^{p} N_i \varepsilon_i = NE \right\}.$$

Using the method of Lagrange multipliers (see Appendix B), one searches for the maximum of the target function

$$\Phi(N_1, \dots, N_p) = \log(\omega) + \alpha \sum_{i=1}^{p} N_i + \beta \sum_{i=1}^{p} N_i \varepsilon_i.$$

We are interesting in the behaviour of the macroscopic system with $N \to \infty$ (and therefore $N_1, \dots, N_p$ are very large). Using Stirling's formula[4], we only keep the first order in the asymptotic expansion of Φ

---

[2] This is an idea from L.Boltzmann; k_B is the Boltzmann constant.

[3] The letter ω comes from the German word *Wahrscheinlichkeit* which means probability.

[4] Stirling's formula states that

$$N! \sim \left(\frac{N}{e}\right)^N \sqrt{2\pi N}, \qquad \text{as } N \to \infty.$$

which reads as

$$\Phi(\mathrm{N}_1,\ldots,\mathrm{N}_p) \simeq \log \mathrm{N}! - \sum_{i=1}^{p} \mathrm{N}_i \log(\mathrm{N}_i) + (\alpha+1)\sum_{i=1}^{p} \mathrm{N}_i + \beta \sum_{i=1}^{p} \mathrm{N}_i \varepsilon_i.$$

Computing the derivatives with respect to N_i and solving its zeros, it yields

$$\mathrm{N}_i = e^{\alpha+\beta\varepsilon_i}.$$

The Lagrange multipliers can be found from (3). Finally, the probability that one particle is found in ith state is

$$p_i = \frac{\mathrm{N}_i}{\mathrm{N}} = \frac{1}{\mathrm{Z}(\beta)} e^{-\beta\varepsilon_i} \qquad \text{where } \mathrm{Z}(\beta) = \sum_{i=1}^{p} e^{-\beta\varepsilon_i},$$

which is the discrete version of the Boltzmann-Gibbs distribution (2), and β is implicitly given by

$$\frac{\mathrm{d}(\log \mathrm{Z})}{\mathrm{d}\beta}(\beta) = -\mathrm{E}. \tag{4}$$

Notice that p_i is actually the *density* of particles which are in ith state. The macro-state of the infinite system is defined as $(p_1,\ldots,p_p)$. We also easily check the relations

$$\sum_{i=1}^{p} p_i = 1, \qquad \sum_{i=1}^{p} p_i\varepsilon_i = \mathrm{E}.$$

In the following we denote by ω_{eq} the number of possible micro-states associated to $\mathrm{N}_i = \mathrm{N}e^{-\beta\varepsilon_i}/\mathrm{Z}(\beta)$.

Definition 1.2. *The parameter* $\mathrm{T} = (k_{\mathrm{B}}\beta)^{-1}$*, where* β *is solution to* (4)*, is called* temperature *of the thermodynamical system. In particular, the temperature is only defined* at equilibrium *and* in the thermodynamic limit $\mathrm{N} \to \infty$.

Definition 1.3. *The quantity*

$$\mathrm{S}(p_1,\ldots,p_p) := -k_{\mathrm{B}} \sum_{i=1}^{p} p_i \log(p_i)$$

is called thermodynamic entropy[5].

Proposition 1.1. *The thermodynamic entropy can be written as*

$$\mathrm{S}(p_1,\ldots,p_p) = \mathrm{S}(\mathrm{E},\beta) = k_{\mathrm{B}}\big(\beta\mathrm{E} + \log \mathrm{Z}(\beta)\big).$$

We also have

$$\lim_{\mathrm{N}\to\infty} \frac{k_{\mathrm{B}}}{\mathrm{N}} \log(\omega_{eq}) = \mathrm{S}(\mathrm{E},\beta).$$

In particular, the equilibrium macro-state is the one that maximises the *finite dimensional entropy* $\mathrm{S}_{\mathrm{N}}(\mathrm{N}_1,\ldots,\mathrm{N}_p) := k_{\mathrm{B}} \log(\omega)$ (among all ω that satisfy the conditions (3)), and that is what Boltzmann called *entropy*. This formula is now written on his gravestone. Indeed, when N is very large, one can use Stirling's formula to write

$$\frac{\mathrm{S}_{\mathrm{N}}}{k_{\mathrm{B}}} \underset{\mathrm{N}\to\infty}{\sim} \mathrm{N}\log(\mathrm{N}) - \alpha\mathrm{N} + \beta\mathrm{E} = \log \mathrm{Z}(\beta) + \beta\mathrm{E}.$$

Finally, one can easily compute

$$\frac{\partial \mathrm{S}}{\partial \mathrm{E}}(\mathrm{E},\beta) = \frac{1}{\mathrm{T}}.$$

This is a standard formula of thermodynamics.

---

[5] That entropy takes the same form than the *Shannon entropy* which is used in information theory.

