NOTES ON STATISTICAL MECHANICS

Leonardo Pacciani Mori

Life is short and truth works far
and lives long: let us speak the truth.

———————

A. Schopenhauer

Last update: 31st March 2019

Contents

About this document

In my initial intentions this document was meant to be a collection of *all* I have ever learnt about statistical mechanics as a student, written in the most unitary and self-consistent possible way. However, as I gradually wrote down the notes of the courses I was attending at the time I started working on this document I realised how much effort this goal would have really required. Therefore, I have partially "rescaled" my ambitions: for the moment I would like this document to be a partially re-elaborated transcription of the notes of the courses about statistical mechanics that I have attended during my Master's Degree.

My hope is that one day I will manage to fill all the gaps and include also the material relative to what I have studied during my Bachelor's Degree, which basically consist in some introductory kinetic theory and, most importantly, basic quantum statistical mechanics, which is the greatest lack of this document to date.

As I have said I've tried to be as much self-consistent as possible, since I have combined material relative to different courses; for example I have tried to use always the same notation and to highlight the connections between different topics as much clearly as I could, but there still may be some parts which can turn out to be not really complete. I will surely work on this document again in the future, so I hope I will fix them.

I have also decided to release this document together with its LATEX source code under the Creative Commons *Attribution-ShareAlike 4.0 International* license. In brief this means that this document can be freely modified and redistributed, provided that it is released again under the same license and that proper credit is given to the original author.

Everything is available at leonardo.pm/teaching, which is the website where I will upload the new versions of this document in the future.

The content of these notes has also been used to form the Statistical Mechanics course on WikiToLearn, which is a website that allows users to create collaborative textbooks, and I have decided to do so because the philosophy of this community perfectly reflects the reasons for which I have decided to write down this document. I strongly encourage anyone who would like to correct or improve the material contained in this work to do it directly on WikiToLearn.

I have tried to be as much ruthless as possible in finding and correcting errors and mistakes, and I apologize if some have survived.

I hope that the overall result will anyway be satisfactory.

Padua, July 2016
Leonardo Pacciani Mori

Part I

Introductory topics

Chapter 1

Review of thermodynamics

This chapter is a review of the basic concepts of thermodynamics, in particular those that we will use in the following chapters. Therefore it has by no means the intention to be a complete review. For a *really* complete and rigorous exposition of thermodynamics see [2].

Thermodynamics is the theory that describes the *thermodynamic properties* of *macroscopic systems* at *equilibrium*. It can be generally introduced in two ways, an empirical and an axiomatic one; we will follow the latter.

By *macroscopic systems* we mean systems with many degrees of freedom (like the atoms or molecules that constitute a gas or a fluid), but such that only a few of them are measurable and relevant in order to describe the bulk[1] behaviour of the system itself. These relevant degrees of freedom are the *thermodynamic properties* of the system; for example, in the case of a liquid or a gas, these variables can be the pressure P, the volume V or the temperature T, while in the case of magnetic systems we can consider the magnetic field $\vec{H}$, the magnetization $\vec{M}$ (which is the total magnetic dipole moment) and the temperature T, or finally in electrostatic systems the electrical field $\vec{E}$, the electric polarization $\vec{P}$ and the temperature T.
If these thermodynamic variables do not change over time, we say that the system is in *equilibrium*[2].

The aim of thermodynamics is to find relations between the thermodynamic variables of a system in a given state of equilibrium, so that the value of any other variable of interest can be

---

[1]It is in fact impossible to describe the *exact* behaviour of the system, due to the terribly large number of degrees of freedom (of the order of 10^{23} in the case of gases or fluids).

[2]This statement should rather be reformulated this way: "If these thermodynamic variables do not change sensibly during the time we take to observe the system, we say that the system is in equilibrium". In fact, the thermodynamic properties of a system can change if we observe it for shorter or longer periods of time. Consider this example: if some boiling water is poured into a tea cup, after some seconds it will reach an unchanging state of rest in which its thermodynamic properties (like volume and temperature) do not change sensibly within some seconds, and thus we can say that during this observation time the cup of water is at equilibrium. However, if we observe it for a longer period of time, the temperature will obviously decrease sensibly and so we cannot say that the system is still in equilibrium; after an hour, say, the temperature of the water will be the same of that of the room, and if the room temperature remains constant for several hours then also the temperature of the water will do so, and therefore if the observation time is within this range of several hours the cup can again be considered in equilibrium. However, if we observe the system for even longer periods, such as a couple of days, then the volume of the water will decrease until it evaporates completely and so within this observation time the system cannot be regarded as being in an equilibrium state. After a few days, then, the cup can be again considered in equilibrium, but strictly speaking this is not a real equilibrium because the molecules of the cup can evaporate, even if it could take years to measure a significant change in the system.

obtained by the initial ones (and generally one tries to chose experimentally accessible variables). In other words, if we have a system in a known state of equilibrium (namely we know all of its thermodynamic variables) and then change the configuration of the system (for example changing its shape, or giving it a certain amount of energy), in general it will reach another equilibrium state: the scope of thermodynamics is to determine the values of the thermodynamic variables in this new equilibrium.

More in general we can say that the central problem of thermodynamics is to find the final state of equilibrium of a system from a given initial equilibrium state of several thermodynamic systems that can interact.

1.1 Thermodynamics of equilibrium

1.1.1 State variables

We will call *state variable* (or *function*) any quantity which at equilibrium depends only on the thermodynamic variables of a system, rather than on its history. In other words if $X_1, \ldots, X_N$ are the thermodynamic variables of a given system, any state variable A will be a function $A(X_1, \ldots, X_N)$ of them.

State variables can be either *extensive* or *intensive*:

- by **extensive** we mean that it is proportional to the size of the system. In other words, if a system is composed of N subsystems and X_α is the value of the thermodynamic variable X relative to the α-th subsystem, then the value $\overline{X}$ of the variable relative to the whole system is:

$$\overline{X} = \sum_{\alpha=1}^{N} X_\alpha$$

Considering them as functions, a state variable A will be extensive if it is a homogeneous function of degree one:

$$A(\lambda X_1, \ldots, \lambda X_N) = \lambda A(X_1, \ldots, X_N)$$

where λ is a positive real number.

Examples of extensive state variables are the internal energy[3], the mass or the volume of the system.

- by **intensive** we mean that it is independent of the size of the system, namely that the value of the variable relative to a subsystem is equal to that of the whole system. Intensive state variables are homogeneous functions of degree zero:

$$A(\lambda X_1, \ldots, \lambda X_N) = A(X_1, \ldots, X_N)$$

Examples of intensive variables are the pressure and the temperature of a system.

The fundamental hypothesis of thermodynamics is that any thermodynamic system can be characterized at equilibrium with a *finite* set of thermodynamic variables.

---

[3]Note that, strictly speaking, internal energy is *not* extensive, but it is so for large enough systems. In fact if we put two systems 1 and 2 in contact, the internal energy of the total system U_{TOT} will be the sum of the energies U_1 and U_2 of the two initial systems *and* the energy $U_{\text{s.i.}}$ due to the surface interaction between the two, assuming that the components of the two systems interact via a short-range potential. However, the internal energies of the two systems are proportional to their volumes, while the interaction energy is proportional to their interaction surface; since the surface grows much slower than the volume with respect to the size (unless the systems have *very* exotic and strange shapes, a case which of course we ignore), for large systems this interaction energy can be serenely ignored.

Finally, we also know that thermodynamic theory is based upon four laws (or postulates). For example, it is well known that the first law of thermodynamics is a restatement of the conservation of energy, namely a change of the internal energy dU of a system can only occur if there is an exchange of heat δQ and/or work δW, such that:

$$dU = \delta Q - \delta W$$

with the convention that $\delta Q > 0$ if the heat is absorbed by the system and $\delta W > 0$ if the work is done *by* the system; the difference in notation is due to the known fact that dU is an exact differential, while δQ and δW aren't.

1.1.2 Entropy

We now introduce the following postulate, that will allow us to solve the general problem of thermodynamics:

> *Given a thermodynamic system there exists a state function S called **entropy**, depending on the variables $(U, X_1, \ldots, X_N)$ and which is extensive, convex, monotonically increasing with respect to the internal energy U and such that the equilibrium states of the system are its maxima, compatibly with the constraints put on the system itself.*

The assumption that S is a monotonically increasing function of the internal energy U implies (Dini's theorem) that it is locally invertible; namely we can write at least locally U as a function of S and the other thermodynamic variables:

$$U = U(S, X_1, \ldots, X_N)$$

From now on we will consider systems characterized by the internal energy U, the volume V and the temperature T (also called *PVT systems*).
It can also be shown that S vanishes if and only if $\partial U / \partial S_{|V,N} = 0$, and since $U = U(S, V, N)$ we have:

$$dU = \frac{\partial U}{\partial S}_{|V,N} dS + \frac{\partial U}{\partial V}_{|S,N} dV + \frac{\partial U}{\partial N}_{|S,V} dN$$

These derivatives of U are intensive quantities[4], defined as:

$$T = \frac{\partial U}{\partial S}_{|V,N} \qquad P = -\frac{\partial U}{\partial V}_{|S,N} \qquad \mu = \frac{\partial U}{\partial N}_{|S,V}$$

where T is the temperature of the system, P its pressure and μ its *chemical potential* (note that from what we have previously stated, $S = 0$ if and only if $T = 0$).
We thus have:

$$dU = TdS - PdV + \mu dN$$

and the equations:

$$T = T(S, V, N) \qquad P = P(S, V, N) \qquad \mu = \mu(S, V, N)$$

are called *state equations*, which are the relations we were looking for that bound the thermodynamic variables of a system.

Let us note that once the entropy $S(U, V, N)$ of a system is known, all its thermodynamics can be straightforwardly derived: it is in fact sufficient to invert S and express U as $U = U(S, V, N)$ and then take some derivatives in order to obtain all the state equations of the system.

---

[4]Since they are derivatives of a homogeneous function of degree one, they are homogeneous functions of degree zero.

Let us also briefly see that, for example, T is indeed the temperature of a system, namely that if two systems with different temperatures are allowed to exchange internal energy they will finally reach an equilibrium where both have the same temperature. Let us call 1 and 2 these systems, which can be represented for example by two compartments of a box, each of volume V_i, containing N_i particles of gas (not necessarily of the same kind), at temperature T_i and with internal energy U_i (with $i = 1, 2$), separated by a wall that allows the exchange of heat (namely internal energy), but not particles (it's impermeable) or volume (it's fixed).

Once the systems are in thermal contact, they will reach a new equilibrium, which will be a maximum for the entropy, namely $dS = 0$ in the final state. However, since the whole system is isolated the total internal energy $U_1 + U_2$ must remain constant in the process. Thus (remembering that $dV = dN = 0$):

$$\begin{cases} dS = 0 \\ U_1 + U_2 = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} dS_1 + dS_2 = \frac{1}{T_1}dU_1 + \frac{1}{T_2}dU_2 = 0 \\ dU_1 = -dU_2 \end{cases} \quad \Rightarrow$$

$$\Rightarrow \quad dS = \left(\frac{1}{T_1} - \frac{1}{T_2}\right)dU_1 = 0$$

and since dU_1 is an arbitrary (non null) quantity, in the final state we have:

$$\frac{1}{T_1} - \frac{1}{T_2} = 0 \quad \Rightarrow \quad T_1 = T_2$$

and so the two systems will have the same temperature.

It can also be shown that the flow of heat goes, as intuition suggests, from the hotter to the colder system.

1.1.3   Thermodynamic potentials

Thermodynamic potentials are extremely useful tools, whose name derives from an analogy with mechanical potential energy: as we will later see, in certain circumstances the work obtainable from a macroscopic system is related to the change of an appropriately defined function, the thermodynamic potential. They are useful because they allow one to define quantities which are experimentally more easy to control and to rewrite the fundamental thermodynamic relations in terms of them.

Mathematically all the thermodynamic potentials are the result of a Legendre transformation of the internal energy, namely they are a rewriting of the internal energy so that a variable has been substituted with another.
For a PVT system $U = U(S, V, N)$; more in general we can write:

$$U = U(S, \{x_i\}, \{N_j\})$$

where x_i are called *generalized displacements* and N_j is the number of the j-th type of particle, and so we have:

$$dU = TdS + \sum_i X_i dx_i + \sum_j \mu_j dN_j$$

where X_i are called *generalized forces*.
The thermodynamic potentials that are generally used are:

 – **Helmholtz free energy**. It is defined by:

$$F = U - TS$$

Let's see that it is indeed a function of T instead of S:

$$dF = dU - d(TS) = dU - SdT - TdS = TdS - PdV + \mu dN - SdT - TdS \qquad \Rightarrow$$

$$\Rightarrow \qquad dF = -SdT - PdV + \mu dN \qquad \Rightarrow \qquad F = F(T, V, N)$$

Let us note that from this we have:

$$S = -\frac{\partial F}{\partial T}_{|V,N} \qquad\qquad P = -\frac{\partial F}{\partial V}_{|T,N} \qquad\qquad \mu = \frac{\partial F}{\partial N}_{|T,V}$$

We now show that F is in effect a thermodynamic potential.
For a general infinitesimal process we have:

$$dF = dU - d(TS) = \delta Q - \delta W - TdS - SdT$$

and thus:

$$\delta W = \delta Q - TdS - SdT - dF$$

For a reversible transformation $\delta Q - TdS = 0$, and if it is also isothermal then $dT = 0$; thus, for a reversible isothermal process $\delta W = -dF$: the quantity $-\Delta F$ is the amount of work that can be obtained from a reversible isothermal process, and hence the name "thermodynamic potential".
However, if the process is isothermal but irreversible then $\delta Q - TdS \leq 0$, so:

$$(\delta W)_{\text{irr,isot}} = \delta Q - TdS - dF \leq -dF$$

therefore in general $-\Delta F$ is the maximum work that can be extracted from a system at constant temperature.
We also see that if the process is spontaneous $\delta W = 0$ and so $dF \leq 0$: a spontaneous transformation can only decrease the Helmholtz free energy of a system, and thus we can conclude that equilibrium states of a system at fixed T, $\{x_i\}$, $\{N_j\}$ are the global minima of F.

– **Gibbs free energy**. It is defined by:

$$G = U - TS + PV = F + PV \qquad \Rightarrow \qquad G = G(T, P, N)$$

and as we have done for F it can be shown that G is the thermodynamic potential for transformations at fixed temperature and pressure.
Note that in general, if a system can be characterised by a set $\{x_i\}$ of generalized displacements (and $\{X_i\}$ are the corresponding generalized forces) and is composed of different types of particles, then:

$$G(T, \{X_i\}, \{N_j\}) = \sum_j \mu_j N_j \tag{1.1}$$

In order to show that, we must consider the *Euler identity* associated to the internal energy U. As a reminder, $f(x_1, \ldots, x_N)$ is a homogeneous function of degree ℓ if and only if we have:

$$\sum_{k=1}^{N} \frac{\partial f}{\partial x_k} x_k = \ell f(x)$$

Taking f as the internal energy $U(S, \{x_i\}, \{N_j\})$, since it is homogeneous of degree one we have:

$$\frac{\partial U}{\partial S}_{|x_i, N} S + \sum_i \frac{\partial U}{\partial x_i}_{|S,N} x_i + \sum_j \frac{\partial U}{\partial N_j}_{|S,x_i} N_j = U$$

Namely:

$$U = TS + \sum_i X_i x_i + \sum_j \mu_j N_j$$

Considering the general definition of the Gibbs free energy, we get precisely (1.1).

– **Entalpy**. It is defined as:

$$H = U + PV \qquad \Rightarrow \qquad H = H(S, P, N)$$

and it is the thermodynamic potential for isobaric transformations.

– **Grand potential**. It is defined as:

$$\Phi = U - TS - \mu N = F - \mu N \qquad \Rightarrow \qquad \Phi = \Phi(T, V, \mu)$$

and is useful for the description of open systems, namely systems that can exchange particles with their surroundings.

1.1.4   Gibbs-Duhem and Maxwell relations

Gibbs-Duhem relation

We start from the differential of the Gibbs free energy:

$$dG = \sum_i \frac{\partial G}{\partial x_i}_{|T,N} dx_i + \frac{\partial G}{\partial T}_{|P,N} dT + \sum_i \frac{\partial G}{\partial N_i}_{|P,T,N_{j \neq i}} dN_i = \sum_i X_i dx_i - SdT + \sum_j \mu_j dN_j$$

(where we are considering a system composed of different types of particles and that can be characterized with different generalized displacements). However, differentiating (1.1) we get:

$$dG = \sum_j N_j d\mu_j + \sum_j \mu_j dN_j$$

and equating with the previous expression of dG we obtain:

$$SdT - \sum_i X_i dx_i + \sum_j N_j d\mu_j = 0$$

which is called *Gibbs-Duhem relation*. We can see that as a consequence of this relation the intensive variables of the system are not all independent from each other: in particular, an r-component PVT system will have $r - 1$ independent intensive thermodynamic variables; for a simple PVT system the Gibbs-Duhem relation reduces to:

$$SdT - VdP + Nd\mu = 0$$

Maxwell relations

We have previously seen that for a PVT system we have:

$$T = \frac{\partial U}{\partial S}_{|V,N} \qquad\qquad -P = \frac{\partial U}{\partial V}_{|S,N}$$

we also know (Schwartz's theorem) that (since U is a sufficiently regular function) we must have:

$$\frac{\partial^2 U}{\partial V \partial S} = \frac{\partial^2 U}{\partial S \partial V}$$

which means:

$$\frac{\partial T}{\partial V}_{|S,N} = -\frac{\partial P}{\partial S}_{|V,N}$$

or similarly, deriving U with respect to S and N, or V and N:

$$\frac{\partial T}{\partial N}_{|S,V} = \frac{\partial \mu}{\partial S}_{|N,V} \qquad -\frac{\partial P}{\partial N}_{|S,V} = \frac{\partial \mu}{\partial V}_{|S,N}$$

These are called *Maxwell relations*.

If we consider F instead of U, we get:

$$\frac{\partial S}{\partial V}_{|T,N} = \frac{\partial P}{\partial T}_{|V,N} \qquad \frac{\partial S}{\partial N}_{|V,T} = \frac{\partial \mu}{\partial P}_{|T,N} \qquad \frac{\partial S}{\partial N}_{|P,T} = -\frac{\partial \mu}{\partial T}_{|P,N}$$

or, with G:

$$\frac{\partial S}{\partial P}_{|T,N} = -\frac{\partial V}{\partial T}_{|P,N} \qquad \frac{\partial S}{\partial N}_{|P,T} = -\frac{\partial \mu}{\partial T}_{|P,N} \qquad \frac{\partial V}{\partial N}_{|P,T} = \frac{\partial \mu}{\partial P}_{|T,N}$$

other similar relations can be found using other thermodynamic potentials or also for magnetic systems.

The usefulness of these relations will become clear in the next section.

1.1.5 Response functions

Response functions are quantities that express how a system reacts when some external parameters are changed.

Some important response functions for a PVT system are:

- The **isobaric thermal expansion coefficient**:

$$\alpha_P = \frac{1}{V}\frac{\partial V}{\partial T}_{|P,N}$$

- The **isothermal** and **adiabatic compressibilities**:

$$K_T = -\frac{1}{V}\frac{\partial V}{\partial P}_{|T,N} = -\frac{1}{V}\frac{\partial^2 G}{\partial P^2}_{|T,N}$$

$$K_S = -\frac{1}{V}\frac{\partial V}{\partial P}_{|S,N} = -\frac{1}{V}\frac{\partial^2 H}{\partial P^2}_{|S,N}$$

(the minus sign is needed to make them positive)

- The **specific heats at constant volume** and **pressure**:

$$C_V = \frac{\delta Q}{\partial T}_{|V,N} = T\frac{\partial S}{\partial T}_{|V,N} = -T\frac{\partial^2 F}{\partial T^2}_{|V,N}$$

$$C_P = \frac{\delta Q}{\partial T}_{|P,N} = T\frac{\partial S}{\partial T}_{|P,N} = -T\frac{\partial^2 G}{\partial T^2}_{|P,N}$$

- For a magnetic system, a very important response function is the **isothermal magnetic susceptibility**:

$$\chi_T = \frac{\partial M}{\partial H}_{|T} = -\frac{\partial^2 G}{\partial H^2}_{|T}$$

where of course H is the external field. To be more precise, this should be a tensor instead of a scalar:

$$\chi_{T\alpha\beta} = \frac{\partial M_\alpha}{\partial H_\beta}_{|T}$$

As we can see, response functions are related to the second derivatives of thermodynamic potentials.

With the response functions and Maxwell relations we can express many quantities otherwise hard to guess.
For example, suppose we want to know what

$$\frac{\partial S}{\partial V}_{|T,N}$$

is. Using Maxwell relations, we have:

$$\frac{\partial S}{\partial V}_{|T,N} = \frac{\partial P}{\partial T}_{|V,N}$$

and using the fact that:

$$\frac{\partial P}{\partial T}_{|V,N} \frac{\partial V}{\partial P}_{|T,N} \frac{\partial T}{\partial V}_{|P,N} = -1 \tag{1.2}$$

(which comes from mathematical analysis[5]), then:

$$\frac{\partial P}{\partial T}_{|V,N} = -\frac{1}{\frac{\partial V}{\partial P}_{|T,N} \frac{\partial T}{\partial V}_{|P,N}} = -\frac{\frac{\partial V}{\partial T}_{|P,N}}{\frac{\partial V}{\partial P}_{|T,N}} = \frac{\alpha_P}{K_T}$$

Now, response functions must obey some simple yet important inequalities, which come from the thermal or mechanical stability of the system.
For example, $C_P = \delta Q/\partial T_{|P} \geq 0$ and $C_V = \delta Q/\partial T_{|V} \geq 0$, since giving heat to a system will increase its temperature[6]. Similarly K_T, $K_S \geq 0$ since an increase in pressure always decreases the volume.

With Maxwell's relations we can also obtain some equations that will be useful in the future. Let's start considering a system with a fixed number of particles (namely $dN = 0$) and such that S is explicitly expressed in terms of T and V. Then:

$$dS = \frac{\partial S}{\partial T}_{|V} dT + \frac{\partial S}{\partial V}_{|T} dV$$

Dividing by dT both sides keeping the pressure constant, and then multiplying by T:

$$T\frac{\partial S}{\partial T}_{|P} - T\frac{\partial S}{\partial T}_{|V} = T\frac{\partial S}{\partial V}_{|T} \frac{\partial V}{\partial T}_{|P} \qquad \Rightarrow \qquad C_P - C_V = T\frac{\partial S}{\partial V}_{|T} \frac{\partial V}{\partial T}_{|P}$$

---

[5]In fact, let us call x_1, x_2 and x_3 three variables, and define:

$$x_1 = f_1(x_2, x_3) \qquad x_2 = f_2(x_1, x_3) \qquad x_3 = f_3(x_1, x_2)$$

Now, we have that $x_3 = f_3(x_1, f_2(x_1, x_3))$ and so deriving with respect to x_3 and x_1 we get:

$$1 = \frac{\partial f_3}{\partial x_2}\frac{\partial f_2}{\partial x_3} \quad \Rightarrow \quad \frac{\partial f_3}{\partial x_2} = \frac{1}{\frac{\partial f_2}{\partial x_3}} \qquad\qquad 0 = \frac{\partial f_3}{\partial x_1} + \frac{\partial f_3}{\partial x_2}\frac{\partial f_2}{\partial x_1} \quad \Rightarrow \quad \frac{\partial f_3}{\partial x_1} = -\frac{\partial f_3}{\partial x_2}\frac{\partial f_2}{\partial x_1}$$

Therefore we have indeed:

$$\frac{\partial f_3}{\partial x_1}\frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_3} = -1$$

This way we can find relations similar to equation (1.2), for example:

$$\frac{\partial S}{\partial V}_{|E,N} \frac{\partial V}{\partial E}_{|S,N} \frac{\partial E}{\partial S}_{|V,N} = -1 \qquad\qquad \frac{\partial S}{\partial N}_{|E,V} \frac{\partial N}{\partial E}_{|S,V} \frac{\partial E}{\partial S}_{|N,V} = -1$$

[6]There are a few cases where the opposite occurs, but are rather exotic. The most notable ones are black holes.