2 The ideal gas in one dimension

2.1 The micro-canonical distribution

We give another simple (and highly unrealistic) example. Let us consider N particles, all of mass 1, that move on a one-dimensional interval $[0, \mathrm{L}]$ and do not interact with each other. For each particle i, we denote by $p_i \in \mathbb{R}$ its momentum and by $q_i \in [0, \mathrm{L}]$ its position. When reaching the boundary of the interval, particles are reflected.

Assume the right boundary of the interval (the *piston*) to be movable: a constant force f is acting, and the energy due to this external force is given by $\mathrm{E}_{\mathrm{ext}} = f\,\mathrm{L}$, if the piston's position is L. At initial time, the interior energy of the system of particles is given by

$$\sum_{i=1}^{\mathrm{N}} \frac{p_i^2(0)}{2} = \mathrm{E}_{\mathrm{int}}.$$

When time evolves, we assume that the dynamics of the system is such that it runs uniformly over all configurations that are compatible with the constraint that the total energy is kept constant, equal to

$$\mathrm{E}(t) = \mathrm{H} = \mathrm{E}_{\mathrm{int}} + f\,\mathrm{L}.$$

Since $\mathrm{E}_{\mathrm{int}} \geqslant 0$, the maximal value for the position of the piston is $\mathrm{L}_{\max} = \mathrm{H}/f$. Therefore, the state space for the whole system $(\mathrm{L}, q_1, ..., q_{\mathrm{N}}, p_1, ...p_{\mathrm{N}})$ is

$$\Omega := [0, \mathrm{L}_{\max}] \times [0, \mathrm{L}]^{\mathrm{N}} \times \mathcal{S}^{\mathrm{N}}\Big(\big[2(\mathrm{H} - f\,\mathrm{L})\big]^{1/2}\Big).$$

Assumption 2.1. *We assume that the initial configuration of the process is chosen uniformly at random in the set Ω, and that when time evolves, the configuration remains uniformly distributed on this set.*

The probability distribution on Ω is called the *micro-canonical distribution* of the system (see Section 3 below for the general definition). Assume that L is known, in that case we can compute the normalization constant for the uniform probability on $[0, \mathrm{L}]^{\mathrm{N}} \times \mathcal{S}^{\mathrm{N}}(\sqrt{2(\mathrm{H} - f\,\mathrm{L})})$ as

$$\mathrm{Z}_{\mathrm{N}}(\mathrm{L}, \mathrm{H}, f) := \mathrm{L}^{\mathrm{N}} \frac{2\pi^{\mathrm{N}/2}}{\Gamma(\mathrm{N}/2)} \big(2(\mathrm{H} - f\,\mathrm{L})\big)^{(\mathrm{N}-1)/2}.$$

This is indeed an easy consequence of Exercise 3.

2.2 Finding the equilibrium position

Thermodynamics is concerned with macroscopic variables (or macro-states): in that case, we are looking for the distribution of the piston position L. Precisely, the probability that L belongs to a small interval $(\mathrm{L}_0, \mathrm{L}_0 + \mathrm{dL})$ equals

$$\begin{aligned} \mathbb{P}[\mathrm{L} \in (\mathrm{L}_0, \mathrm{L}_0 + \mathrm{dL})] &= \frac{\mathrm{Z}_{\mathrm{N}}(\mathrm{L}_0, \mathrm{H}, f) \times \mathrm{dL}}{\int_0^{\mathrm{L}_{\max}} \mathrm{Z}_{\mathrm{N}}(\mathrm{L}, \mathrm{H}, f)\mathrm{dL}} \\ &= \frac{\exp\big\{\mathrm{N}\log(\mathrm{L}_0) + (\mathrm{N}-1)\log(2(\mathrm{H} - f\,\mathrm{L}_0))/2)\big\}\mathrm{dL}}{\int_0^{\mathrm{L}_{\max}} \exp\big\{\mathrm{N}\log(\mathrm{L}) + (\mathrm{N}-1)\log(2(\mathrm{H} - f\,\mathrm{L}))/2)\big\}\mathrm{dL}} \end{aligned} \tag{5}$$

When N is very large[6] we can estimate the behaviour of the integral above thanks to Laplace's method (see Appendix A). For that purpose, let us look for the maximal value of the function

$$f(\mathrm{L}) = \log(\mathrm{L}) + \frac{1}{2}\log(2(\mathrm{H} - f\,\mathrm{L})).$$

---

[6] N should be thought as the number of atoms, which is of order 10^{23} (Avogadro number).

The maximum of f is achieved at

$$\mathrm{L}^\star = \frac{2}{3}\frac{\mathrm{H}}{f}.$$

Therefore, with very high probability (close to 1), the piston stays at its *equilibrium position* $\mathrm{L}^\star$.