Now, using the Maxwell relation $\partial S/\partial V_{|T} = \partial P/\partial T_{|V}$ and (1.2), namely:

$$\frac{\partial P}{\partial T}_{|V} = -\frac{\partial P}{\partial V}_{|T}\frac{\partial V}{\partial T}_{|P}$$

we get:

$$C_P - C_V = -T\frac{\partial P}{\partial V}_{|T}\left(\frac{\partial V}{\partial T}_{|P}\right)^2 = \frac{TV}{K_T}\alpha_P^2$$

Similarly, for magnetic systems we have:

$$C_H - C_M = \frac{T}{\chi_T}\left(\frac{\partial M}{\partial T}\right)^2 \tag{1.3}$$

Since all these quantities are positive, we also see that:

$$C_P \geq C_V \qquad\qquad C_H \geq C_M$$

1.2 Thermodynamics of phase transitions

1.2.1 Phase transitions and phase diagrams

Experimentally, any element or compound can be found, depending on the thermodynamic conditions in which it is, in different *phases*. When we say that a system is in a particular *phase* we mean that its physical properties (like density or magnetization) are uniform.
We have also seen that the states of equilibrium for a thermodynamic system are minima of appropriate thermodynamic potentials.
Let us consider for example a PVT system and its Gibbs free energy $G(T, P)$; the molar Gibbs free energy will be:

$$g(T, P) = \frac{G(T, P)}{N}$$

and we can describe the thermodynamic properties of the system in (g, T, P) space.
Let us suppose for example that the system can be found in two phases α and β (for example liquid and solid), and call g_α and g_β their respective molar Gibbs potentials; then for given values of T and P the stable phase will be that with the lowest value of g: for example, if we have $g_\alpha(T, P) < g_\beta(T, P)$ then the system will be in phase α. Therefore there will be regions in (T, P) space were the most stable phase will be α and others in which it will be β.
A *phase diagram* is a representation of these regions in the appropriate parameter space (for the PVT system we are considering, (T, P) space).
If we now plot the values of g as a function of T and P in (g, T, P) space for every phase of the system, we can determine the regions where the two phases will be the stable ones, namely we can determine the phase diagram of the system, as shown in figure 1.1.
The very interesting region of this space (and the one on which we will focus our attention in this section) is the line where the surfaces of the two phases intersect: along this the two phases coexist, and when the system crosses it we say that it undergoes a *phase transition*.

Obviously, what we have stated can be generalized to systems which exhibit an arbitrary number of phases[7].

For a PVT system the phase diagram can be represented by a surface in PVT space described by a state equation like $f(P, V, T) = 0$ (with f a function depending on the particular system considered).

---

[7]There are however constraints on how many phases can coexist at a given temperature and pressure, see 1.2.3.
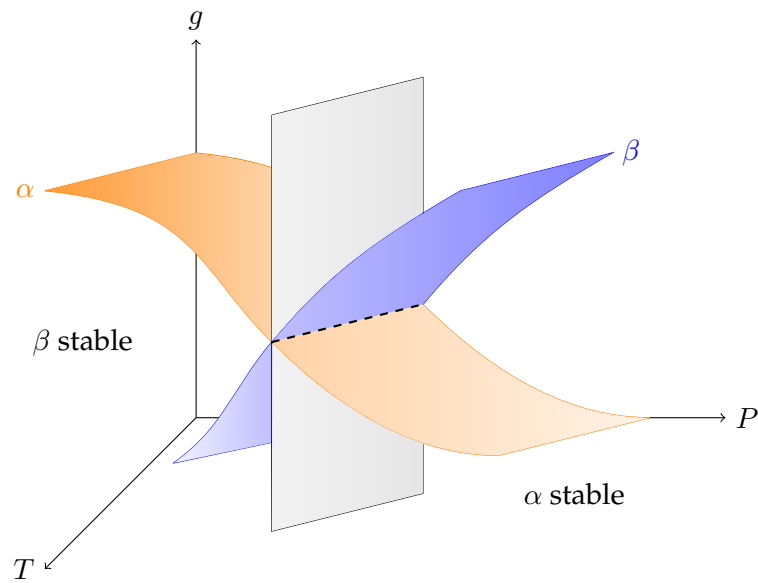
Figure 1.1: Stability of phases

(a) (T, P) projection
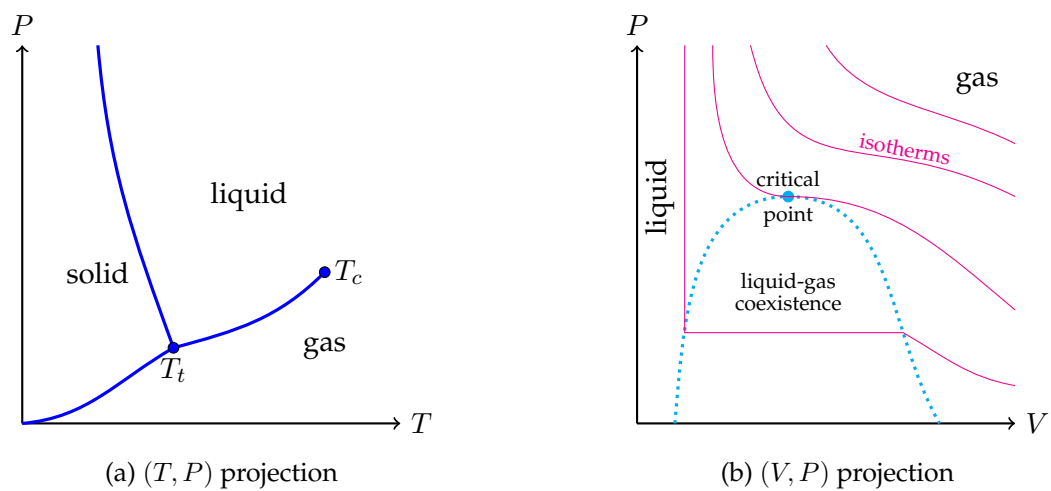
(b) (V, P) projection

Figure 1.2: Projections of phase diagram for PVT systems

(a) Phase diagram for a magnetic system in (T, H) space

(b) Magnetization as a function of the external field
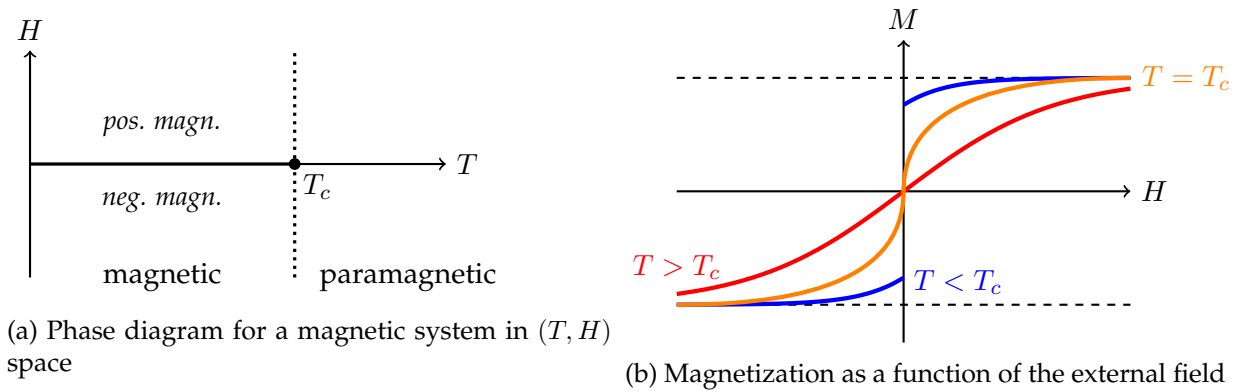
Figure 1.3: Magnetic systems seen in different parameter spaces

In this case it is more useful to consider the projection of this surface on (T, P) and (V, P) planes. Generally these projections look like those shown in figure 1.2a (there are however notable exceptions, like water for example). In particular, PVT systems are generally characterized by the existence of a *triple point* (see 1.2.3) and a *critical point* in the (T, P) phase diagram (in figure 1.2a these are determined, respectively, by the temperatures T_t and T_c).

The existence of a critical point has a very intriguing consequence: since the liquid-gas coexistence line ends in a point, this means that a liquid can continuously be transformed in a gas (or viceversa), and in such a way that the coexistence of liquid and gaseous phases is *never* encountered[8].

Obviously, our considerations are valid also for different thermodynamic systems.

For example, if we consider a magnetic system and its (T, H) phase diagram (where H is the external magnetic field[9]), then we find another critical point as shown in figure 1.3a.

In this case when $T > T_c$ the system is paramagnetic, namely it has no spontaneous magnetization when $H = 0$; when $T < T_c$ however, the system is magnetic and its magnetization can be positive (the total magnetic dipole moment $\vec{M}$ points upwards) or negative (it points downwards), and depends on the sign of H.

The same system in (H, M) space is represented in figure 1.3b for different values of the temperature; in particular, we can see that when $T < T_c$ the magnetization M has a jump discontinuity at $H = 0$.

1.2.2 Phase coexistence and general properties of phase transitions

In order to understand a little bit more about phase transitions and the coexistence of different phases, let us consider a PVT system in the liquid-gaseous transition.

The projection of the phase diagram in (V, P) space is shown in figure 1.2b, where we can also see clearly the presence of a critical point.

Let us now focus on the region of liquid-gas coexistence; physically, what happens (and what can also be understood from the figure) is that the system exhibits both a liquid and a gaseous

---

[8]It can be legitimately asked if such a point exists also for the solid-liquid transition. As far as we know this doesn't happen, and a reasonable explanation for this has been given by Landau: critical points can exist only between phases that differ quantitatively and not qualitatively. In the case of liquid and gas, in fact, the two phases have the same internal symmetry (both are invariant under *continuous* spatial translations) and differ only for the mean distance between the particles, while the solid and liquid phases have qualitatively different internal symmetries (solids are invariant only under *discrete* spatial translations).

[9]For simplicity, we are supposing that the real magnetic field $\vec{H}$, which is a vector, is directed along an axis, for example the vertical axis, so that we can consider only its magnitude H and thus treat it as a scalar.

phase, and it is possible to change the volume of the system without changing its pressure at fixed temperature. This means that the work we do on the system is used only to change the proportion between the phases (breaking or forming molecular bonds, on a microscopic level).

Let us now consider a slightly more general situation: suppose we have an isolated single-component thermodynamic system described by the internal energy U and extensive variables x_i (generalized displacements); if the system is subject to reversible processes then from the expression of dU we get:

$$dS = \frac{1}{T}dU - \sum_i \frac{X_i}{T}dx_i \tag{1.4}$$

where X_i are the generalized forces relative to the generalized displacements x_i.
Let us now suppose that two phases α and β coexist and that they can exchange U and x (namely, they are systems in contact with each other). At equilibrium the total entropy $S = S_\alpha + S_\beta$ must be maximized, and since the whole system is in equilibrium we also have $dU_\alpha + dU_\beta = 0$ and $dx_{i,\alpha} + dx_{i,\beta} = 0$. Substituting in (1.4) we get:

$$dS = \left(\frac{1}{T_\alpha} - \frac{1}{T_\beta}\right)dU_\alpha - \sum_i \left(\frac{X_{i,\alpha}}{T_\alpha} - \frac{X_{i,\beta}}{T_\beta}\right)dx_{i,\alpha} = 0$$

Therefore, since dU_α and $dx_{i,\alpha}$ are arbitrary positive quantities, we must have:

$$T_\alpha = T_\beta \qquad\qquad X_{i,\alpha} = X_{i,\beta}$$

Considering our PVT system again, since the generalized displacements that are needed in order to describe it are the volume V and the number of particles N, when two phases coexist we have:

$$T_\alpha = T_\beta \qquad\qquad P_\alpha = P_\beta \qquad\qquad \mu_\alpha = \mu_\beta$$

Since, as we have shown in 1.1.3, the Gibbs free energy is $G(T, P) = \mu N$ (for a single-component PVT system) then $g(T, P) = \mu$ and so when two phases coexist we also have:

$$g_\alpha(T, P) = g_\beta(T, P)$$

This equality must hold along the whole coexistence line in (T, P) space and so if we know a point in this space where the two phases coexist we can, at least locally, "reconstruct" the coexistence line. In fact we have $dg_\alpha = dg_\beta$, namely:

$$dg_\alpha = -s_\alpha dT + v_\alpha dP = dg_\beta = -s_\beta dT + v_\beta dP \qquad \Rightarrow \qquad \frac{dP}{dT}\Big|_{\text{coex.}} = \frac{s_\alpha - s_\beta}{v_\alpha - v_\beta} = \frac{L_{\alpha,\beta}}{T(v_\alpha - v_\beta)} \tag{1.5}$$

where $s = S/N$ and $v = V/N$, and by definition $L_{\alpha,\beta}$ is the molar latent heat needed to bring the system from phase β to phase α. This is known as *Clausius-Clapeyron equation*.

From the expression of $dg_\alpha = dg_\beta$ we can also understand some very general properties of phase transitions. In fact, from the first part of (1.5) at phase coexistence we have:

$$\frac{\partial g_\beta}{\partial T}\Big|_P - \frac{\partial g_\alpha}{\partial T}\Big|_P = \Delta s > 0 \qquad\qquad \frac{\partial g_\beta}{\partial P}\Big|_T - \frac{\partial g_\alpha}{\partial P}\Big|_T = \Delta v > 0$$

This means that when the system undergoes a phase transition its volume and its entropy have a jump discontinuity; as we will later see in more detail, since in this transition the first derivatives of a thermodynamic potential have a jump discontinuity, we call it a *first order transition*. A similar behaviour can be encountered in magnetic systems, where the magnetization M has a jump at $H = 0$ for temperatures lower that the critical one; in this case since $M = -\partial F/\partial H$

we see that the first derivative of the free energy F with respect to H has a jump discontinuity.

Furthermore, if we consider our system at the critical point we see from figure 1.2b that the isothermal compressibility:

$$K_T = -\frac{1}{V}\frac{\partial V}{\partial P}_{|T}$$

diverges when $V = V_c$.
Similarly, the magnetic susceptibility χ_T of a magnetic system at its critical temperature diverges when $H = 0$ and $M = 0$.
As we will see further on in 5.1.2, the divergence of response functions is a typical behaviour of thermodynamic systems in the neighbourhood of critical points and has important consequences.

1.2.3 Gibbs phase rule

Until now we have not considered any constraint on the number of phases that can coexist in a system at given values of T and P. However such constraints exist and are given by the *Gibbs phase rule*, which we now derive.

Let us consider a system composed of c different components and with φ coexisting phases (which we will label with roman numbers I, II etc.); from what we have previously stated T and P must be common to all phases, and all their chemical potentials must be equal. However, in order to specify the composition of the system we need only $c - 1$ variables, namely the ratios of the concentrations of the various components with respect to the concentration of a previously chosen component. Therefore the chemical potentials will depend on T, P and $c-1$ relative concentrations $(x^1, \ldots, x^{c-1})$, not necessarily common to all phases. We therefore have:

$$\mu_I^1(T, P, x_I^1, \ldots, x_I^{c-1}) = \mu_{II}^1(T, P, x_{II}^1, \ldots, x_{II}^{c-1}) = \cdots = \mu_\varphi^1(T, P, x_\varphi^1, \ldots, x_\varphi^{c-1})$$

$$\vdots$$

$$\mu_I^j(T, P, x_I^1, \ldots, x_I^{c-1}) = \mu_{II}^j(T, P, x_{II}^1, \ldots, x_{II}^{c-1}) = \cdots = \mu_\varphi^j(T, P, x_\varphi^1, \ldots, x_\varphi^{c-1})$$

$$\vdots$$

$$\mu_I^c(T, P, x_I^1, \ldots, x_I^{c-1}) = \mu_{II}^c(T, P, x_{II}^1, \ldots, x_{II}^{c-1}) = \cdots = \mu_\varphi^c(T, P, x_\varphi^1, \ldots, x_\varphi^{c-1})$$

We therefore have $2 + \varphi(c - 1)$ variables and $c(\varphi - 1)$ equations; if we want a solution to exist we must have at least as many variables as equations, namely $2 + \varphi(c - 1) \geq c(\varphi - 1)$. This means that:

$$\varphi \leq c + 2$$

which means that the maximum number of coexisting phases in a system composed of c different substances at fixed temperature and pressure is $c + 2$. In particular, we see that for a single-component PVT system we can have at most three coexisting phases for a particular choice of T and P: this is the *triple point* that we have already encountered.

Equivalently, we can say that in general the number of independent variables with which we can choose to describe the system at the coexistence of phases is:

$$f = 2 + \varphi(c - 1) - c(\varphi - 1) = c + 2 - \varphi$$

called *thermodynamic degrees of freedom*.
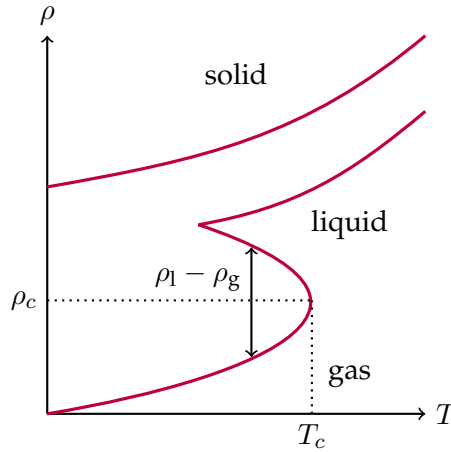
Figure 1.4: Projection of the phase diagram of a PVT system in (T, ρ)space

1.3   Order parameters

A very important feature of a thermodynamic system that, as we will see later on, is very important if we want to study its behaviour near a phase transition or a critical point is that of *order parameter*. By *order parameter* we mean a characteristic quantity of the system which is nonzero for $T < T_c$ and that vanishes when $T \geq T_c$. This quantity can be taken as a sort of "measure" of the order of the system: in fact when the temperature of a system is lower than its critical one the system will exhibit some kind of "order" that was not present at higher temperatures (due to the strong thermal fluctuations), and accordingly the value of the order parameter will keep track of that.

Let us see a couple of examples in order to make these statements more comprehensible.
Let us consider a fluid and the projection of its phase diagram in (T, ρ) space (which is of course equivalent to (T, V) space), which is represented in figure 1.4. As we can see the difference between the densities of the system in its liquid and gaseous phases $\rho_l - \rho_g$ vanishes when $T \geq T_c$ (because for those temperature there is actually no distinction between a liquid and a gas) and grows as the temperature drops under the critical temperature. Thus, this difference of densities can be taken as the order parameter of the system (the "order" consisting in the less chaotic motion that the particles have in the liquid phase).
Let us now consider, on the other hand, a magnetic system and the projection of its phase diagram in (M, T) space when $H = 0$, as shown in figure 1.5 (where the symmetry of the curve is due to the fact that, as we have seen at the end of 1.2.1, the magnetization of the system can point upwards or downwards). In this case, we see that the magnetization M itself satisfies the right requirements in order to be the order parameter of the system (and in this case the "order" is quantified in the amount of spins that point in the same direction).

A very important feature of an order parameter, as we shall see when we will be discussing the statistical mechanics of critical phenomena, is its nature: the properties of the system and its behaviour near a critical point will depend strongly on the fact that the order parameter of the system is either a scalar, a vector, a tensor etc.

Sometimes, order parameters can be distinguished in conserved ones (like in the case of a fluid, where the *average* density of the system remains constant during the transition) and not conserved ones (like the magnetization of a magnetic system).
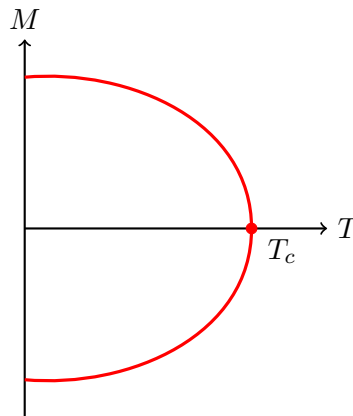
Figure 1.5: Phase diagram for a magnetic system in (T, M) space for $H = 0$

1.4 Classification of phase transitions

As we have partially seen, not all phase transitions are similar since they can exhibit different behaviours and characteristics.

There are in fact several ways that have been introduced in order to classify them:

– **Thermodynamic classification**: according to this classification, phase transitions can be divided in two groups:

  $\rightarrow$ Those which generate latent heat

  $\rightarrow$ Those which don't generate latent heat

– **Ehrenfest classification**: according to this classification, a phase transition is said to be an *n-th order transition* if the n-th derivative of any of the thermodynamic potentials of the system is discontinuous at the transition.

  For example, according to Ehrenfest's classification the solid-liquid or liquid-gaseous transitions for a fluid are all first order transitions. On the other hand, an example of a second order transition is the conductor-superconductor transition for a metal, since in this case there is a jump in the specific heat (which is the second derivative of the free energy, see 1.1.5).

However, Ehrenfest's classification is ultimately incorrect since at the time it was formulated it was still not known that in some transitions there are thermodynamic quantities that actually *diverge* instead of exhibiting a simple discontinuity.

– **Modern classification**: it is essentially a generalization of Ehrenfest's one; according to this classification a phase transition is said to be:

  $\rightarrow$ **first order** if there is a jump discontinuity in one (or more) of the first derivatives of an appropriate thermodynamic potential (which depends on the system considered)

  $\rightarrow$ **higher order**, **continuous** or **critical** if the first derivatives of the thermodynamic potentials are continuous, but the second derivatives are discontinuous or diverge

1.5 Critical exponents and universality

We know that when a system is in the neighbourhood of a critical point its thermodynamic quantities have peculiar behaviours (they change with the temperature, eventually diverging).

We can however ask ourselves *how* this quantities depend on the temperature, and for this we introduce the concept of *critical exponents*; as we will see their study will give rise to surprising and fundamental observations, and we will also understand why they are so important in the study of critical phenomena.

Since we are only interested in the proximity of the critical point we define the *reduced tempera-ture*:

$$t = \frac{T - T_c}{T_c}$$

so that when $t = 0$ the system is at its critical temperature.

If $F(t)$ is a generic function of t, the critical exponent λ associated to this function is defined as:

$$\lambda := \lim_{t \to 0} \frac{\ln |F(t)|}{\ln |t|}$$

In other words, we can write:

$$F(t) \overset{t \approx 0}{\sim} |t|^{\lambda}$$

Note, however, that we are unjustifiably supposing that the critical exponent λ is the same for both the limits $t \to 0^+$ and $t \to 0^-$; to be explicit, more in general we should have written:

$$F(t) \overset{t \approx 0}{\sim} |t|^{\lambda}(1 + at^{\lambda_1})$$

where a is a generic constant.

At this level, we are assuming that the two limits give the same result only for the sake of simplicity: however, in the framework of the scaling theory (see chapter 7) and of the Renor-malization Group (see chapter 8) this can be rigorously proved, so we are *a posteriori* justified to make this assumption.

Depending on the thermodynamic quantity that we are considering, the relative critical expo-nent has a different name. For example, considering a fluid we know that its specific heat at constant volume diverges at the critical point, and we call its relative critical exponent α so that:

$$C_V \overset{t \approx 0}{\sim} |t|^{-\alpha}$$

(where obviously $\alpha > 0$, as will always be in all the other cases).