2.3 Computing the pressure on the piston

The *pressure* on the piston is defined as the average force that the gas molecules exercise when they are reflected on it. Each time a molecule i is reflected, its momentum changes from p_i to $-p_i$. The time between two hits is $(2\mathrm{L}/p_i)$. Therefore, during a time interval $[0,t]$, the molecule i hits the piston around $tp_i/(2\mathrm{L})$ times. The average force exerted during a time interval $[0,t]$ equals:

$$p(t) = \frac{1}{t}\int_0^t \Bigg(\sum_{\substack{i\in\{1,\dots,\mathrm{N}\}\\ q_i(s)=\mathrm{L}^\star}} \frac{dp_i}{ds}\Bigg) \mathrm{d}s = \frac{1}{t} \sum_{\substack{i\in\{1,\dots,\mathrm{N}\}\\ q_i(s)=\mathrm{L}\\ s\in[0,t]}} (2p_i(s)) = \frac{1}{t}\sum_{i=1}^{\mathrm{N}} 2p_i \times \frac{tp_i}{2\mathrm{L}^\star} = \frac{2\mathrm{E}_{\mathrm{int}}}{\mathrm{L}^\star} = \frac{2\mathrm{H}-2f\mathrm{L}^\star}{\mathrm{L}^\star} = f.$$

In other words, in equilibrium, the thermodynamic pressure $p = \lim_{t\to\infty} p(t)$ is equal to the external force f acting on the piston.

2.4 Concepts of entropy and temperature

The pressure is the first *intensive* quantity that we met: it is independent of the system size (or the amount of material in the system). The other quantities (mass, internal energy) are classified as *extensive*. Let us now introduce two new intensive quantities: the *temperature* and the *entropy*. For that purpose, we define the rescaled variables

$$\ell = \frac{\mathrm{L}}{\mathrm{N}}, \qquad e = \frac{\mathrm{E}_{\mathrm{int}}}{\mathrm{N}}, \qquad h = \frac{\mathrm{H}}{\mathrm{N}}.$$

The probability distribution (5) rewrites as

$$\mathbb{P}[\ell \in (\ell_0, \ell_0 + \mathrm{d}\ell)] = \frac{\exp\{\mathrm{N}\log(\ell_0) + (\mathrm{N}-1)\log(2(h-f\ell_0))/2)\}\mathrm{d}\ell}{\int_0^{\ell_{\max}} \exp\{\mathrm{N}\log(\ell) + (\mathrm{N}-1)\log(2(h-f\ell))/2)\}\mathrm{d}\ell}.$$

Define

$$s(h,f,\ell) := \log(\ell) + \frac{1}{2}\log(2(h-f\ell)).$$

Then, up to negligible terms,

$$\mathbb{P}[\ell \in (\ell_0, \ell_0 + \mathrm{d}\ell)] = \frac{\exp\{\mathrm{N}s(h,f,\ell_0)\}\mathrm{d}\ell}{\int_0^{\ell_{\max}} \exp\{\mathrm{N}s(h,f,\ell)\}\mathrm{d}\ell}.$$

The functions s is called *entropy function*. When computed at the equilibrium value $\ell^\star$ it is called *entropy*. In our case it reads

$$s(h,f) := s(h,f,\ell^\star) = \frac{1}{2}\log\frac{2h}{3} + \log\frac{2h}{3f}.$$

The *temperature* is defined as the coefficient that relates the change of entropy to the change of internal energy e. Since, at equilibrium, (h,f) completely determine (ℓ,e), we can express s in the variables (ℓ,e) as

$$s(\ell,e) = \log(\ell\sqrt{2e}).$$

As a result, the temperature T can be defined as

$$\mathrm{T}^{-1} = \frac{\partial s}{\partial e}(\ell,e) = 2e.$$

2.5 Conclusion

In this simple example, we understand how some famous equations of thermodynamics arise. The equilibrium state of the system is governed by the external force f, and the intrinsic probability of the system to find itself in a state with a given value of the macroscopic parameter (ℓ). The properties of this probability distribution give rise to some effective force – the pressure $p = 2e/\ell$ – that has to be equilibrated against the macroscopic force.

We also understand that: in the constant volume ensemble (L fixed), the conserved energy is the kinetic energy $\mathrm{E}_{\mathrm{int}}$; while in the constant pressure ensemble ($\mathrm{E}_{\mathrm{int}}/\mathrm{L}$ fixed), the conserved energy is the sum of the mechanical and kinetic energy $\mathrm{E}_{\mathrm{int}} + p\mathrm{L}$.

3 Rigorous definitions of micro-canonical objects

With the two examples above in mind, we now define what we understand by a thermodynamic system and give precise definitions.

Definition 3.1 (Thermodynamic system). *A* thermodynamic system *involves:*

1. *a parameter* $\mathrm{N} \in \mathbb{N}$*, called the* particle number*,*
2. *a state space* X *and its product space* X^{N}*,*
3. *a Hamilton function* $\mathrm{H}_{\mathrm{N}} : \mathrm{X}^{\mathrm{N}} \to \mathbb{R}$*,*
4. *a few constraints depending on macroscopic parameters.*

Remark 3.1. In the context of a gas, the space X is the phase space of a single molecule, H_{N} is the interaction between molecules, and the constraint is the indicator function that the position of all molecules should be within the container of volume V. In the following, we will denote this constraint by $x \in \mathrm{V}$.

Definition 3.2 (Micro-canonical ensemble). *The* micro-canonical ensemble *of a thermodynamic system is the collection of all uniform probability distributions* $f_{\mathrm{E,V,N}}$ *on the sets*

$$\Omega_{\mathrm{E,V,N}} := \left\{x \in \mathrm{X}^{\mathrm{N}} \; ; \; \mathrm{H}_{\mathrm{N}}(x) = \mathrm{E} \text{ and } x \in \mathrm{V}\right\}.$$

The micro-canonical partition function *is the function*

$$\mathrm{Z}_{\mathrm{E,V,N}} := \frac{1}{\mathrm{N}!} \int \delta(\mathrm{E} - \mathrm{H}_{\mathrm{N}}(x)) \, \mathbb{1}_{x \in \mathrm{V}} \, \mathrm{d}x, \tag{6}$$

where δ denotes the Dirac delta-function[7] *on* $\mathbb{R}$ *and* $\mathbb{1}_{\mathrm{A}}$ *is the characteristic function*[8] *of the set* A.

---

[7] The delta function is defined as follows: for all smooth functions ϕ, and $a \in \mathbb{R}$.

$$\int_{\mathbb{R}} \phi(y)\delta(y-a)\,\mathrm{d}y = \phi(a).$$

It follows that, if ψ is a function on X^{N}, then

$$\int_{\mathrm{X}^{\mathrm{N}}} \int_{\mathbb{R}} \delta(y - \mathrm{H}_{\mathrm{N}}(x))\psi(x)\phi(y)\,\mathrm{d}y\mathrm{d}x = \int_{\mathrm{X}^{\mathrm{N}}} \psi(x)\phi(\mathrm{H}_{\mathrm{N}}(x))\,\mathrm{d}x.$$

[8] The characteristic function $\mathbb{1}_{\mathrm{A}}$ is defined as

$$\mathbb{1}_{\mathrm{A}}(x) = \begin{cases} 1 & \text{if } x \in \mathrm{A}, \\ 0 & \text{otherwise.} \end{cases}$$

The thermodynamic entropy *is defined as*

$$\mathrm{S}(\mathrm{E},\mathrm{V},\mathrm{N}) := \log(\mathrm{Z}_{\mathrm{E},\mathrm{V},\mathrm{N}}).$$

Notice that the factor $1/\mathrm{N}!$ *is introduced to take into account that particles are indistinguishable and ensures that* S *is proportional to* N *(and additive). Finally, the* micro-canonical probability density *of the micro-states* $x \in \mathrm{X}^{\mathrm{N}}$ *associated to the macro-state* $(\mathrm{E},\mathrm{V},\mathrm{N})$ *is*

$$f_{\mathrm{E},\mathrm{V},\mathrm{N}}(x) := \frac{1}{\mathrm{Z}_{\mathrm{E},\mathrm{V},\mathrm{N}}} \times \frac{1}{\mathrm{N}!}\delta(\mathrm{E}-\mathrm{H}_{\mathrm{N}}(x))\mathbb{1}_{x\in\mathrm{V}}.$$

In many physics textbooks, this last measure is replaced with the Lebesgue measure (thought as an equivalent of the *volume* in $\mathbb{R}^{\mathrm{N}}$) of the set $\{x \in \mathrm{X}^{\mathrm{N}} \; ; \; |\mathrm{H}_{\mathrm{N}}(x)-\mathrm{E}| \leqslant \varepsilon\}$, which gives the same practical results. From now on we assume that $\mathrm{S}(\mathrm{E},\mathrm{V},\mathrm{N})$ is a differentiable function and we set Boltzmann's constant k_{B} equal to 1.

Definition 3.3. *The* pressure p, temperature T *and* chemical potential μ *are defined as*

$$\frac{\partial \mathrm{S}}{\partial \mathrm{E}}(\mathrm{E},\mathrm{V},\mathrm{N}) = \frac{1}{\mathrm{T}} = \beta, \qquad \frac{\partial \mathrm{S}}{\partial \mathrm{V}}(\mathrm{E},\mathrm{V},\mathrm{N}) = \frac{p}{\mathrm{T}}, \qquad \frac{\partial \mathrm{S}}{\partial \mathrm{N}}(\mathrm{E},\mathrm{V},\mathrm{N}) = \frac{1}{\mu}.$$