On the other hand, the critical exponent associated to the order parameter of the system is called β, so again for a fluid we will have:

$$\rho_{\text{liq}} - \rho_{\text{gas}} \sim (-t)^{\beta}$$

(where the minus sign is due to the fact that the order parameter is non null only for $T < T_c$, and for example looking at figure 1.4 we could guess $\beta \approx 1/2$).

Of course all these considerations can be extended to other types of systems.

For completeness, in tables 1.1 and 1.2 we have written the most used and important critical exponents for fluid and magnetic systems.

Note that we have also included two quantities, the *correlation length* and the *correlation func-tion*[10], which haven't been previously introduced. We will encounter them later on (see 5.1.2), and we have inserted them here only for the sake of completeness.

---

[10]This is a function of $\vec{r}$, the distance between two points of the system; furthermore, d is the dimensionality of the system itself.

Thermodynamic quantity	*Critical exponent*
Specific heat at constant volume	$C_V \approx |t|^{-\alpha}$
Density difference (order parameter)	$\rho_{\text{liq}} - \rho_{\text{gas}} \approx (-t)^{\beta}$
Isothermal compressibility	$K_T \approx |t|^{-\gamma}$
Pressure at critical isotherm ($t = 0$)	$P - P_c \approx |\rho_{\text{liq}} - \rho_{\text{gas}}|^{\delta} \operatorname{sgn}(\rho_{\text{liq}} - \rho_{\text{gas}})$
Correlation length	$\xi \approx |t|^{-\nu}$
Correlation function	$G(\vec{r}) \approx 1/|\vec{r}|^{d-2+\eta}$

Table 1.1: Critical exponents for a fluid

Thermodynamic quantity	*Critical exponent*
Zero-field specific heat	$C_H \approx |t|^{-\alpha}$
Zero-field magnetization (order parameter)	$M \approx (-t)^{\beta}$
Zero-field isothermal susceptibility	$\chi_T \approx |t|^{-\gamma}$
Field at critical isotherm ($t = 0$)	$H \approx |M|^{\delta} \operatorname{sgn} M$
Correlation length	$\xi \approx |t|^{-\nu}$
Correlation function	$G(\vec{r}) \approx 1/|\vec{r}|^{d-2+\eta}$

Table 1.2: Critical exponent for a magnetic system

Now that we have defined critical exponents, we must justify their importance in the study of critical phenomena.

What makes them really interesting is that they appear to be *universal*: what we mean by this is that their values are *independent of the system chosen*, and depend only on some very general properties of the system; for example, if the components of the system interact via short-range potentials, it turns out[11] that critical exponents depend only on the dimensionality d of the system and on the symmetry of the order parameter.

In other words, critical exponents are *universal* characteristics of thermodynamic systems, while some other properties (like the value of the critical temperature) depend strongly on its microscopic details.

As an example, let us consider the results of a famous experiment done in 1945 by Guggenheim. In this experiment several different chemical compounds have been taken, and their liquid-gaseous phase coexistence curves have been measured; plotting the values of the reduced temperatures T/T_c and densities ρ/ρ_c of these coexistence curves for the different compounds, it turns out that *all* the data lie on the same curve in the proximity of the critical point, and surprisingly also far from it[12]; as an example of a similar case, see figure 1.6. This fact is also known as *law of corresponding states* (we will encounter it again in 6.4).

All these systems can therefore be described with the same critical exponent, which fitting the data results approximately equal to $1/3$. Note that the systems studied in this experiment are microscopically different from each other (there are monoatomic and diatomic gases, and even CH_4 which has a *much* more complex structure compared to them).

Now, the very interesting fact is that (compatibly with experimental errors) the *same* critical exponents appear in *completely* different systems: the measure of the β critical exponent for the magnetization of MnF_2 gives $\beta = 0.335(5)$ as a result[13], while the same exponent for the phase separation in a mixture of CCl_4 and C_7F_{16} gives[14] $\beta = 0.33(2)$.

This property of thermodynamic systems is called *universality*, and systems with the same set

---

[11] From the study of several exactly solvable models, or also from numerical simulations.

[12] E. A. Guggenheim, *The Principle of Corresponding States*, Journal of Chemical Physics, 13, 253 (1945).

[13] Heller and Benedek, *Physical Review Letters*, 13, 253, 1962.

[14] Thompson and Rice, *Journal of the American Chemical Society*, 86, 3547, 1964.
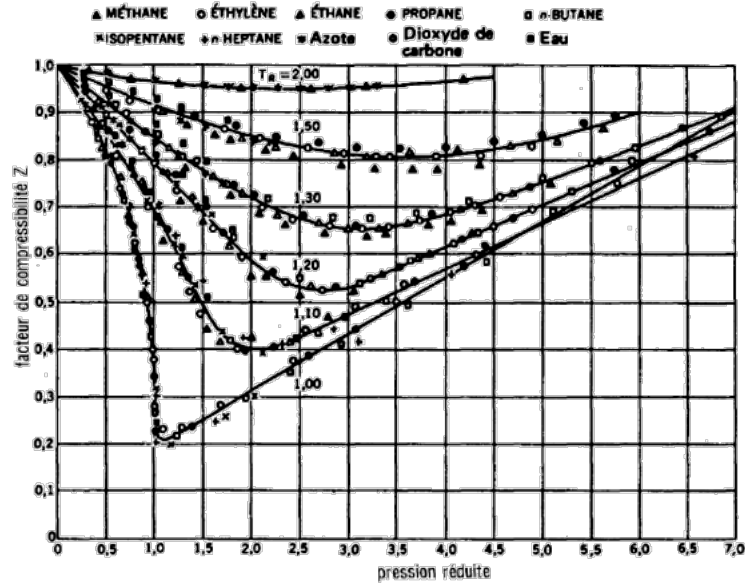
Figure 1.6: Law of corresponding states (source)

of critical exponents are said to belong to the same *universality class* (we will be able to actually explain this phenomenon only when we will be studying the Renormalization Group, see chapter 8).

This fact justifies the use of very simple and minimal models, with the right general properties, in order to describe the behaviour of thermodynamic systems in the neighbourhood of critical points, regardless of their microscopic details.

1.5.1   Inequalities between critical exponents

Critical exponents are not all completely independent: as we shall now show, from relations that we have previously found we can deduce some inequalities that involve critical exponents. The two main such relations are the *Rushbrooke* and *Griffiths* inequalities. We will limit ourselves to show explicitly only the derivation of the first one.

In 1.1.5 we have shown a relation, namely equation (1.3), for magnetic systems that involve the specific heats at constant field and magnetization:

$$C_H - C_M = \frac{T}{\chi_T} \left(\frac{\partial M}{\partial T} \right)^2$$

We thus have (since C_M is surely positive, because giving heat to a system at constant magnetization will *always* increase its temperature):

$$C_H = C_M + \frac{T}{\chi_T} \left(\frac{\partial M}{\partial T} \right)^2 \geq \frac{T}{\chi_T} \left(\frac{\partial M}{\partial T} \right)^2 \qquad \Rightarrow \qquad C_H \geq \frac{T}{\chi_T} \left(\frac{\partial M}{\partial T} \right)^2$$

If we now suppose $H = 0$ and T slightly lower than the critical temperature T_c, from the definitions of critical exponents we have:

$$C_H \sim |t|^{-\alpha} \qquad\qquad \chi_t \sim |t|^{-\gamma} \qquad\qquad \frac{\partial M}{\partial T} \sim (-t)^{\beta-1}$$

Thus substituting and remembering that $T < T_c$:

$$(T_c - T)^{-\alpha} \geq A \cdot T (T_c - T)^{\gamma - 2(\beta-1)} \qquad \Rightarrow \qquad (T_c - T)^{-\alpha-\gamma+2-2\beta} \geq A \cdot T$$

where A is a (positive) constant. Taking the limit $T \to T_c^-$:

$$\lim_{T \to T_c^-} (T_c - T)^{-\alpha - \gamma + 2 - 2\beta} \geq A \cdot T_c$$

and this inequality can be satisfied only if the left hand side doesn't tend to zero. This happens if $-\alpha - \gamma + 2 + 2\beta \leq 0$, namely:

$$\alpha + 2\beta + \gamma \geq 2$$

This is the *Rushbrooke inequality*.

The *Griffiths inequality*, on the other hand, is:

$$\alpha + \beta(1 + \delta) \geq 2$$

and can be obtained from the fact that the free energy is a convex function.

At this level they are both inequalities; however, when the first critical transitions were studied, numerical simulations showed that the critical exponents satisfied them as *exact equalities*. We will show much later on (see 7.1.1), that once assumed the *static scaling hypothesis* Rushbrooke and Griffiths inequalities will become *exact* equalities.

Part II

Equilibrium statistical mechanics

Chapter 2

Brownian motion

2.1 Random walks and diffusion

Random walks have been introduced historically in order to explain the completely random motion exhibited by pollen grains in water (the so called *Brownian motion*). Einstein was the first to show that this particular behaviour could be explained, in terms of random walks, as the result of the impacts of the pollen grains with water molecules: since at the time the discrete nature of matter was still not universally accepted, this was one of the first theoretical predictions of the existence of atoms and molecules, which was later proofed experimentally.

2.1.1 General properties of random walks

We will now study the general properties of random walks.
Consider a point particle in the real axis starting form the origin and moving by steps s_i (which can be thought to be done at times $i\Delta t$) independent from each other. Supposing that $s_i = \pm 1$ (namely the particle can make one step to the right or left each time) the position of the particle after N steps (note that the number of steps done is proportional to the time passed) will be:

$$x_N = s_1 + s_2 + \cdots + s_N$$

What we would like to understand is the following: if we make the same particle start from the origin for N times and move freely, what is its average behaviour? Equivalently, if we make N different independent particles start at the same time from the origin, how will the system evolve and which will be its properties?
Since the particles at every step have the same probability to go right or left, namely $p(s_i = 1) = p(s_i = -1) = 1/2$, then:

$$\langle s_i \rangle = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (-1) = 0 \qquad \Rightarrow \qquad \langle x_N \rangle = \sum_{i=1}^{N} \langle s_i \rangle = 0$$

so the mean position is not really relevant. We can however compute the variance from this mean position to understand how the particles are distributed:

$$\sigma_N^2 = \langle x_N^2 \rangle - \langle x_N \rangle^2 = \langle x_N^2 \rangle = \left\langle \left(\sum_{i=1}^{N} s_i \right)^2 \right\rangle = \left\langle \sum_{i,j=1}^{N} s_i s_j \right\rangle = \sum_{i,j=1}^{N} \langle s_i s_j \rangle$$

However, since the steps are independent we have $\langle s_i s_j \rangle = \langle s_i \rangle \langle s_j \rangle$ if $i \neq j$. In other words:

$$\langle s_i s_j \rangle = \begin{cases} \langle s_i \rangle \langle s_j \rangle = 0 & i \neq j \\ \langle s_i^2 \rangle = 1 & i = j \end{cases} \qquad \Rightarrow \qquad \langle s_i s_j \rangle = \delta_{ij}$$

so that:

$$\sigma_N^2 = \sum_{i,j=1}^{N} \langle s_i s_j \rangle = \sum_{i,j=1}^{N} \delta_{ij} = N$$

Therefore, after N steps the particles are distributed in a region of width $\sigma_N = \sqrt{N}$.
This allows us to highlight a great difference with the more familiar uniform linear motion: for random walks the mean displacement does *not* grow linearly with time, but depends on its *square root*.

If we now move to three dimensions, calling $\vec{s}_i$ the random steps the position of one particle after N steps will be:

$$\vec{r}_N = \sum_{i=1}^{N} \vec{s}_i$$

and supposing $|\vec{s}_i| = \ell \quad \forall i$, then:

$$\langle \vec{r}_N \rangle = \sum_{i=1}^{N} \langle \vec{s}_i \rangle = 0$$

since the $\vec{s}_i$-s are uniformly distributed on a sphere.
We will also have $\langle \vec{s}_i{}^2 \rangle = \ell^2$ and therefore this time:

$$\langle \vec{s}_i \cdot \vec{s}_j \rangle = \begin{cases} \langle \vec{s}_i \rangle \langle \vec{s}_j \rangle = 0 & i \neq j \\ \langle \vec{s}_i{}^2 \rangle = \ell^2 & i = j \end{cases} \qquad \Rightarrow \qquad \langle \vec{s}_i \cdot \vec{s}_j \rangle = \ell^2 \delta_{ij}$$

so that the variance after N steps is:

$$\sigma_N^2 = \langle \vec{r}_N{}^2 \rangle = \left\langle \sum_{i=1}^{N} \vec{s}_i \cdot \sum_{j=1}^{N} \vec{s}_j \right\rangle = \sum_{i,j=1}^{N} \langle \vec{s}_i \cdot \vec{s}_j \rangle = N\ell^2$$

Therefore, this time the particles will be distributed in a region of linear dimension $\sigma_N = \ell \sqrt{N}$; as in the previous case, the variance increases with the square root of the number of steps (namely, the square root of time).

2.1.2   The diffusion equation

Let us now return to the one-dimensional case. The question we would like to answer now is the following: in the continuum limit, if we make many different independent particles start from the origin at the same instant, supposing that they undergo a random walk where the length of each step is regulated by a probability distribution $\chi(\ell)$, how many particles will there be at time t in the interval $[x, x + dx]$?
We will see that the answer to this question leads us to the *diffusion equation*.

Let us suppose that $\langle \ell \rangle = 0$ and $\langle \ell^2 \rangle = a^2$, with a finite (which is rather reasonable[1]). The position of the particle at time $t + \Delta t$ given the one at time t will be:

$$x(t + \Delta t) = x(t) + \ell(t)$$

---

[1]The fact that $\langle \ell \rangle = 0$ of course means that the random walk is *unbiased*. As we will later see in 2.1.3, we can also consider cases where the random walk is actually biased by an external force.
Note also that, however, there can be cases where $\chi(\ell)$ does not have a finite variance, for example power law distributions. In this case the resulting motion is called *Lévy flight*, and the main difference with a "normal" random walk is that in a Lévy flight the particles sometimes make very long steps. Such systems have been also used to model the movement of animal herds.

If we call $\rho(x,t)$ the particle density, then at time $t + \Delta t$ (with Δt a small time interval, and we will later take the limit $\Delta t \to 0$) we will have:

$$\rho(x, t + \Delta t) = \int \rho(x', t)\chi(x - x')dx'$$

In fact, a particle starting from x' at time t will be in x at time $t + \Delta t$ with probability $\chi(x - x')$, and so integrating over all the possible initial positions of the particles (given by the particle density at time t) we get exactly the new particle distribution. Defining the variable $\ell = x - x'$:

$$\rho(x, t + \Delta t) = \int \chi(\ell)\rho(x - \ell, t)d\ell$$

Let us now suppose that the step size ℓ is small compared to the scales on which ρ varies (which is another reasonable assumption); this way we can expand ρ in a Taylor series[2]:

$$\rho(x - \ell, t) = \rho(x, t) - \ell\frac{\partial}{\partial x}\rho(x, t) + \frac{\ell^2}{2}\frac{\partial^2}{\partial x^2}\rho(x, t) - \frac{\ell^3}{3!}\frac{\partial^3}{\partial x^3}\rho(x, t) + \cdots$$

so that we can write:

$$\rho(x, t + \Delta t) = \int \chi(\ell)\left(\rho - \ell\frac{\partial\rho}{\partial x} + \frac{\ell^2}{2}\frac{\partial^2\rho}{\partial x^2} - \frac{\ell^3}{3!}\frac{\partial^3\rho}{\partial x^3} + \cdots\right)d\ell =$$

$$= \rho - \frac{\partial\rho}{\partial x}\langle\ell\rangle + \frac{\partial^2\rho}{\partial x^2}\frac{\langle\ell^2\rangle}{2} - \frac{\partial^3\rho}{\partial x^3}\frac{\langle\ell^3\rangle}{3!} + \cdots$$

If we now also expand ρ for small Δt and suppose for simplicity that $\langle\ell^3\rangle = ca^3$ with c a constant[3], then:

$$\Delta t\frac{\partial\rho}{\partial t} + \frac{\Delta t^2}{2}\frac{\partial^2\rho}{\partial t^2} = \frac{a^2}{2}\frac{\partial^2\rho}{\partial x^2} - c\frac{a^3}{6}\frac{\partial^3\rho}{\partial x^3} + \cdots$$

and dividing by Δt:

$$\frac{\partial\rho}{\partial t} + \frac{\Delta t}{2}\frac{\partial^2\rho}{\partial t^2} = \frac{a^2}{2\Delta t}\frac{\partial^2\rho}{\partial x^2} - \frac{ca^3}{6\Delta t}\frac{\partial^3\rho}{\partial x^3} + \cdots \tag{2.1}$$

We are now interested in the continuum limit, namely $\Delta t \to 0$ and $a \to 0$; this means that in general the limit of $a^2/(2\Delta t)$ is not well determined. In particular, we could have:

1. $a^2/\Delta t \to 0$, in which case $\partial\rho/\partial t = 0$: there is no evolution

2. $a^2/\Delta t \to \infty$, in which case the evolution would be instantaneous

3. $a^2/\Delta t \to$ const.

The most interesting case is the last one: in fact, if we require $a^2/2\Delta t$ to be *always* (so also in the limits $a, \Delta t \to 0$) equal to a constant D, called *diffusion constant*, then in equation (2.1) the terms proportional to $\partial^2\rho/\partial t^2$ and $\partial^3\rho/\partial x^3$ and all the other terms in the expansion vanish[4]. Therefore, we find that $\rho(x, t)$ must satisfy the so called *diffusion equation*[5]:

$$\frac{\partial}{\partial t}\rho(x, t) = D\frac{\partial^2}{\partial x^2}\rho(x, t)$$

---

[2]We also keep the third order in the expansion, to make explicit that all the terms beyond the second order vanish in the limits we will take.

[3]An example of probability distribution that satisfies this requirement is the Poisson distribution.

[4]In particular, the term proportional to $\partial^3\rho/\partial x^3$ vanishes because:

$$\frac{ca^3}{6\Delta t} = \frac{c}{6}aD \to 0$$

[5]Fun fact: the diffusion equation with an imaginary time is the Schrödinger equation for a free particle (of course provided that ρ is interpreted as a wave function), with a diffusion constant equal to $D = \hbar/2m$.

which in three dimensions can be rewritten as:

$$\frac{\partial}{\partial t}\rho(\vec{x},t) = D\nabla^2\rho(\vec{x},t) \tag{2.2}$$

Diffusion and continuity equation

We have until now not considered a very important property of ρ: since it is a particle density, it must satisfy a *continuity equation*. This is an equation that expresses the fact that particles cannot "disappear" and "reappear" in different points of space, but must continuously flow from point to point.
To make things clearer let us consider a region of volume V of our system: this will contain a certain number of particles, which will change with time because some particles will go out of V while others will enter into the volume from the outside, due to the continuous random motion they are subjected to. Therefore, in terms of $\rho(\vec{r},t)$ the number of particles contained in the volume at time $t + \Delta t$ will be:

$$\int_V \rho(\vec{r},t+\Delta t)d\vec{r} = \int_V \rho(\vec{r},t)d\vec{r} + \text{in} - \text{out}$$

where by "in" and "out" we mean the number of particles that have entered or gone out of V. If we call $\vec{J}$ the flow of particles[6], and $\hat{n}$ the unit vector orthogonal to the surface S enclosing V pointing outward, we can rewrite this equation as:

$$\int_V \rho(\vec{r},t+\Delta t)d\vec{r} = \int_V \rho(\vec{r},t)d\vec{r} - \int_S \vec{J}\cdot\hat{n}\Delta t dS$$

(where the last term is the net outgoing flow of particles). Dividing by Δt:

$$\int_V \frac{\rho(\vec{r},t+\Delta t) - \rho(\vec{r},t)}{\Delta t}d\vec{r} = -\int_S \vec{J}\cdot\hat{n}dS$$

Taking the limit $\Delta t \to 0$ and using Gauss theorem we get:

$$\int_V \frac{\partial\rho}{\partial t}d\vec{r} = -\int_V \vec{\nabla}\cdot\vec{J}d\vec{r}$$

Since the volume V is arbitrary, the integrands must be always equal:

$$\frac{\partial}{\partial t}\rho(\vec{r},t) = -\vec{\nabla}\cdot\vec{J}(\vec{r},t)$$

This is the *continuity equation* associated to ρ, with flow $\vec{J}$.

Now, how are continuity and diffusion equations related?
It is immediate to see that they are equivalent if:

$$\vec{J}(\vec{r},t) = -D\vec{\nabla}\rho(\vec{r},t) \tag{2.3}$$

In fact, this way:

$$\frac{\partial\rho}{\partial t} = -\vec{\nabla}\cdot\left(-D\vec{\nabla}\rho\right) = D\nabla^2\rho$$

---

[6]Remember that in general the flow of particles $\vec{J}$ is defined as a vector such that $|\vec{J}|$ is the number of particles that pass through the surface orthogonal to $\vec{J}$ per unit time and unit surface. More in general, if $\hat{n}$ is a unit vector orthogonal to a surface, then $\vec{J}\cdot\hat{n}$ is the number of particles passing through the surface along the direction of $\hat{n}$ per unit time and surface. If $\vec{J}\cdot\hat{n} < 0$, this means that the particles are passing through the surface along the direction of $-\hat{n}$.

The physical meaning of equation (2.3) is that a set of particles subjected to a random walk will move from areas of high to areas of low concentration. In other words, the diffusion of particles tends to "flatten" concentration inhomogeneities[7].

Random walks and central limit theorem

We now want to show a particular property of random walks, and so also of diffusive phenomena: we want to see that they are coherent with the central limit theorem. In fact, this states that if $x_N = \sum_i \ell_i$ is a sum of independent random variables, then x_N for large N is distributed along a Gaussian: what we want to show is that this is indeed the case.
To be more precise, let us reconsider the one-dimensional discrete random walk with constant step length a. Let us call x the position of the particle at the N-th step, n_+ the number of steps that it has done to the right and n_- to the left; of course $N = n_+ + n_-$, and we also call $m = n_+ - n_-$, so that $m = x/a$. What we want to do is to determine the probability $P(m, N)$ that the particle is at position m after N steps, and see how it behaves for large N.

Now, since the probabilities for a step to be done to the right or to the left are the same, and equal to $1/2$, the probability $P(m, N)$ will be a binomial one[8]:

$$P(m, N) = \binom{N}{n_+} \frac{1}{2^{n_+}} \frac{1}{2^{N-n_+}} = \binom{N}{\frac{N+m}{2}} \frac{1}{2^N} = \frac{N!}{\left(\frac{N+m}{2}\right)! \left(\frac{N-m}{2}\right)!}$$

It is easier to handle this expression if we take its logarithm:

$$\ln P(m, N) = \ln N! - \ln \left(\frac{N+m}{2}\right)! - \ln \left(\frac{N-m}{2}\right)!$$

Using Stirling's approximation (see appendix C):

$$\ln N! \sim N \ln N - N + \frac{1}{2} \ln(2\pi N)$$

and supposing that m is small compared to N, we can use the logarithm expansion $\ln(1 + x) \sim x - x^2/2 + \cdots$; in the end we get:

$$\ln P(m, N) = -\frac{m^2}{2N} + \ln \frac{1}{\sqrt{2\pi N}}$$

Therefore, exponentiating we have that:

$$P(m, N) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{m^2}{2N}} \qquad N \text{ large}$$

In other words, the probability of finding a particle in a given position for large times is distributed along a Gaussian, which is what we expected from the central limit theorem.
Note that the result we have just found is *completely general*: we have *not* solved the diffusion equation in order to find it, so we can argue that the density of diffusing particles will be a Gaussian for large enough times *independently of the initial conditions*. In other words no matter

---

[7]However, there are situations were diffusion (despite the presence of the negative sign in equation (2.3)) can actually highlight and increase concentration inhomogeneities, spontaneously bringing a system from a homogeneous to an inhomogeneous configuration. These are called *patterning phenomena*.