In particular, if we want to couple the thermodynamic system to a mechanical source of energy of strength f, it amounts to assume that the quantity $\mathrm{H} = \mathrm{E}+f\mathrm{V}$ is conserved (meaning that $\mathrm{H}_{\mathrm{N}}(x) = \mathrm{H}-f\mathrm{V}$ becomes the new constraint, and E can vary). Then, the distribution of V is given by

$$\mathbb{P}[\mathrm{V} \in (\mathrm{V}_0, \mathrm{V}_0+\mathrm{dV})] = \frac{\exp\left(\mathrm{S}(\mathrm{V}_0, \mathrm{H}-f\mathrm{V}_0, \mathrm{N})\right) \times \mathrm{dV}}{\int \exp\left(\mathrm{S}(\mathrm{V}_0, \mathrm{H}-f\mathrm{V}_0, \mathrm{N})\right) \mathrm{dV}}.$$

From Laplace's method, the *equilibrium position* of V is determined by the maximal entropy, and in particular is the solution of the equation

$$\frac{\mathrm{dS}(\mathrm{V},\mathrm{H}-f\mathrm{V},\mathrm{N})}{\mathrm{dV}} = 0, \tag{7}$$

with H and f fixed.

4 The theorem of equivalence of ensembles

The main difficulty in the computations involving the micro-canonical ensemble comes from the constrained integral (6) over some space $\{\mathrm{H}_{\mathrm{N}}(x) = \mathrm{E}\}$ in very high dimension, which turns out to be a pretty difficult geometric problem. We are able to do it for the sphere (Exercise 3), but in general this is hopeless. That is the reason why it is more convenient to change ensembles: we are going to consider a system where the energy is no longer fixed, but allowed to vary, while the temperature is fixed.

Definition 4.1. *Let* $\beta > 0$ *be a parameter. The* canonical partition function *associated to* $(\beta,\mathrm{V},\mathrm{N})$ *is given by*

$$\zeta_{\beta,\mathrm{V},\mathrm{N}} = \int e^{-\beta\mathrm{E}}\, \mathrm{Z}_{\mathrm{E},\mathrm{V},\mathrm{N}}\, \mathrm{dE} = \int e^{-\beta\mathrm{E}+\mathrm{S}(\mathrm{E},\mathrm{V},\mathrm{N})}\, \mathrm{dE} = \frac{1}{\mathrm{N}!}\int e^{-\beta\mathrm{H}_{\mathrm{N}}(x)}\, \mathrm{d}x,$$

and is associated to the canonical probability density *(or* canonical ensemble *or* Gibbs measure*)*

$$\mathcal{G}_{\beta,\mathrm{V},\mathrm{N}}(x) = \frac{\int e^{-\beta\mathrm{E}}\, \mathrm{Z}_{\mathrm{E},\mathrm{V},\mathrm{N}}\, f_{\mathrm{E},\mathrm{V},\mathrm{N}}(x)\, \mathrm{dE}}{\zeta_{\beta,\mathrm{V},\mathrm{N}}} = \frac{e^{-\beta\mathrm{H}_{\mathrm{N}}(x)}}{\int e^{-\beta\mathrm{H}_{\mathrm{N}}(x)}\, \mathrm{d}x}.$$

The function F *defined as*

$$\mathrm{F}(\beta,\mathrm{E},\mathrm{V},\mathrm{N}) := \mathrm{E}-\beta^{-1}\mathrm{S}(\mathrm{E},\mathrm{V},\mathrm{N})$$

is called the free energy functional*.*

We are now ready to state the main result of that section:

THEOREM 4.1 (Equivalence of ensembles). *Assume that the micro-canonical entropy satisfies*

$$\lim_{N\to\infty} \frac{S(E,V,N)}{N} = s(e,v), \tag{8}$$

where

$$\frac{V}{N} \xrightarrow[N\to\infty]{} v \quad \textit{and} \quad \frac{E}{N} \xrightarrow[N\to\infty]{} e.$$

Assume that s is strictly concave, continuous and satisfies, for all $\beta \geqslant 0$,

$$\int_{\{e\,;\,s(e,v)-\beta e\leqslant a\}} \exp\big(N[s(e,v)-\beta e]\big)\,de \leqslant C\exp(Na). \tag{9}$$

Define the function $f(\beta,v)$ by

$$\beta f(\beta,v) = \min_{e\in\mathbb{R}} \big\{e\beta - s(e,v)\big\}. \tag{10}$$

Finally, assume that the convergence in (8) is such that, uniformly in e,

$$\frac{S(E,V,N)-\beta E}{N[s(e,v)-\beta e]} \xrightarrow[N\to\infty]{} 1. \tag{11}$$

Then, for any $\beta \geqslant 0$,

$$\lim_{N\to\infty} \frac{\log\zeta(\beta,V,N)}{\beta N} = -f(\beta,v). \tag{12}$$

REMARK 4.1. The function (βf) is called the *Legendre transform* of s. One can easily show that, if s is differentiable and strictly concave, then

$$f(\beta,v) = e^\star(\beta,v) - \beta^{-1} s(e^\star(\beta,v),v) \tag{13}$$

where $e^\star$ is the unique solution of the equation

$$\beta = \frac{\partial s(e,v)}{\partial e}.$$

Proof. Let us denote

$$D_\delta := \big\{e\in\mathbb{R}\,;\, s(e,v)-\beta e > s(e^\star,v)-\beta e^\star - \delta\big\},$$

and D_δ^c is the complement of D_δ. First of all, the assumption on uniform convergence (11) ensures that, for large N, we can replace the integrand in

$$\int e^{-\beta E+S(E,V,N)}\,dE \tag{14}$$

by its limit, since, for any $\varepsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that, for all $N \geqslant N_0$, for all $e\in\mathbb{R}$,

$$\Big|N^{-1}\big[S(E,V,N)-\beta E\big] - \big[s(e,v)-\beta e\big]\Big| \leqslant \varepsilon\Big|s(e,v)-\beta e\Big|.$$

More precisely, for such N, we make the change of variables $E = eN$ in (14) and we bound

$$\begin{aligned}\int_{D_\delta^c} e^{-\beta eN+S(eN,V,N)} de &\leqslant \int_{D_\delta^c} e^{N[s-\beta e]+[S-\beta E-N(s-\beta e)]} de \\ &\leqslant \int_{D_\delta^c} e^{N[(s-\beta e)+\varepsilon|s-\beta e|]} de \\ &\leqslant C\exp\big(N[s(e^\star,v)-\beta e^\star+\varepsilon|s(e^\star,v)-\beta e^\star|-\delta]\big).\end{aligned}$$

On the other hand, for N large enough, the integral on D_δ can be bounded from (9) as follows:

$$\int_{D_\delta} e^{-\beta e N + S(eN,V,N)} de \leqslant C \exp\big(N[s(e^\star, v) - \beta e^\star](1+\varepsilon)\big).$$

Moreover, since the function s has bounded derivatives, on a set of size N^{-1} the integrand cannot vary by more than a constant, for some $c > 0$, so that we have

$$\int_{D_\delta} e^{-\beta e N + S(eN,V,N)} de \geqslant c N^{-1} \exp\big(N[s(e^\star, v) - \beta e^\star](1+\varepsilon)\big).$$

Taking the logarithm, dividing by N, for any $\varepsilon > 0$ we obtain that

$$-\beta f(\beta, v)(1-\varepsilon) \leqslant \liminf_{N\to\infty} \frac{1}{N} \log\left(N \int e^{-\beta e N + S(eN,V,N)} de\right)$$

and

$$\limsup_{N\to\infty} \frac{1}{N} \log\left(N \int e^{-\beta e N + S(eN,V,N)} de\right) \leqslant -\beta f(\beta, v)(1+\varepsilon).$$

This implies (12). □

Conclusion. The theorem of equivalence of ensembles justifies the use of the canonical ensemble to compute thermodynamic quantities, from the canonical rather than the micro-canonical partition function. We can define the free energy in terms of the logarithm of the partition function and derive all thermodynamic quantities (including the entropy) from it via Legendre transformations.

It is important to note that this equivalence holds in the limit of infinite particle number (and, in consequence, infinite volume, energy, ...). Once again, this illustrates the fact that statistical mechanics in interested in understanding what happens when the size of the system tends to infinity.

The beauty of the equivalence of ensembles is that, computationnally, it is much easier (even though still hard enough) to work with the Gibbs distribution than with the micro-canonical distribution: working with constraints is always hard, and the canonical ensemble allows us to get rid of one annoying constraint. The theorem tells us that not fixing the energy is fine.

Statistical Mechanics: Exercices

EXERCISE 1 (Exact and inexact differentials). Consider the infinitesimal quantity $\mathrm{A}dx + \mathrm{B}dy$ where A and B are functions of x and y. It this quantity is to be an *exact* differential, then there must exist a function $\mathrm{F}(x, y)$ such that $\mathrm{dF} = \mathrm{A}dx + \mathrm{B}dy$.

1. Show that in this case, A and B must satisfy

$$\frac{\partial \mathrm{A}}{\partial y} = \frac{\partial \mathrm{B}}{\partial x}.$$

2. Is $(x^2 - y)\mathrm{d}x + x\mathrm{d}y$ an exact differential?
3. Is $(1 - y/x^2)\mathrm{d}x + (1/x)\mathrm{d}y$ an exact differential?
4. Prove that if $\mathrm{A}dx + \mathrm{B}dy$ is an exact differential, then its integral from (x_1, y_1) to (x_2, y_2) depends on the endpoints only, and not on the path taken to go from one point to the other.

EXERCISE 2 (Normal distribution). 1. Prove[9] that

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}.$$

2. Check that the function N (given in (1)) defines a probability distribution, in the sense that

$$\int_{\mathbb{R}} \mathrm{N}(x)\mathrm{d}x = 1,$$

and compute its mean value.