[8]We can also justify this thinking that the choice of the direction of the step is regulated by the flips of a coin.

what the initial shape of ρ was, for large enough times it will have become a Gaussian.

A final remark: the expression of P for large N is formally defined also for $m > N$. However, this is clearly impossible: if the particle has done N steps on the right its final position can't be $m > N$. Should we care about this problem?
Not really much, in reality: it is in fact true that formally we should also have non-null probabilities to find the particles for $m > N$, but in practice these are ridiculously small (we are considering events beyond N sigmas from the mean of the Gaussian!), so they do not constitute a real problem.

The solution of the diffusion equation

We now want to solve the diffusion equation. For the sake of simplicity, we will do that in one dimension, and the first approach we will use is that of a Fourier decomposition.

We begin by noting that the operators $\partial^2/\partial x^2$ and the translation operator T_a, defined as $(T_a f)(x) = f(x + a)$, commute; in fact:

$$\left(\frac{\partial^2}{\partial x^2} T_a f \right)(x) = \frac{\partial^2}{\partial x^2} f(x + a) \qquad\qquad \left(T_a \frac{\partial^2}{\partial x^2} f \right)(x) = \frac{\partial^2}{\partial x^2} f(x + a)$$

Therefore, since $\partial^2/\partial x^2$ and T_a commute they have a common basis of eigenfunctions. Thus, if we find the eigenfunctions of T_a:

$$T_a \varphi = \lambda \varphi \qquad \Rightarrow \qquad \varphi(x + a) = \lambda \varphi(x) \qquad \Rightarrow \qquad \varphi_k(x) = e^{ikx} \quad \text{and} \quad \lambda = e^{ika}$$

we have also found the eigenfunctions of $\partial^2/\partial x^2$:

$$\frac{\partial^2}{\partial x^2} \varphi_k(x) = -k^2 e^{-ikx} = -k^2 \varphi_k(x)$$

These are plain waves, and in order to make them orthonormal we redefine them as:

$$\varphi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}$$

so that we have indeed:

$$\int_{-\infty}^{+\infty} \varphi_k(x) \varphi_k^*(y) dk = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x-y)} dk = \delta(x - y)$$

Therefore, we can expand any given ρ in terms of plane waves:

$$\rho(x, t) = \frac{1}{2\pi} \int \rho_k(t) e^{ikx} dk$$

where $\rho_k(t)$ are the *Fourier coefficients*, given by:

$$\rho_k(t) = \int e^{-ikx} \rho(x, t) dx$$

This way we have:

$$\frac{\partial}{\partial t} \rho(x, t) = \frac{1}{2\pi} \int \frac{\partial \rho_k}{\partial t} e^{ikx} dk \qquad\qquad D \frac{\partial^2}{\partial x^2} \rho(x, t) = \frac{1}{2\pi} \int \left(-Dk^2 \right) \rho_k(t) e^{ikx} dk$$

and the validity of the diffusion equation implies that the integrands must be equal, i.e.:

$$\frac{\partial}{\partial t}\rho_k(t) = -Dk^2\rho_k(t) \qquad \Rightarrow \qquad \rho_k(t) = e^{-Dk^2(t-t_0)}\rho_k(t_0)$$

Note that this means that the various Fourier modes of the initial conditions are "dampened" as time passes, and the larger is k the the more dampened $\rho_k(t)$ will be.

Therefore, substituting in $\rho(x,t)$:

$$\rho(x,t) = \frac{1}{2\pi}\int\rho_k(t)e^{ikx}dk = \frac{1}{2\pi}\int\rho_k(t_0)e^{ikx-Dk^2(t-t_0)}dk$$

Everything we have stated so far is abslutely general; we now consider a very special case: let us suppose $t_0 = 0$ for simplicity and set $\rho(x,0) = \delta(x)$, namely at the beginning all the particles are located at the origin. We have:

$$\rho_k(0) = \int e^{-ikx}\delta(x)dx = 1 \qquad \Rightarrow \qquad \rho(x,t) = \frac{1}{2\pi}\int e^{-Dk^2t+ikx}dk \qquad (2.4)$$

This is a *Fresnel integral*, which can be computed with complex analysis using Cauchy's theorem; however, we use a "trick" to determine ρ without explicitly calculating it, based on the fact that:

$$\int_{-\infty}^{+\infty}e^{-z^2}dz = \sqrt{\pi}$$

Deriving ρ in equation (2.4) with respect to x, we have:

$$\frac{\partial\rho}{\partial x} = \frac{i}{2\pi}\int_{-\infty}^{+\infty}\underbrace{ke^{-Dk^2t}}_{-\frac{1}{2Dt}\frac{\partial}{\partial k}e^{-Dk^2t}}e^{ikx}dk = -\frac{i}{4\pi Dt}\int_{-\infty}^{+\infty}\left(\frac{\partial}{\partial k}e^{-Dk^2t}\right)e^{ikx}dk =$$

$$= -\frac{i}{4\pi Dt}\left(\underbrace{e^{-Dk^2t+ikx}\Big|_{-\infty}^{+\infty}}_{0} - \underbrace{\int_{-\infty}^{+\infty}ixe^{-Dk^2t+ikx}dk}_{ix2\pi\rho(x,t)}\right) = -\frac{x}{2Dt}\rho(x,t)$$

Therefore:

$$\rho(x,t) = \text{const.}\cdot e^{-\frac{x^2}{4Dt}}$$

and the constant can be determined from the normalization condition, which leads to:

$$\rho(x,t) = \frac{1}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}$$

This is the solution of the diffusion equation for the initial condition $\rho(x,0) = \delta(x)$. Let us see its properties:

0. After $t = 0$, ρ is a Gaussian of height $1/\sqrt{4\pi Dt}$ and width $\sqrt{2Dt}$: as time passes this Gaussian lowers and widens

1. As $t \to 0$, $\rho(x,t) \to \delta(x)$ (comprehensibly)

2. The mean value $\langle x\rangle_t$ of x at time t is null:

$$\langle x\rangle_t = \int\rho(x,t)xdx = 0$$

since the integrand is odd. Then calling $\alpha = 1/4Dt$ for simplicity, we have[9]:

$$\sigma_t^2 = \langle x^2 \rangle_t = \int \rho(x,t) x^2 dx = \int \frac{x^2 e^{-\alpha x^2}}{\sqrt{\pi/\alpha}} dx = -\sqrt{\frac{\alpha}{\pi}} \frac{\partial}{\partial \alpha} \int e^{-\alpha x^2} dx =$$

$$= -\sqrt{\frac{\alpha}{\pi}} \frac{\partial}{\partial \alpha} \sqrt{\frac{\pi}{\alpha}} = -\frac{\partial}{\partial \alpha} \ln \sqrt{\frac{\pi}{\alpha}} = \frac{1}{2\alpha} = 2Dt$$

Therefore $\sigma_t = \sqrt{2Dt}$, a result we already knew from the general properties of random walks.

If we use as initial condition $\rho(x,0) = \delta(x-y)$, then by definition $\rho(x,t)$ is the *Green's function* of the diffusion equation:

$$\rho(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{(x-y)^2}{4Dt}} := G(x-y,t)$$

Green's functions can be used to solve differential equations, like the diffusion one, with an approach different from the Fourier decomposition.
In fact, if we can (at least formally) write ρ as a superposition in the following form:

$$\rho(x,0) = \int \rho(y,0)\delta(x-y)dy$$

then the general solution of the diffusion equation is:

$$\rho(x,t) = \int G(x-y,t)\rho(y,0)dy$$

Of course, both Fourier's and Green's methods are equivalent, but depending on the situation and the initial conditions considered one can result more convenient than the other.

2.1.3   Fokker-Planck and Langevin's equations

In studying the diffusion equation we have neglected the possible presence of any external force. We could ask, however, what happens if we include them; as we will see here, we will obtain the so called *Fokker-Planck equation*.
So, let's go back to the beginning and start with the same assumptions we made in 2.1.2, with the difference that this time we also include an external force F. Therefore if $x(t)$ is the position of the particle at time t, after a little time interval Δt it will be:

$$x(t+\Delta t) = x(t) + \ell(t) + F\gamma\Delta t \tag{2.5}$$

where γ is the so called *mobility* of the particles.

Depending on the nature of the system, the expression of the mobility γ can vary. For example, if our particles are diffusing through a dense fluid, the mobility will be essentially related to the viscosity of the fluid itself; in fact, it is a known fact that the equation of motion of a body in a fluid is (in one dimension):

$$m\ddot{x} = F - k\eta v$$

where $v = \dot{x}$ is the velocity of the body, η the viscosity of the fluid and k a geometric factor (for example, from Stokes's law in the case of spherical bodies we have $k = 6\pi r$, with r the radius of the sphere). In this case if F is constant the body will rapidly acquire a drift velocity:

$$\ddot{x} = 0 \quad \Rightarrow \quad \dot{x} = v_\infty = \frac{F}{k\eta} \quad \Rightarrow \quad x(t+\Delta t) = x(t) + \frac{F}{k\eta}\Delta t$$

---

[9]This result can be obtained much more easily from the fact that $\rho(x,t)$ is a Gaussian.

so in the case of particles diffusing through a viscous fluid we have $\gamma = 1/k\eta$.

On the other hand, in the case of a dilute fluid, supposing again that F is constant (or equivalently that the free mean path of the particles is much smaller than the scales over which F varies) we will have:

$$x(t + \Delta t) = x(t) + \frac{1}{2}\Delta t^2 \frac{F}{m}$$

where we are only considering the drift due to F (we have neglected the possible term $v\Delta t$, where v is the velocity of the particle after a scattering, since it is isotropic and so it does not contribute to the net displacement of the particles). Now, since $D = a^2/2\Delta t$:

$$\frac{1}{2}\Delta t^2 \frac{F}{m} = \frac{1}{2}\Delta t^2 \frac{F}{m}\frac{2\Delta t D}{a^2} = \frac{F\Delta t}{m(a/\Delta t)^2}D := F\gamma\Delta t$$

and remembering also that $\langle \ell^2 \rangle = a^2$, so that the mean displacement after a scattering is $\Delta x = \sqrt{\langle \ell^2 \rangle} = a$ and therefore $(a/\Delta t)^2 = \langle v^2 \rangle$ (the mean square velocity of the particles) we have:

$$\gamma = \frac{D}{m \langle v^2 \rangle} = \frac{D}{2E_{\text{kin}}}$$

where of course E_{kin} is the kinetic energy of the particles.

Now that we also know what the mobility is, we can proceed to determine how the diffusion equation changes with the presence of the external and static force F.

Consider an arbitrary function $f(x)$; then:

$$\langle f(x(t)) \rangle = \int \rho(x, t) f(x) dx$$

Therefore, computing $\langle f(x(t + \Delta t)) \rangle$ we get:

$$\langle f(x(t + \Delta t)) \rangle = \langle f(x(t) + \ell + F\gamma\Delta t) \rangle \qquad \Rightarrow$$

$$\Rightarrow \qquad \int \rho(x, t + \Delta t) f(x) dx = \int \rho(x, t) f(x + \ell + \gamma F\Delta t)\chi(\ell) dx d\ell$$

where the presence of $\chi(\ell)$ and the integration over ℓ are due to the fact that at time $t + \Delta t$ the coordinate x will be equal to $x + \ell + \gamma F\Delta t$ with probability $\chi(\ell)$ (just like we have seen in the derivation of the diffusion equation in 2.1.2).

Since Δt and ℓ are small, we can expand f in the right hand side:

$$\int \rho(x, t + \Delta t) f(x) dx =$$

$$= \int \chi(\ell)\rho(x, t) \left[f(x) + \frac{\partial f}{\partial x}(\gamma F\Delta t + \ell) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\gamma F\Delta t + \ell)^2 + \cdots \right] dx d\ell =$$

$$= \int \rho(x, t) \left[f(x) + \frac{\partial f}{\partial x}\gamma F\Delta t + \frac{\partial f}{\partial x}\langle \ell \rangle + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\gamma F\Delta t)^2 + \right.$$

$$\left. + \frac{\partial^2 f}{\partial x^2}\gamma F\Delta t \langle \ell \rangle + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\langle \ell^2 \rangle + \cdots \right] dx$$

Neglecting all the higher order terms and remembering that $\langle \ell \rangle = 0$ and $\langle \ell^2 \rangle = a^2$, we get:

$$\int \rho(x, t + \Delta t) f(x) dx = \int \rho(x, t) \left(f(x) + \gamma F\Delta t \frac{\partial f}{\partial x} + \frac{a^2}{2}\frac{\partial^2 f}{\partial x^2} \right) dx$$

Integrating by parts:

$$\int \rho(x, t + \Delta t) f(x) dx = \int f(x) \left(\rho - \gamma \Delta t \frac{\partial}{\partial x} (F\rho) + \frac{a^2}{2} \frac{\partial^2 \rho}{\partial x^2} \right) dx + \rho(x, t) f(x) F \gamma \Delta t \Big|_{-\infty}^{+\infty} +$$

$$+ \frac{a^2}{2} \left(\rho(x, t) \frac{\partial f}{\partial x} + \frac{\partial \rho}{\partial x} f(x) \right) \Big|_{-\infty}^{+\infty}$$

Assuming that ρ and $\partial \rho / \partial x$ vanish sufficiently fast at infinity, the last two terms are null so that we are left with:

$$\int \rho(x, t + \Delta t) f(x) dx = \int f(x) \left(\rho - \gamma \Delta t \frac{\partial}{\partial x} (F\rho) + \frac{a^2}{2} \frac{\partial^2 \rho}{\partial x^2} \right) dx$$

Since $f(x)$ is arbitrary, this equality holds if the remaining part of the integrands are equal; therefore, dividing also by Δt:

$$\frac{\rho(x, t + \Delta t) - \rho(x, t)}{\Delta t} = -\gamma \frac{\partial}{\partial x} (\rho F) + \frac{a^2}{2\Delta t} \frac{\partial^2 \rho}{\partial x^2}$$

and taking the limits $a \to 0$ and $\Delta t \to 0$ with $a^2 / 2\Delta t = D = \text{const.}$, we have:

$$\frac{\partial}{\partial t} \rho(x, t) = \frac{\partial}{\partial x} \left(-\gamma \rho(x, t) F + D \frac{\partial}{\partial x} \rho(x, t) \right) \tag{2.6}$$

This is the so called *Fokker-Planck equation*, while (2.5) is the *Langevin's equation*.

Let us note that this is still compatible with a continuity equation, but this time the flow of particles is given (in three dimensions) by:

$$\vec{J} = \gamma \rho \vec{F} - D \vec{\nabla} \rho$$

where the first term is the flow due to the external force, while the second one is the usual diffusive term due to the density inhomogeneities of the system.

Just to see the Fokker-Planck equation at work in a very simple case, let us determine its stationary solution $\rho^*(x)$ when F is a uniform gravitational force, with x being the vertical coordinate (which is the only significant one; we assume that the positive x direction is that pointing downwards); since $F = mg$ (with m the mass of the particles), then $\partial \rho^* / \partial t = 0$ if[10]:

$$-\rho^* \gamma mg + D \frac{\partial \rho^*}{\partial x} = 0 \qquad \Rightarrow \qquad \rho^*(x) = e^{\frac{m\gamma g}{D} x} \rho^*(0) = e^{-\frac{\gamma}{D} U(x)} \rho^*(0)$$

where $U(x) = -mgx$ is the gravitational potential.

---

[10]Of course, we could have equivalently solved the equation from a direct application of (2.6), namely from $-\gamma mg \partial \rho^* / \partial x + D \partial^2 \rho^* / \partial x^2 = 0$.

Chapter 3

Ensemble theory

3.1 A bridge between the microscopic and the macroscopic

The fundamental question from which statistical mechanics has risen is the following: *where does thermodynamics come from?*

In fact, we know that fluids and gases are made of particles (atoms or molecules) and in principle we could use the tools of classical mechanics in order to study their motion; therefore, we could theoretically describe the system at least from a microscopic perspective. We can however wonder how this microscopic point of view is related to the macroscopic description of systems given by thermodynamics. In other words: how do the thermodynamic laws we know come from the microscopic motion of particles?

What we want to do now is exactly to establish this link, i.e. to derive the thermodynamics of a macroscopic system at equilibrium from its microscopic properties. This is the final purpose of equilibrium statistical mechanics.

We now therefore outline the general theoretical framework that will be needed in order to develop this theory.

Let us consider an isolated system composed of N particles, with volume V and energy E. Since it is isolated its energy, momentum $\vec{P}$ and angular momentum $\vec{L}$ are conserved; however, considering the system as fixed and still we can set $\vec{P} = 0$ and $\vec{L} = 0$ so that the energy E is its only non null conserved quantity. If we call $\vec{q}_i$ and $\vec{p}_i$, respectively, the position and momentum of the i-th particle the dynamics of the system can be obtained from its Hamiltonian:

$$\mathcal{H}(\vec{q}_1, \ldots, \vec{q}_N, \vec{p}_1, \ldots, \vec{p}_N) = \sum_{i=1}^{N} \frac{|\vec{p}_i|^2}{2m_i} + \mathcal{V}(\vec{q}_1, \ldots, \vec{q}_N)$$

(where $\mathcal{V}$ is a generic interaction potential acting between the particles) through Hamilton's equations:

$$\frac{d\vec{q}_i}{dt} = \vec{\nabla}_{\vec{p}_i} \mathcal{H} \qquad \frac{d\vec{p}_i}{dt} = -\vec{\nabla}_{\vec{q}_i} \mathcal{H}$$

However, solving these equations for a macroscopic system is impractical for two reasons:

1. The number of particles in the system is insanely huge, in general of the order of Avogadro's number, i.e. $\sim 10^{23}$. We therefore should solve a system of approximately 10^{23} coupled differential equations, which is rather impossible (also from a computational point of view)

2. Even if we could solve them the solutions of Hamilton's equations would give no significant information about the system; for example it is much more interesting to know the average number of particles that hit a wall of the system per unit time than knowing exactly which particle hits the wall at a given instant

Furthermore a lot of interesting systems exhibit chaotic behaviours, namely their time evolution depends strongly on initial conditions making essentially useless any exact solution of Hamilton's equations.

Therefore, it is clear how a *statistical* treatment of many-particle systems is essential in order to obtain relevant information and ultimately derive their thermodynamics.
The fundamental concept that allows one to develop such statistical description is that of *ensemble*, which we now introduce.
Consider the generic many-particle system that we have introduced earlier, and for the sake of simplicity call $\mathbb{Q} = (\vec{q}_1, \ldots, \vec{q}_N)$ and $\mathbb{P} = (\vec{p}_1, \ldots, \vec{p}_N)$ the set of all the coordinates and momenta of the particles. The $3N$-dimensional spaces where $\mathbb{Q}$ and $\mathbb{P}$ live are called, respectively, *configuration* and *momentum space*, while the $6N$-dimensional space where $(\mathbb{Q}, \mathbb{P})$ lives is called *phase space* of the system, often referred to as Γ.
Once all the positions and momenta of the particles have been given (i.e. once we have all the possible information about its microscopic configuration) the whole system is identified with a unique point $(\mathbb{Q}, \mathbb{P})$ in phase space, sometimes called *microstate* or *representative point* of the system, and as the system evolves (i.e. the particles move, thus changing their positions, and interact with each other, thus changing their momenta) this point moves in phase space describing a trajectory. The exact solution of Hamilton's equations for this system would give us the expression of this trajectory, but as we have seen before this is not a really useful information.
We therefore change point of view: if we look at our system under a "macroscopic perspective", in general it will be subjected to some constraints like the conservation of energy (in case it is isolated) or the conservation of volume etc., and therefore the macroscopic properties of a system have precise values. This suggests that we can define a *macrostate* of the system, i.e. describe it only with some of its bulk properties (which is exactly the approach of thermodynamics).
We thus have two substantially different ways to describe the same system: a macroscopic and a microscopic one.
Now, for a given macrostate of the system there will be multiple microstates which are compatible with the first one, namely there are many microscopic configurations of the system that satisfy the same macroscopic constraints (have the same energy, volume etc.) and obviously they are all equivalent from a macroscopic point of view.
The set of all the possible microstates which are compatible with a given macrostate of the system is called *ensemble*.

The approach of statistical mechanics consists, essentially, in studying the *average* behaviour of the elements of an ensemble rather than the *exact* behaviour of a single particular system.

Depending on the constraints set on a system its ensemble changes name, and in particular there are three kind of ensembles:

– **microcanonical ensemble**: when the system is completely isolated and has fixed values of energy E, volume V and number of particles N

– **canonical ensemble**: when the system can exchange energy with its surroundings

– **grand canonical ensemble**: when the system can exchange energy and particles with its surroundings

We will now proceed to study the properties of such ensembles and see how we can link them with thermodynamics.

3.2 The microcanonical ensemble

Now that we have laid out the general framework we needed we can proceed to study the properties of the *microcanonical ensemble*; we therefore have a system with fixed values of energy E, volume V and number of particles N.

We have just seen that the ensemble of such a system is constituted by a multitude of equivalent microstates; since we have no additional information we can assume that all these microscopic configurations are equally probable. In other words, we introduce statistics in our treatment by formulating the so called *a priori equal probability postulate*:

> *If a system is in a given macroscopic configuration it can be found with equal probability in any of the microstates of its ensemble.*

Mathematically this means introducing a constant probability density $\rho(\mathbb{Q}, \mathbb{P})$ in the ensemble of the system, namely:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{\delta(\mathcal{H} - E)}{\Omega(E, V, N)}$$

where $\Omega(E, V, N)$ is the volume occupied by the ensemble in phase space (i.e. the volume of the set of points corresponding to all the microstates of a given ensemble, which is a hypersurface of constant energy E):

$$\Omega(E, V, N) = \int_{\mathcal{H}=E} d\Gamma = \int d\Gamma \delta(\mathcal{H} - E)$$

where $d\Gamma$ is a short notation for $d\mathbb{Q}d\mathbb{P} = \prod_{i=1}^{N} d\vec{q}_i d\vec{p}_i$. If we now divide both sides by $\Omega(E, V, N)$ we see that $\rho(\mathbb{Q}, \mathbb{P})$ as we have defined it has indeed the meaning of a probability density[1] :

$$1 = \int d\Gamma \frac{\delta(\mathcal{H} - E)}{\Omega(E, V, N)} := \int d\Gamma \rho(\mathbb{Q}, \mathbb{P})$$

In particular the Dirac delta is needed to make ρ vanish everywhere but on the hypersurface of energy E in phase space, and Ω in order to correctly normalize ρ.

Therefore since we have introduced a probability density in phase space, if in general we define an observable $O(\mathbb{Q}, \mathbb{P})$ as a function of all the positions and momenta of the particles we can define its *mean value* in the ensemble as:

$$\langle O(\mathbb{Q}, \mathbb{P}) \rangle = \int \rho(\mathbb{Q}, \mathbb{P}) O(\mathbb{Q}, \mathbb{P}) d\Gamma$$

---

[1]What we are seeing now could also have been derived in a different but equivalent way.