EXERCISE 3 (The surface of a sphere). Let $\mathrm{S}^{\mathrm{N}}(r)$ denote the $(\mathrm{N}-1)$–dimensional sphere of radius $r > 0$, i.e.

$$\mathrm{S}^{\mathrm{N}}(r) := \left\{ (x_1, ..., x_{\mathrm{N}}) \in \mathbb{R}^{\mathrm{N}} \; ; \; \sum_{i=1}^{\mathrm{N}} x_i^2 = r \right\}.$$

Let $\mathcal{A}_{\mathrm{N}}(r)$ be the surface area of $\mathrm{S}^{\mathrm{N}}(r)$.

1. Compute

$$\int_{\mathbb{R}^{\mathrm{N}}} e^{-(x_1^2 + \cdots + x_{\mathrm{N}}^2)} \, \mathrm{d}x_1 \cdots \mathrm{d}x_{\mathrm{N}}.$$

2. Deduce that

$$\mathcal{A}_{\mathrm{N}}(r) = \frac{2\pi^{\mathrm{N}/2}}{\Gamma(\mathrm{N}/2)} r^{\mathrm{N}-1}, \qquad \text{where } \Gamma(s) := \int_0^{+\infty} x^{s-1} e^{-x} \, \mathrm{d}x.$$

Remind (or prove) that the function Γ satisfies

$$\begin{aligned} \Gamma(x+1) &= x\Gamma(x) \quad (x \in \mathbb{R}), \\ \Gamma(1/2) &= \sqrt{\pi}, \\ \Gamma(n+1) &= n! \quad (n \in \mathbb{N}). \end{aligned}$$

---

[9] **Hint:** First of all, use the polar coordinates to compute

$$\int_{\mathbb{R}^2} e^{-x^2-y^2} dxdy.$$

As a result, the *uniform probability density on the sphere* $S^N(r)$ is the function that constantly equals the inverse of the partition function

$$Z_N(r) = \int_{S^N(r)} dx_1 \cdots dx_N = \frac{2\pi^{N/2}}{\Gamma(N/2)} r^{N-1}$$

on $S^N(r)$ and vanishes elsewhere.

EXERCISE 4 ($\star$). Now we consider a random vector $X \in \mathbb{R}^N$ that has random coordinates $(X_1, \ldots, X_N)$. We choose to generate each coordinate by using a standard normal distribution. We are generating each coordinate *independently* – i.e., the value of one coordinate has no effect on the value of any other coordinate. Show that the probability distribution of the random vector U given by

$$U = \frac{X}{\|X\|}, \qquad \text{where } \|X\|^2 = \sum_{i=1}^{N} X_i^2,$$

is invariant under any rotation in $\mathbb{R}^N$. This means that U has the uniform distribution on the surface of the sphere $S^N(r)$.

EXERCISE 5. Generalize Example 1.2 for a system of N particles that can have $p \in \mathbb{N}$ different states, denoted by $\{1, 2, \ldots, p\}$: more precisely,

- show that the number of possible micro-states equals p^N;
- compute the number of micro-states associated to one macro-state $(N_1, \ldots, N_p)$;
- what is the most probable macro-state?

EXERCISE 6. Write explicitely the list of elements given in Definition 3.1 for

- the placement of N particles in p boxes (Subsection 1.4 and Exercise 5);
- the ideal gas in one dimension (Section 2).

EXERCISE 7. Using Definition 3.3, show that

$$\frac{dS(V, H - fV, N)}{dV} = 0,$$

is equivalent to $p = f$.

EXERCISE 8 (Ideal lattice gas). Let Λ be a finite subset of the discrete lattice $\mathbb{Z}^d$. We denote by $V = |\Lambda|$ the number of vertices of this subset. We consider of fixed number N of particles, for which we are only interested in their positions $x_i \in \Lambda$. We assume that two particles can occupy the same vertex, and the total energy of the system, which is conserved, is proportional to the number of particles: $E = \mu N$. Therefore, the micro-canonical partition function is simply the number of ways to arrange N particles on V vertices of the lattice.

1. Compute $Z_{E,V,N}$ and show that $S(E, V, N)$ is equivalent as $N \to \infty$ to $N \log(1 + V/N)$.
2. Prove that the pressure is given by $p = E/(\mu V)$.
3. Consider the lattice gas with the additional constraint that no more than one particle can occupy the same site. What will be changed?