Let us suppose that the energy of the system instead of being exactly equal to E can belong to the interval $[E, E + \Delta E]$ with $\Delta E \ll E$. This means that in phase space the system will occupy the region enclosed by the two hypersurfaces of energy E and $E + \Delta E$; the volume of this region can be written as:

$$\Omega(E, V, N)\Delta E = \int d\Gamma \left\{ \Theta\left[\mathcal{H}(\mathbb{Q}, \mathbb{P}) - E\right] - \Theta\left[\mathcal{H}(\mathbb{Q}, \mathbb{P}) - (E + \Delta E)\right] \right\}$$

where Θ is the Heaviside step function. Within the *theory of distributions* it can be shown that, formally, the derivative of Θ is the Dirac δ function, i.e. $d\Theta(x)/dx = \delta(x)$, so that:

$$\Theta\left[\mathcal{H} - E\right] - \Theta\left[\mathcal{H} - (E + \Delta E)\right] = -\frac{d}{dE}\Theta(\mathcal{H} - E)\Delta E = -\Delta E\left[-\delta(\mathcal{H} - E)\right] = \Delta E\delta(\mathcal{H} - E)$$

and thus:

$$\Omega(E, V, N) = \int d\Gamma \delta(\mathcal{H} - E)$$

On the other hand, we could have equivalently obtained the expression of $\rho(\mathbb{Q}, \mathbb{P})$ from the general definition of the mean value of an observable O over this ensemble:

$$\langle O(\mathbb{Q}, \mathbb{P}) \rangle = \frac{1}{\Omega(E, V, N)} \int d\Gamma \left\{ \Theta\left[\mathcal{H}(\mathbb{Q}, \mathbb{P}) - E\right] - \Theta\left[\mathcal{H}(\mathbb{Q}, \mathbb{P}) - (E + \Delta E)\right] \right\} O(\mathbb{Q}, \mathbb{P})$$

and proceeding in the same way.

This is the value of O that we actually measure: since the microstate of the system is continuously moving through the ensemble and since the time that macroscopic measurements require is many orders of magnitude longer than the time intervals typical of microscopic dynamics, we will only be able to measure ensemble averages[2].

Now, the mean value of an observable is significant if its variance is small, otherwise the results of a measure of the same quantity in the same conditions would fluctuate over a wide range making the observable essentially meaningless; for example, since we know that a system in equilibrium has a constant value of energy we expect that, in order to be a good theory, statistical mechanics can show that the statistical fluctuations of energy (at least for macroscopic systems) are very small and thus negligible, but of course this consideration can be extended to any observable (like the number of particles of a system)[3]. We will show that (fortunately!) this is always the case.

We can however already get a taste of that in a very simple situation: consider a gas of $2N$ particles in a cubic box of side L and let us mentally divide this box into two halves, the right and the left one. We ask: what is the probability to find $N - m$ particles in the right half and $N + m$ in the left one?

Since intuitively the probability to find a single particle in one half of the box is $1/2$ we will have:

$$P(m) = \frac{1}{2^{N+m}} \frac{1}{2^{N-m}} \binom{2N}{N+m} = \frac{1}{2^N} \frac{(2N)!}{(N+m)! \cdot (N-m)!}$$

where we have introduced the binomial factor because all the configurations which differ for the exchange of particles are equivalent, since they are identical. The original question we asked can now be rephrased as: how does $P(m)$ change with m?

If we call p the probability to find a particle in the left half and q the probability to find it in the right one (we will later set both equal to $1/2$, but let's distinguish them for now), we have:

$$P(m) = p^{N+m} q^{N-m} \binom{2N}{N+m}$$

and then:

$$\langle N + m \rangle = \sum_m (N+m) P(m) = p \frac{\partial}{\partial p} \underbrace{\sum_m P(m)}_{(p+q)^{2N}} = p \cdot 2N (p+q)^{2N-1}$$

Setting $p = q = 1/2$ we get:

$$\langle N + m \rangle = N \qquad \Rightarrow \qquad \langle m \rangle = 0$$

which is rather reasonable. Therefore, we expect that the configuration where N particles are in the right half and N in the left one is the most probable for our system. However, *how much more probable* is this configuration with respect to the other ones?

In order to understand that let us also compute the standard deviation from the mean value $\langle N + m \rangle$. We have:

$$p^2 \frac{\partial^2}{\partial p^2} \sum_m P(m) = \sum_m (N+m)(N+m-1) P(m) = \langle (N+m)^2 \rangle - \langle N+m \rangle \qquad \Rightarrow$$

---

[2]In other words, it is *impossible* to measure a macroscopic quantity relative to a *single* microstate of the system since in the time the measurement takes the system will have acquired many other different microscopic configurations and what we measure is the average (of course weighted with the probability density of the ensemble) over all the microscopic configurations acquired. This also relates to what we will see in 3.2.5 and appendix D.

[3]Of course this is not possible in the microcanonical ensemble, since both E and N are fixed, and will become possible in the other ensembles.

$$\Rightarrow \qquad \langle (N+m)^2 \rangle = p^2 \frac{\partial}{\partial p} \left[2N(p+q)^{2N-1} \right] + p2N(p+q)^{2N-1} =$$

$$= p^2 2N(2N-1)(p+q)^{2N-2} + p2N(p+q)^{2N-1}$$

and setting $p = q = 1/2$:

$$\langle (N+m)^2 \rangle = N^2 + \frac{N}{2}$$

Therefore:

$$\sigma^2_{\langle N+m \rangle} = \langle (N+m)^2 \rangle - \langle N+m \rangle^2 = \frac{N}{2} \qquad \Rightarrow \qquad \sigma_{\langle N+m \rangle} = \sqrt{\frac{N}{2}}$$

This means that the relative fluctuation is:

$$\frac{\sigma_{\langle N+m \rangle}}{\langle N+m \rangle} = \frac{1}{\sqrt{2N}}$$

which turns out to be astonishingly small: in fact if $N \sim 10^{23}$ this relative fluctuation is of the order of 10^{-11}. We can therefore conclude that the fluctuations of the number of particles in the two halves from their mean values are absolutely negligible (we never observe the gas spontaneously occupying only one half of the system!).

We can also obtain the same result in a slightly more complicated way, but which allows us to extract some more interesting information on the system.
In order to do that let us consider the logarithm of $P(m)$:

$$\ln P(m) = -2N \ln 2 + \ln(2N)! - \ln(N-m)! - \ln(N+m)!$$

Using Stirling's approximation (see appendix C) for large N we get:

$$\ln P(m) \sim -2N \ln 2 + 2N \ln(2N) + \ln \sqrt{2\pi \cdot 2N} - (N-m)\ln(N-m) -$$
$$- \ln \sqrt{2\pi(N-m)} - (N+m)\ln(N+m) - \ln \sqrt{2\pi(N+m)}$$

and with some algebraic reshuffling we obtain:

$$\ln P(m) \sim -\frac{1}{2}\ln(N\pi) - N\left[\left(1 - \frac{m}{N}\right)\ln\left(1 - \frac{m}{N}\right) + \left(1 + \frac{m}{N}\right)\ln\left(1 + \frac{m}{N}\right) \right] -$$
$$- \frac{1}{2}\left[\ln\left(1 - \frac{m}{N}\right) + \ln\left(1 + \frac{m}{N}\right) \right]$$

(note that $\ln P(m)$ is even in m, as we could have expected). If we now suppose that $m \ll N$, and the previous computation showed that this is indeed the case, then:

$$\ln\left(1 \pm \frac{m}{N}\right) \sim \pm \frac{m}{N} - \frac{1}{2}\frac{m^2}{N^2} + \cdots$$

and plugging this approximation to the second order in $\ln P(m)$ we get:

$$\ln P(m) = -\frac{1}{2}\ln(N\pi) - \frac{m^2}{N} + O(m^4)$$

and exponentiating:

$$P(m) = \frac{1}{\sqrt{\pi N}} e^{-\frac{m^2}{N}}$$

Therefore, we learn the interesting fact that for macroscopic systems if $m \ll N$ the probability to find m particles in excess or lack in the two halves of the system is distributed along a Gaussian with $\sigma_N = \sqrt{N/2}$; we have therefore found the same result as before, since the relative fluctuation is again of the order $N^{-1/2}$.

3.2.1 The monoatomic ideal gas

We now choose a specific system in order to do some explicit computations: the *monoatomic ideal gas*. This is a system composed of free point particles that don't interact[4] and therefore has the following Hamiltonian:

$$\mathcal{H}(\mathbb{P}) = \sum_{i=1}^{N} \frac{\vec{p_i}^2}{2m}$$

The volume occupied by the system in phase space is thus:

$$\Omega(E, V, N) = \underbrace{\int_{\vec{q_i} \in V} \prod_{i=1}^{N} d\vec{q_i}}_{V^N} \int \prod_{i=1}^{N} d\vec{p_i} \delta \left(\sum_{i=1}^{N} \frac{\vec{p_i}^2}{2m} - E \right) = V^N \int \prod_{i=1}^{N} d\vec{p_i} 2m \cdot \delta \left(\sum_{i=1}^{N} \vec{p_i}^2 - 2mE \right)$$

where we have used the known property of the Dirac δ function:

$$\delta(\lambda x) = \frac{\delta(x)}{|\lambda|} \qquad \lambda \in \mathbb{R}$$

Now, the sum $\sum_i \vec{p_i}^2$ is the square of the distance of the point $\mathbb{P}$ from the origin in momentum space; this means that the δ selects a spherical surface of radius $R = \sqrt{2mE}$ in momentum space. Therefore:

$$\Omega(E, V, N) = V^N 2m \cdot \hat{\Omega}_{3N}(2mE)$$

where we have denoted with $\hat{\Omega}_\ell(R^2)$ the surface of an ℓ-dimensional hypersphere of radius R, namely:

$$\hat{\Omega}_{3N}(R^2) = \int \prod_{i=1}^{N} d\vec{p_i} \delta \left(\sum_{i=1}^{N} \vec{p_i}^2 - R^2 \right)$$

Let us define in general:

$$\hat{\Omega}_\ell(R^2) = \int \delta(\vec{p}^2 - R^2) d^\ell p$$

where $\vec{p}$ is an ℓ-dimensional vector. Using the properties of the δ function we can write:

$$\hat{\Omega}_\ell(R^2) = \int \frac{\delta(|\vec{p}| - R) + \delta(|\vec{p}| + R)}{2|\vec{p}|} d\vec{p} = \int \frac{\delta(|\vec{p}| - R)}{2R} d\vec{p}$$

where we have neglected the contribution of $\delta(|\vec{p}| + R)$ since $|\vec{p}| > 0$. Now, since (as it can be shown in distribution theory) the Dirac δ is the derivative of the Heaviside step function Θ, we have:

$$\hat{\Omega}_\ell(R^2) = \frac{1}{2R} \frac{d}{dR} \underbrace{\int \Theta(R - |\vec{p}|) d\vec{p}}_{V_\ell(R)}$$

and $V_\ell(R)$ is the volume of an ℓ-dimensional hypersphere; its computation is shown in appendix A, and the result is given by equation (A.2). In the end, we have:

$$\hat{\Omega}_\ell(R^2) = \frac{1}{2R} \frac{\ell R^{\ell-1} \pi^{\ell/2}}{\Gamma\left(\frac{\ell}{2} + 1\right)}$$

---

[4]It could be legitimately argued that strictly speaking such a system will never reach an equilibrium: since the particles don't interact nor collide their velocities will always be equal to the initial ones, whatever they were; for example if in the initial state the kinetic energy of the system is held by only one particle (all the others being still), that particle will always be the only one moving and all the remaining ones will always stay still. We can solve this problem in two ways: we can either suppose that some interaction between the particle exists and such that it is sufficient to establish an equilibrium but weak enough to be neglected, or we can suppose that the particles exchange energy with the walls of the system through the collisions and that the walls "give back" this energy to the other particles when they impact them.

where Γ is Euler's gamma function[5]. Therefore:

$$\Omega(E, V, N) = V^N 2m \frac{1}{2\sqrt{2mE}} \frac{3N \left(\sqrt{2mE}\right)^{3N-1} \pi^{3N/2}}{\Gamma(\frac{3}{2}N + 1)}$$

which, after some reshuffling, can be rewritten as:

$$\Omega(E, V, N) = \frac{3N}{2E} V^N \frac{(2\pi mE)^{\frac{3}{2}N}}{\Gamma\left(\frac{3}{2}N + 1\right)} \tag{3.1}$$

If we now compute the probability density in configuration space, i.e. we integrate ρ over the momenta, we get:

$$\rho(\mathbb{Q}) = \int d\mathbb{P}\rho(\mathbb{Q}, \mathbb{P}) = \int d\mathbb{P}\frac{\delta(\mathcal{H} - E)}{\Omega(E, V, N)} = \int d\mathbb{P}\frac{1}{\Omega(E, V, N)}\delta\left(\frac{\mathbb{P}^2}{2m} - E\right) = \frac{1}{V^N}$$

which is intuitively clear since no forces act on the particles and thus they can be found in any point of our system with equal probability.

If we now think of our system as a cubic box and mentally divide it into a right and a left half we will have that the probability to find the i-th particle in one of these two halves is $1/2$:

$$p(\vec{q}_i \in V/2) = \int_{\vec{q}_i \in V/2} d\vec{q}_i \rho(\mathbb{Q}) = \frac{1}{V^N} V^{N-1} \frac{V}{2} = \frac{1}{2}$$

and so the probability that *all* the particles are in the same half of the system is:

$$p(\mathbb{Q} \in V/2) = \frac{1}{2^N}$$

This is exactly what we have already seen in 3.2.

3.2.2 Statistics and thermodynamics

The fundamental postulate of statistical mechanics

We still have not established a link between the ensemble formalism that we are developing and the thermodynamics of a system. This is what we are now going to do.

From thermodynamics we know (see 1.1.2 and 1.1.3) that all the properties of a system can be obtained from its entropy through appropriate differentiations; we would therefore like to define the entropy of a system within the microcanonical ensemble.

However, we have no clues on what we can do in order to define it; of course we must link S with some property of the microcanonical ensemble of our system, but we have nothing that can suggest us what we might use.

After all the only really new concept that we have introduced with the ensemble formalism is the phase space volume Ω, so we can think that this is what must be related to the entropy.

---

[5]Remember that it is defined as:

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$$

and if $n \in \mathbb{Z}$, then $\Gamma(n) = (n-1)!$.

We can understand that since it looks like we are at a dead end we have to "artificially" introduce something, i.e. we *must* do some assumptions in order to proceed.

The assumption we make is the *fundamental postulate of statistical mechanics*:

$$S(E, V, N) = k_B \ln \Omega(E, V, N) \tag{3.2}$$

where k_B is Boltzmann's constant, needed to give the entropy S the right dimensions[6].

Unfortunately there is no intuitive way to justify this postulate (in general we could have supposed that S was proportional to some power of Ω, or some other function different from the logarithm), it is something we have to take as it is.

However this doesn't mean that we must "religiously" belive in it; it is true that being a postulate we can't really "prove" equation (3.2), but we certainly can try to see if it is "reasonable". In other words what we would like to do now is to verify if S as defined in the fundamental postulate of statistical mechanics is *really* the entropy of a system; this means that we want to see, even if only qualitatively, if the consequences of (3.2) do agree with what we know about the thermodynamics of a system or not.

In particular we have seen in 1.1.2 that in order to obtain all the thermodynamic information we need about a macroscopic system we just need to take appropriate derivatives of the entropy:

$$\frac{1}{T} = \frac{\partial S}{\partial E}_{|V,N} \qquad\qquad \frac{P}{T} = \frac{\partial S}{\partial V}_{|E,N} \qquad\qquad \frac{\mu}{T} = -\frac{\partial S}{\partial N}_{|E,V} \tag{3.3}$$

These relations are equivalent to those that we have actually seen which involve the derivatives of the energy E:

$$T = \frac{\partial E}{\partial S}_{|V,N} \qquad\qquad P = -\frac{\partial E}{\partial V}_{|S,N} \qquad\qquad \mu = \frac{\partial E}{\partial N}_{|S,V}$$

In fact, similarly to what we have seen in 1.1.5 we have that:

$$\frac{\partial S}{\partial V}_{|E,N} \frac{\partial V}{\partial E}_{|S,N} \frac{\partial E}{\partial S}_{|V,N} = -1 \qquad\qquad \frac{\partial S}{\partial N}_{|E,V} \frac{\partial N}{\partial E}_{|S,V} \frac{\partial E}{\partial S}_{|NV} = -1$$

and thus:

$$P = T\frac{\partial S}{\partial V} = -T\frac{1}{\frac{\partial V}{\partial E}\frac{\partial E}{\partial S}} = -T\frac{1}{\frac{\partial V}{\partial E}T} = -\frac{\partial E}{\partial V}$$

$$\mu = -T\frac{\partial S}{\partial N} = T\frac{1}{\frac{\partial N}{\partial E}\frac{\partial E}{\partial S}} = T\frac{1}{\frac{\partial N}{\partial E}T} = \frac{\partial E}{\partial N}$$

Therefore what we are going to do in the following is to see if the relations written in (3.3) follow from (3.2).

Before proceeding let us make an observation that will be useful for the computations.

Since $\ln \Omega \propto S$ and S is extensive, the quantity $\ln \Omega$ is itself extensive and thus we can always write it as[7]:

$$\ln \Omega(E, V, N) = Nf\left(\frac{E}{N}, \frac{V}{N}\right) \tag{3.4}$$

where f is a generic function.

---

[6]In general, we should have used a generic constant k, but at the end of the computations that allow to rederive the thermodynamics of the system we would find out that k is precisely k_B, so we use k_B from the beginning just for the sake of simplicity.

[7]In fact, since $\ln \Omega$ is some extensive function f then factorizing an N we have:

$$\ln \Omega(E, V, N) = f(E, V, N) = f\left(N\frac{E}{N}, N\frac{V}{N}, N\frac{N}{N}\right) = Nf\left(\frac{E}{N}, \frac{V}{N}, 1\right) := Nf\left(\frac{E}{N}, \frac{V}{N}\right)$$
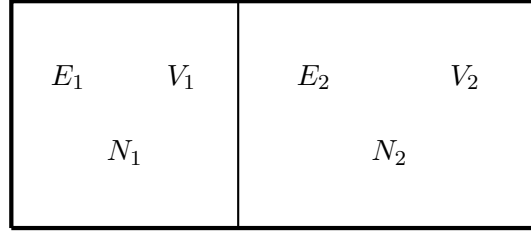
Figure 3.1: Two systems in thermal equilibrium

Temperature and statistical mechanics

Let us now see how the temperature of a system comes into play within the microcanonical ensemble.

Consider two systems, which we call 1 and 2, of volume V_i, energy E_i, at temperature T_i and each composed of N_i particles, with $i = 1, 2$ (see figure 3.1), separated by a fixed wall that allows only the exchange of energy between the two systems. We call $E = E_1 + E_2$ the total energy and $N = N_1 + N_2$ the total number of particles, and we know from thermodynamics that if initially $T_1 \neq T_2$ then after some time the two systems reach an equilibrium and have the same temperature T.

Now, it is intuitively clear[8] that the phase space volume of the whole system is:

$$\Omega(E) = \int \Omega_1(E_1)\Omega_2(E - E_1)dE_1$$

In fact the interpretation of this expression is the following: for a fixed value of E_1 the total number of microstates of the whole system is $\Omega_1(E_1)\Omega_2(E - E_1)$ (remember that $E_2 = E - E_1$), since for *any* of the $\Omega_1(E_1)$ possible microstates of the system 1 the system 2 can be in *any* of its $\Omega_2(E - E_1)$ possible states; therefore the total number of the possible states of the whole system can be obtained integrating over all the possible values of E_1.

Now, since $\ln \Omega$ is extensive and can be written as in equation (3.4), we have (neglecting the dependence of f on the volume, since it is fixed):

$$\Omega(E) = \int dE_1 e^{N_1 f_1\left(\frac{E_1}{N_1}\right) + N_2 f_2\left(\frac{E - E_1}{N - N_1}\right)}$$

and defining $\varepsilon_i = E_i/N_i$, $\varepsilon = E/N$ and $n_i = N_i/N$:

$$\Omega(E) = N_1 \int d\varepsilon_1 e^{N[n_1 f_1(\varepsilon_1) + n_2 f_2(\varepsilon - \varepsilon_1)]}$$

---

[8]If this is not the case we can anyway easily obtain this result in a more "formal" way. In fact, we have:

$$\Omega(E) = \int d\Gamma \delta(\mathcal{H} - E) = \int d\Gamma_1 d\Gamma_2 \delta(\mathcal{H}_1 + \mathcal{H}_2 - E) = \int d\Gamma_1 \Omega_2(E - \mathcal{H}_1)$$

where we have integrated over $d\Gamma_2$ in the last step. Now, using a "trick":

$$\Omega(E) = \int dE_1 \int d\Gamma_1 \Omega_2(E - H_1)\delta(\mathcal{H}_1 - E_1) = \int dE_1 \int d\Gamma_1 \Omega_2(E - E_1)\delta(\mathcal{H}_1 - E) =$$
$$= \int dE_1 \Omega_2(E - E_1) \int d\Gamma_1 \delta(\mathcal{H}_1 - E_1) = \int dE_1 \Omega_2(E - E_1)\Omega_1(E_1)$$

Now, this integral can be approximated using the *saddle point* approximation (see appendix B). The result is[9]:

$$\Omega(E) \approx N_1 \frac{e^{N\left[n_1 f_1(\varepsilon_1^*) + n_2 f_2(\varepsilon - \varepsilon_1^*)\right]}}{\sqrt{2\pi N \left|\left[n_1 f_1''(\varepsilon_1^*) + n_2 f_2''(\varepsilon - \varepsilon_1^*)\right]\right|}}$$

where ε_1^* is the value of ε_1 that maximizes the integrand of $\Omega(E)$, i.e. the exponent $n_1 f_1(\varepsilon_1) + n_2 f_2(\varepsilon - \varepsilon_1)$. We therefore have:

$$n_1 f_1'(\varepsilon_1^*) - n_2 f_2'(\varepsilon - \varepsilon_1^*) = 0$$

which means (remember equation (3.4)):

$$\frac{\partial}{\partial E_1} \ln \Omega_1(E_1)_{|E_1^*} = \frac{\partial}{\partial E_2} \ln \Omega_2(E_2)_{|E_2^*} \tag{3.5}$$

We have therefore found that from the formulation of the problem in the microcanonical ensemble, at thermal equilibrium the only microscopic configurations that significantly contribute to $\Omega(E)$ are those with $E_1 = E_1^*$ and $E_2 = E_2^*$, which are such that equation (3.5) holds.
From the fundamental postulate of statistical mechanics (3.2), equation (3.5) becomes:

$$\frac{\partial S}{\partial E_1}_{|E_1^*} = \frac{\partial S}{\partial E_2}_{|E_2^*}$$

However from thermodynamics we know that $\partial S / \partial E = 1/T$, so in the end we have:

$$T_1 = T_2$$

Thus from the microcanonical definition of entropy we have obtained the well known fact that when two bodies are put in thermal contact they reach an equilibrium state were both have the same temperature[10].
Now, in order to explicitly see that the main contribution to $\Omega(E) = \int \Omega_1(E_1)\Omega_2(E - E_1)dE_1$ comes from the configuration where $E_1 = E_1^*$, let us expand in a Taylor series the integrand around $E_1 = E_1^*$ (and we write explicitly $f_1''(\varepsilon_1^*) = -|f_1''(\varepsilon_1^*)|$ and $f_2''(\varepsilon - \varepsilon_1^*) = -|f_2''(\varepsilon - \varepsilon_1^*)|$ because the second derivatives of f_1 and f_2 are negative in ε_1^* and $\varepsilon - \varepsilon_1^*$, since they are maxima):

$$\Omega_1(E_1)\Omega_2(E - E_1) = e^{N\left[n_1 f_1(\varepsilon_1) + n_2 f_2(\varepsilon - \varepsilon_1)\right]} \approx$$

$$\approx e^{N\left[n_1 f_1(\varepsilon_1^*) + n_2 f_2(\varepsilon - \varepsilon_1^*)\right]} e^{-N\left[n_1 |f_1''(\varepsilon_1^*)| + n_2 |f_2''(\varepsilon - \varepsilon_1^*)|\right](\varepsilon - \varepsilon_1^*)^2} =$$

$$= \text{const.} \cdot e^{-N\left[n_1 |f_1''(\varepsilon_1^*)| + n_2 |f_2''(\varepsilon - \varepsilon_1^*)|\right](\varepsilon - \varepsilon_1^*)^2} = \text{const.} \cdot e^{-\left[n_1 |f_1''(\varepsilon_1^*)| + n_2 |f_2''(\varepsilon - \varepsilon_1^*)|\right]\frac{1}{N} \cdot (E - E_1^*)^2}$$

where we have used the definitions of ε_1 and ε. This is a Gaussian with variance:

$$\sigma_{E_1}^2 := \left\langle (E - E_1^*)^2 \right\rangle \propto N$$

---

[9]The derivatives are intended to be taken with respect to the argument of the function:

$$f_1'(\varepsilon_1) = \frac{\partial}{\partial \varepsilon_1} f_1(\varepsilon_1) \qquad\qquad f_2'(\varepsilon - \varepsilon_1) = \frac{\partial}{\partial(\varepsilon - \varepsilon_1)} f_2(\varepsilon - \varepsilon_1)$$

[10]Alternatively we could have used a different approach to come to the same result. In fact from our analysis we have that two systems at thermal equilibrium in the microcanonical ensemble are such that $\partial S / \partial E$ has the same value for both, and from thermodynamics we know that two systems at thermal equilibrium share the same temperature; thus $\partial S / \partial E$ must be related to the temperature of a system, and since it has the dimensions of the inverse of a temperature we can *define* the temperature in the microcanonical ensemble as:

$$\frac{1}{T} = \frac{\partial S}{\partial E} = k_B \frac{\partial}{\partial E} \ln \Omega(E, V, N)$$

Also in this case we get to the result $T_1 = T_2$.
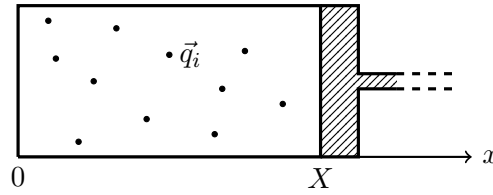
Figure 3.2: System with variable volume

Thus, the relative fluctuation of E_1 with respect to E_1^* is:

$$\frac{\sigma_{E_1}}{E_1} \propto \frac{1}{\sqrt{N}}$$

Again these fluctuations are absolutely negligible in the thermodynamic limit, as we expected, and so indeed E_1^* is the only value of E that significantly contributes to $\Omega(E)$ (in other words, for large N it becomes a very sharply peaked function).

Pressure and statistical mechanics

We now want to see, just like we have done with the temperature, what role does the pressure of a system play in the microcanonical ensemble.
Let us therefore consider a system on which we can act using a piston of cross section A (see figure 3.2). If the pressure of the gas is P, in order to maintain the system in equilibrium we must exert a force $F = PA$ on the piston.
If the gas inside the box is ideal the Hamiltonian of the system can be written as:

$$\mathcal{H}(\mathbb{Q}, \mathbb{P}, X) = E_{\text{kin}} + \sum_{i=1}^{N} u(q_{ix} - X)$$

where E_{kin} is the kinetic energy of the particles and u is the potential that models the presence of the piston, so it will be a function with very small values for $q_{ix} < X$ and that diverges for $q_{ix} \to X^-$. Therefore, the i-th particle will be subjected to the force (which is of course directed along the x direction):

$$F_i = -\frac{\partial}{\partial q_{ix}} u(q_{ix} - X)$$

and the (microscopic and istantaneous) force acting on the piston will be $\sum_i F_i$; therefore the force done by the piston on the system is such that:

$$\left\langle -\sum_i F_i \right\rangle = PA \tag{3.6}$$

We also have:

$$F_i = -\frac{\partial}{\partial q_{ix}} u(q_{ix} - X) = \frac{\partial}{\partial X} u(q_{ix} - X)$$

from which we get:

$$\sum_i F_i = \sum_i \frac{\partial u}{\partial X} = \frac{\partial \mathcal{H}}{\partial X}$$

Therefore, equation (3.6) becomes:

$$PA = \left\langle -\frac{\partial \mathcal{H}}{\partial X} \right\rangle$$
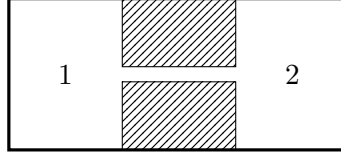
Figure 3.3: System with variable number of particles

and since $V = XA$ we have:

$$P = -\left\langle \frac{\partial \mathcal{H}}{\partial V} \right\rangle$$

Now, by definition:

$$\left\langle \frac{\partial \mathcal{H}}{\partial V} \right\rangle = \frac{1}{\Omega(E, V, N)} \int d\Gamma \frac{\partial \mathcal{H}}{\partial V} \delta(\mathcal{H} - E) \qquad\qquad \Omega(E, V, N) = \int d\Gamma \delta(\mathcal{H} - E)$$

and therefore:

$$\frac{\partial \Omega}{\partial V} = \int d\Gamma \delta'(\mathcal{H} - E) \frac{\partial \mathcal{H}}{\partial V}$$

where with $\delta'(\mathcal{H} - E)$ we mean the derivative of $\delta(\mathcal{H} - E)$ with respect to $\mathcal{H} - E$. Thus, $\delta'(\mathcal{H} - E) = -\partial \delta(\mathcal{H} - E)/\partial E$ and since E is a parameter (it's fixed):

$$\frac{\partial \Omega}{\partial V} = -\frac{\partial}{\partial E} \underbrace{\int d\Gamma \frac{\partial \mathcal{H}}{\partial V} \delta(\mathcal{H} - E)}_{\Omega\langle \partial \mathcal{H}/\partial V \rangle = -\Omega P} = \frac{\partial}{\partial E}(\Omega P) = \Omega \frac{\partial P}{\partial E} + \frac{\partial \Omega}{\partial E} P$$

Dividing both sides of the equation by Ω we get:

$$\frac{\partial}{\partial V} \ln \Omega = \frac{\partial P}{\partial E} + P \frac{\partial}{\partial E} \ln \Omega$$

and considering the definition (3.2) of entropy:

$$\frac{\partial S}{\partial V} = k_B \frac{\partial P}{\partial E} + P \frac{\partial S}{\partial E} = k_B \frac{\partial P}{\partial E} + \frac{P}{T}$$

We know that S, V and E are extensive quantities while P and T are intensive; therefore $\partial S/\partial V$ and P/T are intensive while $\partial P/\partial E \propto N^{-1}$. This means that in the thermodynamic limit $\partial P/\partial E$ vanishes, and thus:

$$\frac{P}{T} = \frac{\partial S}{\partial V}$$

which is exactly what we were looking for (see (3.3)).

Chemical potential and statistical mechanics

We conclude with the chemical potential.

Let us therefore consider a system like that represented in figure 3.3, namely a system divided into two subsystems 1 and 2 connected by a hole that allows the exchange of particles. Its Hamiltonian will be:

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$$

where:

$$\mathcal{H}_i = \sum_{j=1}^{N} \frac{\vec{p}_j^{(i)2}}{2m} + \frac{1}{2} \sum_{j \neq k} \mathcal{V}\left(\vec{q}_j^{(i)} - \vec{q}_k^{(i)} \right) \qquad\qquad i = 1, 2$$

and $\mathcal{V}$ is the interaction potential.

If we suppose that the hole that connects the two subsystems is long enough they will not interact with each other; calling N the total number of particles we will have that:

$$\Omega(E, V, N) = \int d\Gamma_1 d\Gamma_2 \delta(\mathcal{H}_1 + \mathcal{H}_2 - E) = \sum_{N_1=0}^{N} \int_0^E dE_1 \Omega_1(E_1, N_1) \Omega_2(E - E_1, N - N_1)$$

where we are summing over N_1 because the possible configurations of the system now include the number of particles that are in the two subsystems.

If we now divide $[0, E]$ in very small intervals of width Δ, we can write:

$$\int_0^E dE_1 f(E_1) \approx \Delta \sum_{i=1}^{E/\Delta} f(i\Delta)$$

with f a generic function, so that we have:

$$\Omega(E, N) = \Delta \sum_{N_1=0}^{N} \sum_{i=1}^{E/\Delta} \Omega_1(i\Delta, N_1) \Omega_2(E - i\Delta, N - N_1)$$

If we now call $E_1^* = i^*\Delta$ and N_1^* the values of E_1 and N_1 for which the summand has the largest value, we will surely have:

$$\Delta \cdot \Omega_1(E_1^*, N_1^*) \Omega_2(E - E_1^*, N - N_1^*) < \Omega(E, N) < (N+1)\frac{E}{\Delta} \cdot \Delta \cdot \Omega_1(E_1^*, N_1^*) \Omega_2(E - E_1^*, N - N_1^*)$$

(namely the largest term is smaller than the whole sum which is in turn smaller than the sum where every term has been substituted with the largest one).

Therefore, taking the logarithm and multiplying by k_B:

$$k_B \ln \Delta + S_1^* + S_2^* < S < S_1^* + S_2^* + k_B \ln(N+1) + k_B \ln E$$

and since the entropy is extensive in the thermodynamic limit we get:

$$S(E, N) = S_1(E_1^*, N_1^*) + S_2(E - E_1^*, N - N_1^*) + O(\ln N)$$

and we can also neglect the last term.

Using the microcanonical definition of entropy (3.2), since E_1^* and N_1^* are the values of E_1 and N_1 that maximize S we will have:

$$\frac{\partial}{\partial E_1} [S_1(E_1, N_1) + S_2(E - E_1, N - N_1)]_{|E_1^*, N_1^*} = 0$$

$$\frac{\partial}{\partial N_1} [S_1(E_1, N_1) + S_2(E - E_1, N - N_1)]_{|E_1^*, N_1^*} = 0$$

We have already previously encountered the first equation, which has led us to $T_1 = T_2$. Focusing now on the second one, this leads to:

$$\frac{\partial}{\partial N_1} S_1(E_1, N_1)_{|E_1^*, N_1^*} = \frac{\partial}{\partial N_2} S_2(E_2, N_2)\Big|_{\substack{E_2^* = E - E_1^* \\ N_2^* = N - N_1^*}}$$

Therefore if two systems can exchange particles they will have not only the same temperature but also the same value of $\partial S/\partial N$. From thermodynamics we know that this must be related to the chemical potential μ of the system; in particular since $[\mu] = \text{J}$ and $[\partial S/\partial N] = \text{J/K}$, we will have:

$$\frac{\partial S}{\partial N} = -\frac{\mu}{T}$$

where however the minus sign cannot be predicted[11].

---

[11]Remember that this, in the end, is only a qualitative way to show that the fundamental postulate of statistical mechanics is reasonable and leads to results compatible to what we know about the thermodynamics of a system.

Conclusions

Therefore with these qualitative reasonings we have shown that the fundamental postulate of statistical mechanics, i.e. equation (3.2), leads to what we expect about the thermodynamics of a system (namely equation (3.3)).
We can therefore conclude that it is indeed a reasonable assumption.

3.2.3   Some remarks and refinements in the definition of entropy

As we have given them, the definitions of the phase space volume Ω and entropy S are not correct and can give rise to annoying problems and paradoxes.

The dimension of the phase space volume

To begin with we can note that the definition of entropy given in equation (3.2), the fundamental postulate of statistical mechanics, makes no sense since Ω is dimensional. In fact Ω is defined as an integral over $d\Gamma$ and:

$$d\Gamma = \prod_{i=1}^{N} d\vec{q}_i d\vec{p}_i \qquad \Rightarrow \qquad [d\Gamma] = \left(m \cdot \frac{\text{kg m}}{\text{s}} \right)^{3N} = [\Omega]$$

Therefore, if we want to take the logarithm of Ω we must first adimensionalize it; to do so we must introduce a constant h with dimensions $m \cdot (\text{kg m/s}) = J \cdot s$, so that we can actually redefine the phase space volume as:

$$\Omega(E, V, N) = \int \frac{d\Gamma}{h^{3N}} \delta(\mathcal{H} - E)$$

which is now really adimensional and now equation (3.2) makes sense.

A legitimate question could now be: what is the value of h? Which constant is it?
Unfortunately, within classical statistical mechanics it is impossible to establish it; only in quantum statistical mechanics the constant h acquires a precise meaning, and in particular it turns out that h is *Planck's constant*:

$$h = 6.62 \cdot 10^{-34} \text{ J s}$$

Let us finally note that dividing Ω by h^{3N} and integrating over the phase space can also be interpreted as dividing the whole phase space in cells of volume h^3, thus "discretizing" the possible states of the system considering as equivalent all the points inside a cell[12], and summing over these cells[13]: in this sense the entropy gives a measure of the quantity of such possible states.

Extensivity of the entropy and the Gibbs paradox

Even with the introduction of h we still have some problems. In particular, from the expression (3.1) of $\Omega(E, V, N)$ for an ideal gas we have:

$$\ln \Omega(E, V, N) = \ln \left(\frac{3N}{2E} \right) - 3N \ln h + N \ln V + \frac{3}{2} \ln \pi + \frac{3}{2} N \ln(2mE) - \frac{3}{2} \ln \left(\frac{3}{2} N \right) + \frac{3}{2} N$$

---

[12]In other words, we consider all the points inside a cell to represent only a single state. This is just an "approximation" in classical statistical mechanics, but in quantum statistical mechanics it turns out that there *really* is only one state in every of those cell in phase space.

[13]To be a bit more precise: the error that we make approximating the integral over the whole phase space with a sum over cells of linear dimension h is perfectly negligible. This is due to the fact that h is ridiculously small with respect to the common scales of a macroscopic system.

Since E and N are both extensive, E/N is intensive and thus remains constant in the thermodynamic limit; this means that for large N we can neglect the first term[14] so that in the end:

$$\ln \Omega(E, V, N) = N \ln \left[V \left(\frac{4\pi m}{3h^2} \frac{E}{N} \right)^{3/2} \right] + \frac{3}{2} N$$

and therefore the entropy of the system will be:

$$S(E, V, N) = k_B N \ln \left[V \left(\frac{4\pi m}{3h^2} \frac{E}{N} \right)^{3/2} \right] + \frac{3}{2} k_B N \tag{3.7}$$

This expression is rather problematic. In fact, first of all we can note that written in this way it is *not* extensive since it contains terms like $N \ln V$ and $N \ln N$. Furthermore, (3.7) gives rise to a catastrophic paradox known as the *Gibbs paradox*.

In order to understand it let us consider a system divided by a wall into two subsystems each of N_i particles of mass m_i, volume V_i and energy E_i (with $i = 1, 2$). The entropy of the system in this initial state is:

$$S_i = \frac{3}{2} k_B (N_1 + N_2) + k_B \left\{ N_1 \ln \left[V_1 \left(\frac{4\pi m_1}{h^2} \frac{E_1}{N_1} \right)^{3/2} \right] + N_2 \ln \left[V_2 \left(\frac{4\pi m_2}{h^2} \frac{E_2}{N_2} \right)^{3/2} \right] \right\}$$

If we now remove the wall the two gases will mix and the entropy of this final state of the system will be:

$$S_f = \frac{3}{2} k_B (N_1 + N_2) + k_B \ln \left[(V_1 + V_2)^{N_1 + N_2} \left(\frac{4\pi}{3h^2} \right)^{\frac{3}{2}(N_1 + N_2)} \left(\frac{m_1 E_1}{N_1} \right)^{\frac{3}{2} N_1} \left(\frac{m_2 E_2}{N_2} \right)^{\frac{3}{2} N_2} \right]$$

Therefore, the entropy of the system will have changed during this process by the amount:

$$\Delta S = S_f - S_i = k_B \ln \left[\frac{(V_1 + V_2)^{N_1 + N_2}}{V_1^{N_1} V_2^{N_2}} \right] = k_B \left[N_1 \ln \left(\frac{V_1 + V_2}{V_1} \right) + N_2 \ln \left(\frac{V_1 + V_2}{V_2} \right) \right]$$

called *entropy of mixing*, which is always positive. If $m_1 \neq m_2$ this is a correct result, since the mixing of the two gases is an irreversible process.

However, since ΔS doesn't depend on m_1 and m_2 this result holds also in the case $m_1 = m_2$. But this is a paradox: the mixing of two identical gases is a reversible process (we can recover the initial state reinserting the wall) so the entropy of the whole system shouldn't increase if we remove the wall. Furthermore, the fact that $\Delta S > 0$ also when $m_1 = m_2$ is catastrophic because it means that the entropy of a system depends on the *history* of the system itself, rather than on its *state*. But this ultimately means that entropy *doesn't exist at all*: consider a system of energy E made of N particles contained in a volume V; then we can think that this system has been obtained from the union of M pre-existing subsystems, *with M arbitrarily large*. This means that the entropy of the system has increased an arbitrarily large amount of times from its initial value, and therefore the entropy of the system in its final configuration (energy E, volume V, N particles) is greater than any arbitrary number: in other words, assuming (3.7) as the entropy of an ideal gas we would conclude that the entropy of *any* ideal gas is infinite! There's clearly something wrong with the definition of entropy we have given. How can we

---

[14]Note that this means that in (3.1) we can simply drop the term $3N/2E$ (which is what we are going to do in the future), since it divides the phase space volume by a negligible amount. In fact, from the fundamental postulate of statistical mechanics (3.2) we have that multiplying Ω by an intensive factor is equivalent to adding a constant to the entropy; however the entropy of a typical system is so large that adding such a constant doesn't change sensibly its value. In other words the phase space volume Ω is so large that *multiplying* it by a constant does not change significantly its value.

solve this problem?

We know that in classical mechanics identical particles are distinguishable; however in order to solve the Gibbs paradox we must treat them as undistinguishable, just like we would do in quantum mechanics: this way if we exchange two particles the representative point of the system in phase space won't change. Now, since N particles can be exchanged in $N!$ different ways there will be $N!$ different configurations of the system relative to the same representative point. This means that we must redefine the phase space volume of the system reducing it by a factor $N!$ (which is sometimes called *Boltzmann factor*):

$$\Omega(E, V, N) = \frac{1}{N!} \int \frac{d\Gamma}{h^{3N}} \delta(\mathcal{H} - E)$$

In this way we can solve all the problems we have described. In fact, using this definition of the phase space volume in the case of the ideal gas we have:

$$\Omega(E, V, N) = \frac{V^N}{N! h^{3N}} \frac{(2\pi m E)^{3N/2}}{\Gamma\left(\frac{3}{2}N + 1\right)} \qquad\qquad S(E, V, N) = \frac{5}{2} N k_B + N k_B \ln\left[\frac{V}{N}\left(\frac{4\pi m}{3h^2}\frac{E}{N}\right)^{3/2}\right]$$

This expression for the entropy is clearly extensive since it is proportional to N and the logarithm depends only on V/N and E/N. Furthermore, the computation of the entropy of mixing of two gases now gives:

$$\Delta S = k_B \left[(N_1 + N_2) \ln\left(\frac{V_1 + V_2}{N_1 + N_2}\right) - N_1 \ln\frac{V_1}{N_1} - N_2 \ln\frac{V_2}{N_2}\right]$$

If the two gases are different, their densities N_i/V_i will be different and so $\Delta S > 0$ as it must be. However, if the two gases are identical this time we have $V_1/N_1 = V_2/N_2 = V/N$ and the entropy of mixing vanishes.

This solution to the Gibbs paradox, however, is rather an *ad hoc* one. Unfortunately there's no way to understand where does the Boltzmann factor *really* come from within the framework of *classical* statistical mechanics. This can be made clearer within quantum statistical mechanics: in that case the $N!$ comes from the fact that identical particles are intrinsically indistinguishable and does not "disappear" in the classical limit.

Conclusions

To conclude, the correct definition of the phase space volume that eliminates all the problems that we have mentioned, and the one we will *always* use in the future, is:

$$\Omega(E, V, N) = \frac{1}{N!} \int \frac{d\Gamma}{h^{3N}} \delta(\mathcal{H} - E)$$

For the sake of simplicity we now redefine $d\Gamma$ as $d\Gamma_{\text{new}} = d\Gamma_{\text{old}}/N! h^{3N}$ (in other words we "incorporate" the $1/N! h^{3N}$ factor inside $d\Gamma$). This way, for the ideal gas we have:

$$\Omega(E, V, N) = \frac{V^N}{N! h^{3N}} \frac{(2\pi m E)^{3N/2}}{\Gamma\left(\frac{3}{2}N + 1\right)}$$

3.2.4   Some explicit computations for the monoatomic ideal gas

Now that we have fixed everything, in order to understand if what we have done until now makes sense let us use all this machinery to see what happens when we apply it to the monoatomic

ideal gas (which is the simplest system we can think of).
From what we have just seen above the phase space volume of the system is:

$$\Omega(E, V, N) = \frac{V^N}{N! h^{3N}} \frac{(2\pi m E)^{3N/2}}{\Gamma\left(\frac{3}{2}N + 1\right)}$$

and the entropy can be rewritten more shortly as:

$$S(E, V, N) = N k_B \left[\frac{5}{2} - \ln(\rho\lambda^3)\right] \tag{3.8}$$

where:

$$\rho = \frac{N}{V} \qquad\qquad \lambda = \sqrt{\frac{3h^2}{4\pi m} \cdot \frac{N}{E}}$$

Therefore, we can now compute:

$$\frac{1}{T} = \frac{\partial S}{\partial E}\Big|_{V,N} = \frac{3}{2} N \frac{k_B}{E}$$

and thus:

$$E = \frac{3}{2} N k_B T$$

We have really obtained the expression of the energy for an ideal gas!

Now, going further:

$$\frac{P}{T} = \frac{\partial S}{\partial V}\Big|_{E,N} = \frac{N k_B}{V}$$

from which we get the ideal gas law:

$$PV = N k_B T$$

The efforts we have done are therefore not useless, since we are really obtaining all the thermodynamics of the system from the fundamental postulate (3.2)!
This further confirms that this postulate is reasonable, and again justifies its introduction.