EXERCISE 9 (Micro-canonical entropy). Let ρ_0 be the function on $\Omega_{E,V,N}$ defined as

$$\rho_0(x) := \frac{1}{N!} \delta(E - H_N(x)) \mathbb{1}_{x \in V}.$$

Notice that ρ_0 is not a probability density since its integral does not equal 1. Let ρ be any probability density whose support is $\Omega_{E,V,N}$. The *relative entropy* of ρ with respect to ρ_0 is defined as

$$H(\rho,\rho_0) := \int \log\left(\frac{d\rho(x)}{d\rho_0(x)}\right) d\rho(x)$$

Show that the unique minimizer of this function is the uniform probability density on $\Omega_{E,V,N}$, which is denoted by $f_{E,V,N}$ in the lecture notes, and that

$$H(f_{E,V,N},\rho_0) = -S(E,V,N).$$

EXERCISE 10 (Legendre transform). Let $I \subset \mathbb{R}$ be an interval, and $f : I \to \mathbb{R}$ be a convex function. Then, its *Legendre transform* is the function $f^\star : I^\star \to \mathbb{R}$ defined by

$$f^\star(y) = \sup_{x\in I}\{xy - f(x)\}, \qquad y \in I^\star.$$

1. Show that $f^\star$ is well-defined on I if f is convex.
2. Note that if f is defined on $\mathbb{R}$ and everywhere differentiablee, then $f^\star(p)$ can be interpreted as the negative of the y-axis coordinate of the intersection between the tangent line to the graph of f that has slope p and the y-axis.
3. Prove that, if f is twice differentiable such that $f'' > 0$, then $f^\star(p) = px_0 - f(x_0)$ where x_0 is the solution to $f'(x_0) = p$.

EXERCISE 11 (Classical ideal gas in one-dimension). Here the Hamiltonian is

$$H_N(p_1,\dots,p_N,q_1,\dots,q_N) = \sum_{i=1}^{N} \frac{p_i^2}{2}.$$

1. Show that the canonical partition function is

$$\zeta_{\beta,V,N} = \frac{1}{N!}V^N\left(\frac{2\pi}{\beta}\right)^{N/2}.$$

2. Show that

$$f(\beta,v) \underset{N\to\infty}{\sim} -\beta^{-1}\log\left(v\sqrt{\frac{2\pi}{\beta}}\right).$$

3. Compute the thermodynamic entropy using $f(\beta,v)$.
4. Compute the entropy directly from the micro-canonical partition function and compare.

A Laplace's method

Laplace's method is a technique used to investigate the behaviour as N goes to infinity of integrals of the form

$$\int_a^b \exp(\mathrm{N}f(x))\mathrm{d}x,$$

where f is a twice-differentiable function: assume that the function f has a unique global maximum at x_0. Then, significant contributions to the integral of this function will come only from points x in a of x_0, which can then be estimated thanks to Taylor expansions.

Proposition A.1. *Assume that f is twice differentiable on $[a,b]$, and that $x_0 \in [a,b]$ is the unique point where f achieves its maximal value on $[a,b]$. Assume additionally that $f''(x_0) < 0$. Then,*

$$\int_a^b \exp(\mathrm{N}f(x))\mathrm{d}x \underset{\mathrm{N}\to\infty}{\sim} \exp(\mathrm{N}f(x_0)) \times \sqrt{\frac{2\pi}{-\mathrm{N}f''(x_0)}}$$

B Lagrange multipliers

Suppose we want to extremize the function $f(x,y)$ where the coordinates $(x,y) \in \mathbb{R}^2$ are subject to the constraint $g(x,y) = 0$. If f is to be an extremum, then

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y = 0.$$

If the diplacements were all independent, then we would conclude that all the partial derivatives must be zero: $\partial f/\partial x = \partial f/\partial y = 0$. But the displacements are not independent, more precisely $g = 0$ implies that

$$\mathrm{d}g = \frac{\partial g}{\partial x}\mathrm{d}x + \frac{\partial g}{\partial y}\mathrm{d}y = 0,$$

(since the displacements must lie on the surface $g = 0$). This means that $\mathrm{d}y$ can be expressed in terms of $\mathrm{d}x$. Consider the linear combination $\mathrm{d}f + \lambda\mathrm{d}g$, where $\lambda \in \mathbb{R}$ is called the *Lagrange multiplier*. This combination must obviously vanish. Suppose that we can choose λ such that

$$\frac{\partial f}{\partial y} + \lambda\frac{\partial g}{\partial y} = 0, \tag{15}$$

and this implies then

$$\frac{\partial f}{\partial x} + \lambda\frac{\partial g}{\partial x} = 0. \tag{16}$$

Maximisation has been achieved. The *method of Lagrange multipliers* consists in: solving (15) and (16) considering λ as a parameter (x, y are the unknowns); then, determining the value of λ which is consistent with $g(x,y) = 0$.

References

[1] A. Bovier, *Statistical mechanics of disordered systems*, Cambridge University Press (2006), ISBN 0-521-84991-3.

[2] D. Garanin, *Statistical Thermodynamics*, Lecture notes (Fall 2009).

[3] Y.M. Guttmann, *The concept of probability in statistical physics*, Cambridge University Press (1999), ISBN 0-521-62128-3.

[4] E. Poisson, *Statistical Physics II* (PHYS*4240), Lecture Notes (Fall 2000).