To conclude we can also compute the chemical potential for an ideal gas, which turns out to be:

$$\mu = -T \frac{\partial S}{\partial N}\Big|_{E,V} = k_B T \ln\left[\frac{V}{N}\left(\frac{4\pi m E}{3N h^2}\right)^{3/2}\right] = -k_B T \ln(\rho\lambda^3)$$

The Maxwell-Boltzmann distribution

We could now ask: can we determine the probability distribution of the momentum, or equivalently of the velocity, of a single particle?
Since we are only interested in the probability density of the momentum of a single particle, say the j-th one, we take the probability distribution of the *whole* system in phase space and then integrate over all the coordinates we do not care about, i.e. $\vec{q}_i \quad \forall i$ and $\vec{p}_i \quad \forall i \neq j$, with the constraint that $\vec{p}_j$ has some given value $\vec{p}$; this way we are left with the probability distribution of a single particle. We therefore have:

$$\rho(\vec{p}) = \langle\delta(\vec{p}_j - \vec{p})\rangle = \frac{1}{\Omega(E, V, N)} \int d\Gamma_N \delta\left(\sum_{i=1}^{N} \frac{\vec{p}_i^{\,2}}{2m} - E\right) \delta(\vec{p}_j - \vec{p})$$

where we have written $d\Gamma_N$ to show explicitly that it is relative to all the N particles; of course $d\Gamma_N = d\Gamma_{N-1}d\vec{q}_j d\vec{p}_j$ and so integrating over $\vec{q}_j$ and $\vec{p}_j$ we get:

$$\rho(\vec{p}) = \frac{1}{\Omega(E,V,N)}V \int d\Gamma_{N-1}\delta\left[\sum_{i=1}^{N-1}\frac{\vec{p}_i{}^2}{2m} - \left(E - \frac{\vec{p}^2}{2m}\right)\right] = \frac{V}{\Omega(E,V,N)}\Omega\left(E - \frac{\vec{p}^2}{2m}, V, N - 1\right)$$

where for simplicity we have rearranged the indices i that label the particles so that the j-th one is the last (otherwise, the sum over i should have been written as $\sum_{i=1, i\neq j}^{N}$).
Substituting the expressions of Ω for the ideal gas:

$$\rho(\vec{p}) = V\frac{N!h^{3N}\left(\frac{3}{2}N\right)!}{V^N(2\pi mE)^{3N/2}} \cdot \frac{V^{N-1}\left[2\pi m\left(E - \frac{\vec{p}^2}{2m}\right)\right]^{3(N-1)/2}}{(N-1)!h^{3(N-1)}\left[\frac{3}{2}(N-1)\right]!} =$$

$$= h^3 N\frac{\left(\frac{3}{2}N\right)!}{\left[\frac{3}{2}(N-1)\right]!}\left[2\pi m\left(E - \frac{\vec{p}^2}{2m}\right)\right]^{-3/2}\left(1 - \frac{\vec{p}^2}{2mE}\right)^{3N/2}$$

Now, since N is large we can use Stirling's approximation (see appendix C) to see how:

$$A := \frac{\left(\frac{3}{2}N\right)!}{\left[\frac{3}{2}(N-1)\right]!}$$

behaves for large N. We have:

$$\ln A \sim \frac{3}{2}N\ln\left(\frac{3}{2}N\right) - \frac{3}{2}N - \frac{3}{2}(N-1)\ln\left[\frac{3}{2}(N-1)\right] + \frac{3}{2}(N-1)$$

and with some reshuffling we get to:

$$\ln A \sim \frac{3}{2}N\ln\left(\frac{N}{N-1}\right) + \frac{3}{2}\ln\frac{3}{2} + \frac{3}{2}\ln(N-1)$$

If $N \to \infty$, the first term vanishes because the argument of the logarithm tends to 1; in the end, for large N we have:

$$\ln A \sim \frac{3}{2}\ln\frac{3}{2} + \frac{3}{2}\ln(N-1) \sim \frac{3}{2}\ln\frac{3}{2} + \frac{3}{2}\ln N$$

and exponentiating:

$$A \sim \left(\frac{3}{2}\right)^{3/2}N^{3/2}$$

Therefore if we also consider that $\vec{p}^2/2m \ll E$, then:

$$\rho(\vec{p}) = h^3 N^{5/2}\left(\frac{2}{3} \cdot 2\pi mE\right)^{-3/2}\left(1 - \frac{\vec{p}^2}{2mE}\right)^{3N/2}$$

Now, the last term becomes an exponential in the thermodynamic limit; in fact, defining $\varepsilon = E/N$ we have:

$$\left(1 - \frac{\vec{p}^2}{2mE}\right)^{3N/2} = \left(1 - \frac{\vec{p}^2}{2m\varepsilon N}\right)^{3N/2} \overset{M:=3N/2}{=} \left(1 - \frac{3\vec{p}^2}{4m\varepsilon} \cdot \frac{1}{M}\right)^M \overset{M\to\infty}{\longrightarrow} e^{-\frac{3\vec{p}^2}{4m\varepsilon}}$$

Therefore:

$$\rho(\vec{p}) = h^3 N^{5/2}\left(\frac{2}{3} \cdot 2\pi mE\right)^{-3/2}e^{-\frac{3N}{4mE}\vec{p}^2}$$

Since for the ideal gas $E = 3Nk_BT/2$, we have:

$$\rho(\vec{p}) = h^3 N (2\pi m k_B T)^{-3/2} e^{-\frac{3N}{4mE}\vec{p}^2} = (2\pi m k_B T)^{-3/2} e^{-\frac{3N}{4mE}\vec{p}^2 + \ln N + 3\ln h}$$

Now, $\ln N \ll N$ for large N and h is a constant so we can neglect these last two contributions in the exponential; therefore:

$$\rho(\vec{p}) = (2\pi m k_B T)^{-3/2} e^{-\frac{\vec{p}^2}{2mk_B T}}$$

This is the *Mawxell-Boltzmann distribution* for the momentum. If we want to express it in terms of the velocity $\vec{v}$, then from the fact that[15] $\rho(\vec{v})d\vec{v} = \rho(\vec{p})d\vec{p} = m^3\rho(\vec{p})d\vec{v}$ we have:

$$\rho(\vec{v}) = \left(\frac{m}{2\pi k_B T}\right)^{3/2} e^{-\frac{m\vec{v}^2}{2k_B T}}$$

3.2.5 The foundations of statistical mechanics

Before proceeding to the study of the canonical and grand canonical ensembles, let us stop for a moment to reflect on the deep reasons that led us to the formulation of the microcanonical ensemble. This is extremely important since as we shall later see both the canonical and the grand canonical ensembles are based upon the microcanonical one, so understanding *why* the microcanonical ensemble is the way it is can help us determining the *foundations* of statistical mechanics.

In particular one could note that the way we have introduced the microcanonical ensemble is not really "philosophically" satisfying, particularly because of the a priori equal probability postulate. We could in fact ask ourselves: is this postulate realistic? In other words: are all the microstates of any given system *really* equally probable, or can we find cases where this is not true? And if they are really equally probable, *why* is it so?

The topic is very wide and complicated (in its deepest nature it is still an open problem), and since it is not essential in order to understand the rest of this chapter we postpone its study in appendix D.

3.3 The canonical ensemble

Until now we have only considered systems with a fixed value of energy, for which we have seen that the phase space volume is related to the entropy.
This kind of ensemble is however really restricting also because the computations made within it are quite complicated (in 3.2.1 we had to resort to the geometry of hyperspheres in order to compute the phase space volume of the ideal gas, which is the simplest system we can think of!). We therefore need to generalize our considerations to systems which have V and N fixed, but variable energy; in other words we want to see what happens when we remove the constraint of keeping the energy of a system fixed.
This is the essence of the *canonical ensemble*.
However, in order to determine the properties of this ensemble the only thing we can do is to use the tools that we already know, namely the microcanonical ensemble. What we are going to do therefore, in just a few words, is to consider an isolated system (for which the methods of the microcanonical ensemble apply) and see what happens if we consider a small portion of it.

---

[15]In fact, what must remain invariant is the probability to find a particle with a momentum or velocity in a given interval. In other words, if $\rho(\vec{p})dp$ is the probability of finding a particle with momentum in $[\vec{p}, \vec{p} + d\vec{p}]$, or equivalently in $[m\vec{v}, m\vec{v} + md\vec{v}]$, we see that this is equal to the probability of finding a particle with velocity in $[\vec{v}, \vec{v} + d\vec{v}]$, which is exactly $\rho(\vec{v})d\vec{v}$.
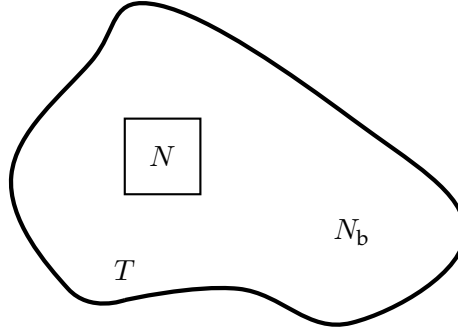
Figure 3.4: Canonical ensemble

Let us therefore consider a system like the one represented in figure 3.4: we have a macroscopic isolated system, which we call the *heat bath*, at temperature T composed of N_b particles, and we also consider a small (but still macroscopic) subsystem of the bath containing N particles (namely $N_b, N \gg 1$ but $N \ll N_b$); for now N and N_b are fixed so the "walls" that enclose the small subsystem are fixed and impermeable and only allow the exchange of energy between the two. We also call E the energy of the whole system.

Now, since the total system (the heat bath and the small subsystem) is isolated we can apply to it the tools of the microcanonical ensemble; in particular the probability density in phase space of the whole system will be:

$$\rho(\mathbb{Q}_b, \mathbb{P}_b, \mathbb{Q}, \mathbb{P}) = \frac{1}{\Omega(E)} \delta \left[\mathcal{H}_b(\mathbb{Q}_b, \mathbb{P}_b) + \mathcal{H}(\mathbb{Q}, \mathbb{P}) + \mathcal{V}(\mathbb{Q}_b, \mathbb{Q}) - E \right]$$

where $\mathcal{V}(\mathbb{Q}_b, \mathbb{Q})$ is the term containing the interaction between the heat bath and the subsystem. This term *must* exist if we want the two systems to exchange energy, but what we now suppose is that it is neglibigle with respect to the other terms (the energy E and the Hamiltonians).

Since we want to study the properties of the small subsystem we are interested in finding a probability density which does not contain any information about the heat bath; in other words we want to determine the so called *marginalized* probability density:

$$\rho(\mathbb{Q}, \mathbb{P}) = \int \rho(\mathbb{Q}_b, \mathbb{P}_b, \mathbb{Q}, \mathbb{P}) d\Gamma_b$$

namely:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{1}{\Omega(E)} \int \delta \left[\mathcal{H}_b(\mathbb{Q}_b, \mathbb{P}_b) - (E - \mathcal{H}(\mathbb{Q}, \mathbb{P})) \right] d\Gamma_b$$

Now, by definition:

$$\Omega_b(E') = \int \delta \left(\mathcal{H}_b(\mathbb{Q}_b, \mathbb{P}_b) - E' \right) d\Gamma_b$$

so we can write:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{\Omega_b(E - \mathcal{H}(\mathbb{Q}, \mathbb{P}))}{\Omega(E)}$$

From the fundamental postulate of statistical mechanics, in terms of the entropy S_b of the heat bath we have:

$$\Omega_b(E - \mathcal{H}) = e^{\frac{1}{k_B} S_b(E - \mathcal{H})}$$

Since the subsystem is much smaller than the heat bath we have[16] $\mathcal{H} \ll E$ and so we can expand S_b around E:

$$\Omega_b(E - \mathcal{H}) = e^{\frac{1}{k_B} \left(S_b(E) - \mathcal{H} \frac{\partial}{\partial E} S_b(E) + \frac{\mathcal{H}^2}{2} \frac{\partial^2}{\partial E^2} S_b(E) + \cdots \right)}$$

---

[16]In fact, $E \propto N + N_b$ while $\mathcal{H} \propto N \ll N_b \approx N + N_b$, so indeed $\mathcal{H} \ll E$.

Since:

$$\frac{\partial S_{\rm b}}{\partial E} = \frac{1}{T} \qquad \frac{\partial^2 S_{\rm b}}{\partial E^2} = \frac{\partial}{\partial E}\frac{1}{T} = -\frac{1}{T^2}\frac{\partial T}{\partial E} = -\frac{1}{T^2}\left(\frac{\partial E}{\partial T}\right)^{-1} = -\frac{1}{T^2 C_V^{\rm b}}$$

where $C_V^{\rm b}$ is the specific heat of the bath at constant volume, then:

$$\Omega_{\rm b}(E - \mathcal{H}) = e^{\frac{1}{k_B}\left(S_{\rm b}(E) - \frac{\mathcal{H}}{T} - \frac{\mathcal{H}^2}{2}\frac{1}{T^2 C_V^{\rm b}} + \cdots\right)}$$

Now, $\mathcal{H} \propto N$ and $\mathcal{H}^2/C_V^{\rm b} \propto N^2/N_{\rm b} = N(N/N_{\rm b}) \ll N$: all the terms beyond the first order can be neglected[17]. Considering that $\Omega(E) = \int d\Gamma d\Gamma_{\rm b}\delta(\mathcal{H}_{\rm b} + \mathcal{H} - E) = \int d\Gamma \Omega_{\rm b}(E - \mathcal{H})$, this leads to the following expression of the *canonical probability density*:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{e^{-\beta\mathcal{H}(\mathbb{Q},\mathbb{P}) + \frac{1}{k_B}S_{\rm b}(E)}}{\int d\Gamma e^{-\beta\mathcal{H}(\mathbb{Q},\mathbb{P}) + \frac{1}{k_B}S_{\rm b}(E)}} = \frac{e^{-\beta\mathcal{H}(\mathbb{Q},\mathbb{P})}}{\int d\Gamma e^{-\beta\mathcal{H}(\mathbb{Q},\mathbb{P})}}$$

where:

$$Z = \int d\Gamma e^{-\beta\mathcal{H}(\mathbb{Q},\mathbb{P})} \tag{3.9}$$

is called *partition function* of the system. Sometimes $e^{-\beta\mathcal{H}}$ is called *Boltzmann weight*.

3.3.1 The canonical and the microcanonical ensemble

We could now ask how the microcanonical and the canonical ensembles are related.
Since in the canonical ensemble we have removed the constraint of having constant energy, the energy of a system will in general fluctuate around its mean value. We can therefore ask if these fluctuations are relevant or not. In fact if it turns out that they are negligible (at least in the thermodynamic limit) then we can conclude that the canonical and microcanonical ensembles are equivalent.

Let us therefore compute $\langle\mathcal{H}\rangle$ and $\sigma_E^2 = \langle(\mathcal{H} - \langle\mathcal{H}\rangle)^2\rangle$.
First of all, from the definition (3.9) of the canonical partition function we have:

$$\langle\mathcal{H}\rangle = \frac{1}{Z}\int e^{-\beta\mathcal{H}}\mathcal{H}d\Gamma = -\frac{1}{Z}\frac{\partial Z}{\partial\beta} = -\frac{\partial \ln Z}{\partial\beta}$$

and:

$$\sigma_E^2 = \langle\mathcal{H}^2\rangle - \langle\mathcal{H}\rangle^2 = \frac{1}{Z}\frac{\partial^2 Z}{\partial\beta^2} - \left(-\frac{\partial \ln Z}{\partial\beta}\right)^2 = \frac{\partial^2 \ln Z}{\partial\beta^2} = -\frac{\partial \langle\mathcal{H}\rangle}{\partial\beta} = -\frac{\partial T}{\partial\beta}\frac{\partial \langle\mathcal{H}\rangle}{\partial T} = k_B T^2 C_V$$

$$\Rightarrow \qquad \sigma_E^2 = k_B T^2 C_V$$

This is a *fluctuation-dissipation* relation, which we couldn't find using only thermodynamics. Therefore, the relative fluctuation of energy is:

$$\frac{\sigma_E}{\langle H\rangle} = \frac{\sqrt{k_B T^2 C_V}}{\langle H\rangle}$$

Both $\langle\mathcal{H}\rangle$ and C_V are extensive quantities, i.e. proportional to N, and therefore:

$$\frac{\sigma_E}{\langle\mathcal{H}\rangle} \propto \frac{1}{\sqrt{N}}$$

---

[17]They are all terms that are smaller the bigger the heat bath is.
This fact could also be justified saying that we can consider the limit where the dimensions of the heat bath tend to infinity, therefore $C_V^{\rm b} \to \infty$ and we can neglect all the terms from the second order on.

Thus, if our system is macroscopic the relative fluctuations of energy are absolutely negligible (as we have already seen, for $N \sim 10^{23}$ this relative fluctuation is of the order of 10^{-11})!

We can therefore conclude that the canonical and microcanonical ensembles are indeed equivalent.

3.3.2   Helmholtz free energy

Since we have just shown that the microcanonical and the canonical ensembles are equivalent we expect that the thermodynamics of a system can be derived also from the canonical ensemble, but how?

What we now want to show is that the *Helmholtz free energy* of a system (see 1.1.3) finds a natural place in the canonical ensemble. In particular we want to show that $Z = e^{-\beta F}$, where F is the free energy of the system. This way, once the canonical partition function Z has been computed we can obtain the Helmholtz free energy as $F = -k_B T \ln Z$, and with appropriate derivatives we can obtain all the thermodynamics of the system (see 1.1.3).

The first thing we can note is that the canonical partition function Z can be written in terms of the microcanonical phase space volume $\Omega(E)$:

$$Z = \int e^{-\beta \mathcal{H}} d\Gamma = \int d\Gamma \int dE \delta(\mathcal{H} - E)e^{-\beta\mathcal{H}} = \int dE e^{-\beta E} \underbrace{\int d\Gamma \delta(E - \mathcal{H})}_{\Omega(E)} = \int dE e^{-\beta E}\Omega(E)$$

(where we have only inserted a $\int dE\delta(\mathcal{H} - E) = 1$). This means that the canonical ensemble can be thought of as an ensemble of microcanonical systems each weighted with $e^{-\beta E}$. Furthermore, if the Hamiltonian $\mathcal{H}$ is bounded below then we can always shift it by a constant amount so that $\mathcal{H} \geq 0$, and thus:

$$Z = \int_0^\infty e^{-\beta E}\Omega(E)dE \tag{3.10}$$

which is a Laplace transform! In other words Z is the Laplace transform of $\Omega(E)$, and β is the variable conjugated to E.

From the microcanonical definition of entropy we have $\Omega(E) = e^{S(E)/k_B} = e^{\beta T S(E)}$, so that (3.10) can be rewritten as:

$$Z = \int dE e^{-\beta(E - TS)}$$

Now, both E and S are extensive so we can use the saddle point approximation (see appendix B) to compute the integral. In particular, in order to do so we need to determine the value $\overline{E}$ of the energy that maximizes $-(E - TS)$, namely minimizes $E - TS$. Therefore:

$$\frac{\partial}{\partial E}(E - TS)_{|\overline{E}} = 0 \qquad \Rightarrow \qquad 1 - T\frac{\partial S}{\partial E}_{|\overline{E}} = 0 \qquad \Rightarrow \qquad \frac{\partial S}{\partial E}_{|\overline{E}} = \frac{1}{T}$$

and:

$$\frac{\partial^2}{\partial E^2}(E - TS)_{|\overline{E}} = -T\frac{\partial^2 S}{\partial E^2}_{|\overline{E}} = \frac{1}{TC_V}$$

Thus, if $C_V > 0$ (which is always the case since it means that giving heat to a system its temperature will increase) then $\overline{E}$ is indeed a minimum.

If we now expand the integrand of Z around $\overline{E}$:

$$Z = \int dE e^{-\beta(\overline{E} - TS(\overline{E})) - \frac{\beta}{2TC_V}(E - \overline{E})^2 + \cdots}$$

and the values of E that contribute significantly to Z are those such that $(E - \overline{E})^2/C_V \lesssim 1$, namely $|E - \overline{E}| \lesssim \sqrt{C_V} \propto \sqrt{N}$: for this reason in the thermodynamic limit we can integrate E

on the whole real axis (even if E does not span $\mathbb{R}$), since in this way we introduce a perfectly negligible error. Therefore:

$$Z = \left(2\pi T^2 C_V k_B\right)^{1/2} e^{-\beta(\overline{E} - TS(\overline{E}))} \qquad \Rightarrow \qquad \ln Z = -\beta(\overline{E} - TS(\overline{E})) + \frac{1}{2}\ln\left(2\pi T^2 k_B C_V\right)$$

The first term of $\ln Z$ is extensive, while the second is proportional to $\ln N$; in the thermodynamic limit this last contribution vanishes and so:

$$-\frac{1}{\beta}\ln Z = \overline{E} - TS(\overline{E}) \tag{3.11}$$

This is indeed the free energy F of the system if we have $\overline{E} = \langle \mathcal{H} \rangle$. However:

$$\langle \mathcal{H} \rangle = \frac{1}{Z}\int E e^{-\beta(E - TS(E))}dE = -\frac{\partial}{\partial \beta}\ln Z = \frac{\partial}{\partial \beta}\left[\beta(\overline{E} - TS(\overline{E}))\right] = \overline{E} + \beta\frac{\partial}{\partial \beta}\left(\overline{E} - TS(\overline{E})\right)$$

and:

$$\beta\frac{\partial}{\partial \beta}\left(\overline{E} - TS(\overline{E})\right) = \beta\frac{\partial\overline{E}}{\partial \beta}\frac{\partial}{\partial E}\left(E - TS(E)\right)_{|\overline{E}} = 0$$

because by definition $\overline{E}$ is the value of E that extremizes $E - TS(E)$.
Therefore $\langle \mathcal{H} \rangle = \overline{E}$, and the expression in (3.11) is indeed the free energy F of the system:

$$Z = e^{-\beta F}$$

We also see that although the partition function formally allows the system to have any value of energy, it is largely "dominated" by the configurations in which the system has energy $\langle \mathcal{H} \rangle$.

3.3.3 The canonical ensemble at work

Suppose we have a system whose Hamiltonian is:

$$\mathcal{H} = \sum_i \frac{\vec{p}_i^{\,2}}{2m} + \sum_i U(\vec{q}_i) + \frac{1}{2}\sum_{i \neq j} U_2(\vec{q}_i, \vec{q}_j)$$

(for example, U can be the gravitational potential and U_2 an interaction between the particles). Then the canonical probability density is:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{1}{Z}e^{-\beta\sum_i \frac{\vec{p}_i^{\,2}}{2m} - \beta\left[\sum_i U(\vec{q}_i) + \frac{1}{2}\sum_{i \neq j} U_2(\vec{q}_i, \vec{q}_j)\right]}$$

Note that $\mathbb{Q}$ and $\mathbb{P}$ don't "get mixed up"; therefore, when calculating the partition function Z we can separate the kinetic and configurational contributions[18]:

$$Z = \int d\Gamma\, \rho(\mathbb{Q}, \mathbb{P}) = \prod_{i=1}^{N}\left(\frac{1}{N!h^{3N}}\int e^{-\beta\frac{\vec{p}_i^{\,2}}{2m}}d\vec{p}_i\right) \cdot \int e^{-\beta\left[\sum_i U(\vec{q}_i) + \frac{1}{2}\sum_{i \neq j} U_2(\vec{q}_i, \vec{q}_j)\right]}d\mathbb{Q}$$

Now, since:

$$\int e^{-\beta\frac{\vec{p}^{\,2}}{2m}}d\vec{p} = \left(\frac{2\pi m}{\beta}\right)^{3/2}$$

---

[18]This "separation" was not possible within the microcanonical ensemble and as we will later see (for example in chapter 5) this will simplify the computations when we will consider non ideal systems, namely if the particles actually interact; in fact since any reasonable interaction does not involve the particles' momenta but only their positions only Q_N (see below) will be affected by the presence of the interaction.

we have:

$$Z = \frac{1}{N!} \left(\frac{2\pi m}{h^2 \beta} \right)^{3N/2} \int e^{-\beta \left[\sum_i U(\vec{q}_i) + \frac{1}{2} \sum_{i \neq j} U_2(\vec{q}_i, \vec{q}_j) \right]} d\mathbb{Q}$$

In order to simplify this expression, we define the *configurational partition function* (or *configurational sum*) as:

$$Q_N = \int e^{-\beta \left[\sum_i U(\vec{q}_i) + \frac{1}{2} \sum_{i \neq j} U_2(\vec{q}_i, \vec{q}_j) \right]} d\mathbb{Q}$$

and the *thermal wavelength* as:

$$\Lambda = \sqrt{\frac{h^2}{2\pi k_B T m}} = \sqrt{\frac{2\pi \hbar^2}{k_B T m}}$$

so that, in general:

$$Z = \frac{1}{N! \Lambda^{3N}} Q_N$$

If we now consider an ideal gas, namely we set $U = U_2 = 0$, then:

$$Q_N = V^N \qquad \Rightarrow \qquad Z = \frac{1}{N!} \left(\frac{V}{\Lambda^3} \right)^N$$

And using Stirling's approximation (see appendix C) the free energy $F = -k_B T \ln Z$ of the system turns out to be:

$$F = k_B T N \left[\ln \left(\frac{N}{V} \Lambda^3 \right) - 1 \right]$$

Since from thermodynamics we know that $P = -\partial F / \partial V$, we get:

$$P = -k_B T N \frac{V}{N\Lambda^3} \left(-\frac{N}{V^2} \Lambda^3 \right) = \frac{k_B T N}{V} \qquad \Rightarrow \qquad PV = N k_B T$$

We have therefore found the state equation of an ideal gas in the canonical ensemble!
We are thus *really* verifying that the canonical and microcanonical ensembles are equivalent. For example, we can also compute the entropy of this ideal gas, and we get:

$$S = -\frac{\partial F}{\partial T} = N k_B \left[\frac{5}{2} - \ln \left(\frac{N}{V} \Lambda^3 \right) \right]$$

which is exactly what we have found in the microcanonical ensemble (see equation (3.8)).

Column of gas in a gravitational potential   Just as another example of the power of the canonical ensemble, we now want to see that in a particular case we can obtain the same results that we can determine from simple kinetic considerations. In particular, we want to see how the density of a gas subjected to a uniform gravitational potential varies with the height from the ground.

We begin with the "kinetic" computation.
Referring to figure 3.5, consider a column of gas subjected to a constant gravitational potential mgz, where z is the height from the ground. The action of gravity will in some way change the density of the particles (we expect that the gas will be more dense near the ground and become thinner and thinner as z grows) so let us call $\rho(z)$ and $P(z)$ the density and the pressure of the gas at height z.
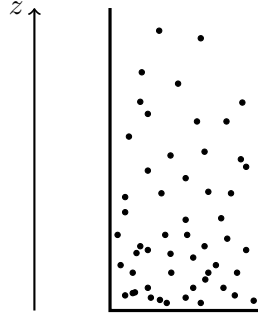
Figure 3.5: Column of gas subjected to gravity

Let us call $n(z)\Delta z$ is the number of particles that are between the heights z and $z + \Delta z$, where $n(z)$ is the particle number density at height z; these will be subjected to the pressures $P(z)$ and $P(z + \Delta z)$, and if we call Σ the cross section of the column then the net force acting on the particles is $P(z)\Sigma - P(z+\Delta z)\Sigma$. Since the system is in equilibrium, this force must be balanced by the weight of the gas itself:

$$P(z)\Sigma - P(z + \Delta z)\Sigma = \rho(z)\Delta z \Sigma g$$

Therefore, dividing by Σ and Δz and taking the limit $\Delta z \to 0$:

$$\frac{\partial P}{\partial z} = -\rho(z)g$$

If we then consider a volume V between z and $z + \Delta z$ we have:

$$P(z)V = N(z)k_B T \qquad \Rightarrow \qquad P(z) = \frac{N(z)}{V}k_B T$$

and since $\rho(z) = N(z)m/V = n(z)m$, then:

$$P(z) = \frac{\rho(z)}{m}k_B T \qquad \Rightarrow \qquad \rho(z) = \frac{m}{k_B T}P(z)$$

Therefore:

$$\frac{\partial P}{\partial z} = -g\rho(z) = -\frac{mg}{k_B T}P(z) \qquad \Rightarrow \qquad P(z) = P(0)e^{-\frac{mg}{k_B T}z}$$

or equivalently:

$$n(z) = n(0)e^{-\frac{mg}{k_B T}z} \tag{3.12}$$

Thus, the density of the gas decreases exponentially with the height.

We now want to show that the tools of the canonical ensemble allow us to obtain the very same result. Let us consider the *single-particle* canonical partition function[19]:

$$Z_{\text{s.p.}} = \int e^{-\beta\left(\frac{\vec{p}^2}{2m} + mgz\right)}\frac{d\vec{p}\,d\vec{q}}{h^3}$$

Since $\vec{q} = (x, y, z)$ is such that $(x, y) \in \Sigma$ and $z \in [0, +\infty]$, integrating we will have:

$$Z_{\text{s.p.}} = \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2}\frac{\Sigma}{\beta mg}$$

---

[19]We use the single-particle partition function because in the end we want to find the single-particle configurational probability density.

The single-particle probability density of having momentum $\vec{p}$ and height z is:

$$\rho_{\text{s.p.}}(\vec{p}, \vec{q}) = \frac{1}{Z_{\text{s.p.}}} e^{-\beta\left(\frac{\vec{p}^2}{2m} + mgz\right)}$$

Therefore we can obtain the spatial probability density integrating over $\vec{p}$, x and y; in the end we get:

$$\rho_{\text{s.p.}}(z) = \frac{mg}{k_B T} e^{-\frac{mg}{k_B T} z} := \rho_{\text{s.p.}}(0) e^{-\frac{mg}{k_B T} z}$$

where we have relabelled the proportionality constant. Now, the particle number density at height z will be $n(z) = N\rho_{\text{s.p.}}(z)$, so in the end we have:

$$n(z) = n(0) e^{-\frac{mg}{k_B T} z}$$

which is exactly (3.12).

3.3.4 The equipartition theorem

We now want to cover an important topic in statistical mechanics, the *equipartition theorem*. In very general words it can be formulated as follows: *the mean energy of a particle is equal to $k_B T/2$ times the number of microscopic degrees of freedom.*

We now see some examples and then we shall prove the theorem in general.

Real gas Let us begin with a gas with the following Hamiltonian:

$$\mathcal{H} = \sum_i \frac{\vec{p_i}^2}{2m} + \sum_i U(\vec{q_i}) + \frac{1}{2} \sum_{i \neq j} U_2(\vec{q_i}, \vec{q_j})$$

(so it is not necessarily an ideal gas since we are also including the interaction potential U_2). Its phase space probability density is of course:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{1}{Z} e^{-\beta\left(\sum_i \frac{\vec{p_i}^2}{2m} + \sum_i U(\vec{q_i}) + \frac{1}{2} \sum_{i \neq j} U_2(\vec{q_i}, \vec{q_j})\right)}$$

Then we will have:

$$\left\langle \frac{\vec{p_i}^2}{2m} \right\rangle = \frac{3}{2} k_B T$$

In fact:

$$\left\langle \frac{\vec{p_i}^2}{2m} \right\rangle = \frac{\int \frac{\vec{p_i}^2}{2m} e^{-\beta\left(\sum_j \frac{\vec{p_j}^2}{2m} + \sum_j U(\vec{q_j}) + \frac{1}{2} \sum_{j \neq k} U_2(\vec{q_j}, \vec{q_k})\right)} d\mathbb{Q} d\mathbb{P}}{\int e^{-\beta\left(\sum_j \frac{\vec{p_j}^2}{2m} + \sum_j U(\vec{q_j}) + \frac{1}{2} \sum_{j \neq k} U_2(\vec{q_j}, \vec{q_k})\right)} d\mathbb{Q} d\mathbb{P}}$$

and the integrals in $\mathbb{Q}$ and $\mathbb{P}$ factorize, since there are no terms that mix them. Therefore:

$$\left\langle \frac{\vec{p_i}^2}{2m} \right\rangle = \frac{\int \frac{\vec{p_i}^2}{2m} e^{-\beta \sum_j \frac{\vec{p_j}^2}{2m}} d\mathbb{P}}{\int e^{-\beta \sum_j \frac{\vec{p_j}^2}{2m}} d\mathbb{P}} = \frac{\int \frac{\vec{p_i}^2}{2m} e^{-\beta \frac{\vec{p_i}^2}{2m}} d\vec{p_i}}{\int e^{-\beta \frac{\vec{p_i}^2}{2m}} d\vec{p_i}}$$

where in the last step we have done the integrals for $j \neq i$, which are all equal and thus simplify. Thus, in the end:

$$\left\langle \frac{\vec{p_i}^2}{2m} \right\rangle = -\frac{\partial}{\partial \beta} \ln\left(\int e^{-\beta \frac{\vec{p}^2}{2m}} d\vec{p}\right) = -\frac{\partial}{\partial \beta} \ln\left[\left(\frac{2\pi m}{\beta}\right)^{3/2}\right] = \frac{3}{2} k_B T$$

In terms of what we have stated at the beginning in this case the microscopic degrees of freedom are three, i.e. the three possible directions of motion.

Harmonic oscillator in a heat bath   Let us now consider a single one-dimensional harmonic oscillator in a heat bath (the presence of the heat bath justifies the use of the canonical partition function); in other words we are considering a single particle with Hamiltonian:

$$\mathcal{H} = \frac{p^2}{2m} + m\frac{\omega^2}{2}q^2$$

The (single-particle) partition function is:

$$Z = \int e^{-\beta\left(\frac{p^2}{2m}+m\frac{\omega^2}{2}q^2\right)}\frac{dqdp}{h} = \frac{1}{h}\left(\int e^{-\beta m\frac{\omega^2}{2}q^2}dq\right)\left(\int e^{-\beta\frac{p^2}{2m}}dp\right) =$$

$$= \frac{1}{h}\sqrt{\frac{2\pi}{\beta m\omega^2}}\sqrt{\frac{2\pi m}{\beta}} = \frac{1}{\beta\hbar\omega}$$

Now, from the expression of the partition function we see that we should have:

$$\left\langle\frac{p^2}{2m}\right\rangle = \left\langle m\frac{\omega^2}{2}q^2\right\rangle = -\frac{\partial}{\partial\beta}\ln Z$$

However, if we simply derive Z as we have found it we determine $\langle p^2/2m + m\omega^2 q^2/2\rangle$, not $\langle p^2/2m\rangle$ and $\langle m\omega^2 q^2/2\rangle$ singularly. In order to avoid this problem we give a different name to the parameters β that multiply the kinetic and the configurational parts in the integrals that define Z, i.e. we set:

$$Z = \frac{1}{h}\left(\int e^{-\beta m\frac{\omega^2}{2}q^2}dq\right)\left(\int e^{-\beta'\frac{p^2}{2m}}dp\right) = \frac{1}{\hbar\omega}\cdot\frac{1}{\sqrt{\beta\beta'}}$$

(and of course in the end we must set $\beta = \beta'$). This way, we have:

$$\left\langle\frac{p^2}{2m}\right\rangle = -\frac{\partial}{\partial\beta}\ln Z = \frac{1}{2}\cdot\frac{1}{\beta\beta'}\beta' = \frac{1}{2\beta} = \frac{k_B T}{2}$$

$$\left\langle m\frac{\omega^2}{2}q^2\right\rangle = -\frac{\partial}{\partial\beta'}\ln Z = \frac{1}{2\beta'}\Big|_{\beta'=\beta} = \frac{k_B T}{2}$$

Again, we see that every degree of freedom of the particle (in this case, the translational and vibrational ones) contributes with $k_B T/2$ to its energy.

We can now move on and prove the equipartition theorem in general.

Theorem (Equipartition). *Every term in the Hamiltonian of a system that appears* only *quadratically contributes to the total energy with $k_B T/2$. In other words, if we can write the Hamiltonian of the system in the form:*

$$\mathcal{H}(\mathbb{Q}, \mathbb{P}) = \mathcal{H}' + kx^2$$

where x is any *of the variables $(\mathbb{Q}, \mathbb{P})$ and $\mathcal{H}'$ is a Hamiltonian that does* not *depend on x, then:*

$$\langle kx^2\rangle = \frac{k_B T}{2}$$

Proof. The partition function of the system is:

$$Z = \int d\Gamma e^{-\beta\mathcal{H}} = \int e^{-\beta\mathcal{H}'}d\Gamma_x\int e^{-\beta' kx^2}dx$$
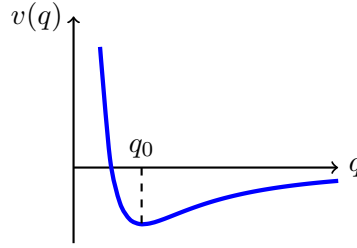
Figure 3.6: Interaction potential

where with $d\Gamma_x$ we mean $d\Gamma$ without dx, and as we have done before we call β' the parameter β that multiplies kx^2 in order to compute $\langle kx^2 \rangle$ more simply.
Therefore:

$$Z = \sqrt{\frac{\pi}{k\beta'}} \int e^{-\beta \mathcal{H}'} d\Gamma_x$$

From the definition of partition function, then:

$$\langle kx^2 \rangle = \int kx^2 e^{-\beta' kx^2} dx = -\frac{\partial}{\partial \beta'} \ln Z = \frac{1}{2\beta'}_{|\beta'=\beta} = \frac{k_B T}{2}$$

$\square$

Therefore, if we have an ideal gas of N particles its energy will be:

$$E = \langle \mathcal{H} \rangle = \frac{3}{2} N k_B T$$

since every particle contributes to the energy with $3k_B T/2$. The specific heat at constant volume of such a system is then $C_V = \partial E/\partial T = 3Nk_B/2$, and if we compute it for a single mole of gas we will have:

$$C_V^{\text{m.}} = \frac{3}{2} R$$

where "m." stands for "molar" and $R = N_A k_B$ is the gas constant.
On the other hand if we have a gas of N one-dimensional harmonic oscillators then the energy of the system will be:

$$E = \langle \mathcal{H} \rangle = k_B T$$

because this time every particle will contribute with $k_B T$ (a $k_B T/2$ for the kinetic part and a $k_B T/2$ for the potential term), and its molar specific heat at constant volume is $C_V^{\text{m.}} = R$.

An application of the equipartition theorem: the specific heat of crystals

Now, we could ask if the model of a gas of harmonic oscillators is actually realistic; in other words, are there cases where a particle can be actually considered a harmonic oscillator?
Let us suppose that the particles of our system are subjected to the potential shown in figure 3.6[20]. Clearly, the equilibrium configurations for the particles will be $q = q_0$; furthermore, if the temperature of the system is very small the fluctuations of the particles around this equilibrium will be very small and therefore we can expand the potential around q_0:

$$v(q) = v(q_0) + (q - q_0) \frac{\partial v}{\partial q}_{|q_0} + \frac{1}{2} (q - q_0)^2 \frac{\partial^2 v}{\partial q^2}_{|q_0} + \cdots$$

---

[20]This is the *Lennard-Jones potential*, which we will cover in some more detail in 6.4.2. It is a very realistic potential for interatomic interactions.

Now, since q_0 is a global minimum for v we will have $\partial v/\partial q_{|q_0} = 0$, and $\partial^2 v/\partial q^2_{|q_0} := k > 0$. Therefore, neglecting $v(q_0)$ since it is just an additive constant, the potential acting on the particles will be:

$$v(q) \sim \frac{k}{2}(q - q_0)^2 := \frac{m\omega^2}{2}(q - q_0)^2$$

where in the last step we have renamed the constant k as $m\omega^2$.

Let us see explicitly that this approximation is good if the system is in contact with a heat bath whose temperature is low, i.e. that the mean displacement of the particles from q_0 is small at low temperatures.

Setting $q_0 = 0$ for simplicity, we have:

$$\langle q^2 \rangle = \frac{2}{m\omega^2} \left\langle m \frac{\omega^2}{2} q^2 \right\rangle = \frac{2}{m\omega^2} \cdot \frac{k_B T}{2} = \frac{k_B T}{m\omega^2}$$

and therefore if $k_B T \ll m\omega^2$ then $\langle q^2 \rangle \approx 0$. This means that for low temperatures the interparticle potential can actually be approximated with a harmonic one.

However, if this is not the case the mean displacement $\langle q^2 \rangle$ from the equilibrium will be large and the particle will "realize" that the potential is actually different from a parabola.

Now, these facts can be used in order to describe the properties of crystals.

Let us in fact suppose to have a system of N particles with Hamiltonian:

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\vec{p_i}^2}{2m} + \underbrace{\frac{1}{2} \sum_{i \neq j} \mathcal{V}(|\vec{q_i} - \vec{q_j}|)}_{:=\mathcal{V}(\mathbb{Q})}$$

and that the interaction potential $\mathcal{V}(\mathbb{Q})$ is like the one in figure 3.6 when considering the distance between two atoms, and such that the particles of the system have positions $\vec{q_i} = \vec{q_i}^{(0)} + \delta\vec{q_i}$, where $\vec{q_i}^{(0)}$ is the equilibrium position of the i-th particle (analogous to the position q_0 in figure 3.6) and $\delta\vec{q_i}$ is the displacement from it. We also suppose that $\vec{q_i}^{(0)}$ are fixed and arranged in a lattice; this is what normally happens in crystalline solids (like metals).

Therefore, just like we have done before, if we suppose the displacements $\delta\vec{q_i}$ to be small we can expand $\mathcal{V}(\mathbb{Q})$ around the equilibrium positions $\mathbb{Q}_0$:

$$\mathcal{V}(\mathbb{Q}) = \mathcal{V}(\mathbb{Q}_0) + \sum_i \sum_\alpha \delta q_{i\alpha} \frac{\partial \mathcal{V}}{\partial q_{i\alpha}}_{|\mathbb{Q}_0} + \frac{1}{2} \sum_{i,j} \sum_{\alpha,\beta} \delta q_{i\alpha} \delta q_{j\beta} \frac{\partial^2 \mathcal{V}}{\partial q_{i\alpha} \partial q_{j\beta}}_{|\mathbb{Q}_0} + \cdots$$

where $i = 1, \ldots, N$ is the index that labels the particle and $\alpha = x, y, z$ the index that labels its coordinates. Note that $i\alpha$ and $j\beta$ are single indexes (i.e., strictly speaking i and α are not independent indexes, like j and β).

Since $\mathbb{Q}_0$ are the equilibrium positions of the particles, we have:

$$\frac{\partial \mathcal{V}}{\partial q_{i\alpha}}_{|\mathbb{Q}_0} = 0$$

Furthermore the second derivative of $\mathcal{V}$ is a matrix, which we call K:

$$\frac{\partial^2 \mathcal{V}}{\partial q_{i\alpha} \partial q_{j\beta}}_{|\mathbb{Q}_0} := K_{i\alpha,j\beta}$$

This way, also neglecting the additive constant $\mathcal{V}(\mathbb{Q}_0)$, we can write the interaction potential as:

$$\mathcal{V}(\mathbb{Q}) = \frac{1}{2} \sum_{i\alpha,j\beta} \delta q_{i\alpha} K_{i\alpha,j\beta} \delta q_{j\beta} = \frac{1}{2} \delta\mathbb{Q}^T K \delta\mathbb{Q} \tag{3.13}$$

Now, by definition K is a symmetric matrix (remember that we must exchange $i\alpha$ with $j\beta$) and so it can be diagonalized, namely we can write:

$$K = \mathbb{S}^T \Lambda \mathbb{S} \qquad \text{with} \qquad \mathbb{S}^T \mathbb{S} = \mathbb{I}, \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{3N}), \quad \lambda_i > 0$$

Therefore the Hamiltonian of the system can be rewritten in the form:

$$\mathcal{H} = \sum_i \frac{\vec{p_i}^2}{2m} + \frac{1}{2} \delta\mathbb{Q}^T \mathbb{S}^T \Lambda \mathbb{S} \delta\mathbb{Q} = \frac{1}{2m} \mathbb{P}^T \cdot \mathbb{I} \cdot \mathbb{P} + \frac{1}{2} \delta\mathbb{Q} \mathbb{S}^T \Lambda \mathbb{S} \delta\mathbb{Q}$$

If we now define the new variables:

$$\delta\hat{\mathbb{Q}} := \mathbb{S}\delta\mathbb{Q} \qquad\qquad \hat{\mathbb{P}} = \mathbb{S}\mathbb{P} \tag{3.14}$$

then:

$$\mathcal{H} = \frac{\hat{\mathbb{P}}^2}{2m} + \frac{1}{2} \delta\hat{\mathbb{Q}}^T \Lambda \delta\hat{\mathbb{Q}}$$

or, explicitly:

$$\mathcal{H} = \sum_{i\alpha} \left[\frac{\hat{p}_{i\alpha}^2}{2m} + \lambda_{i\alpha} \frac{\delta\hat{q}_{i\alpha}^2}{2} \right] \tag{3.15}$$

Therefore, with the change of variables (3.14) (called *normal modes of vibration*) we have rewritten the Hamiltonian as a sum of $3N$ *decoupled* harmonic oscillators.
Note in fact that in (3.13) the positions $\mathbb{Q}$ are coupled by the matrix K, which in general is not diagonal[21], while now the new variables are not. Note however that these new variables are not related to the positions or the momenta of any of the atoms: they are just some generalised coordinates, which allow us to rewrite the Hamiltonian in a simpler way[22].
Thus, since our system is equivalent to a set of $3N$ harmonic oscillators, its energy will be:

$$\langle \mathcal{H} \rangle = 3N \left(\frac{k_B T}{2} + \frac{k_B T}{2} \right) = 3N k_B T$$

which is twice the energy of the ideal gas.
Therefore, the model of crystals we are studying predicts a value for the molar specific heat at constant volume equal to:

$$C_V^{\text{m.}} = 3R$$

This is experimentally verified, but only for high temperatures; we will cover this issue in some more detail in the next section, and as we will see this problem is essentially due to the fact that for low temperatures quantum effects become relevant.

One last application of the equipartition theorem: C_V for biatomic gases

If we apply what we have seen for the equipartition theorem to a biatomic gas, what do we expect?
In this case the atoms in a molecule can oscillate around their equilibrium positions and the

---

[21]The positions and the oscillations of the atoms in a crystal are strongly correlated: when an atom is excited and oscillates faster it "gives" some of its energy to the nearby atoms increasing their energy of oscillation. If this was not the case, the crystal would disgregate.

[22]Note, however, that if our fundamental assumption is not valid, i.e. if the displacements of the atoms from their equilibrium positions are not small, then also these new variables would be coupled: in this case in fact we couldn't have neglected the third derivatives of $\mathcal{V}$, which in the form (3.15) of the Hamiltonian would have been a coupling term between the $\delta\hat{q}$.

molecules can rotate. Choosing a reference frame where the vector that connects the two atoms is directed along the z axis, the Hamiltonian of such a system will be:

$$\mathcal{H} = \sum_i \frac{\vec{p_i}^2}{2m} + \sum_i \left[\left(\frac{p_{\text{rel},i}^2}{2m} + \frac{m\omega^2}{2} q_{\text{rel},i}^2 \right) + \frac{1}{2} \left(I_1 \dot{\theta}_{i1}^2 + I_2 \dot{\theta}_{i2}^2 \right) \right]$$

where q_{rel} is the relative distance between two atoms in the same molecule and I_α and $\theta_{i\alpha}$ are, respectively, the moment of inertia of the molecule and its angle relative to the α-th axis, with $\alpha = x, y$ (the rotations around the direction that connects the two atoms are ineffective). Therefore, we expect the energy of the system to be:

$$\langle \mathcal{H} \rangle = N k_B T \left(\frac{3}{2} + 2\frac{1}{2} + 2\frac{1}{2} \right) = \frac{7}{2} N k_B T$$

and the molar specific heat at constant volume to be $C_V^{\text{m.}} = 7R/2$.

However, experimentally we observe something quite different: for low temperatures $C_V^{\text{m.}}$ turns out to be $3R/2$, then at a certain point increasing the temperature $C_V^{\text{m.}}$ jumps to $5R/2$, and then at higher temperatures there is another jump to $7R/2$. In other words, "new" degrees of freedom become "visible" for high enough temperatures (in the case we have mentioned, increasing T we "see" first the translational degrees of freedom, then the rotational and in the end the vibrational ones).

This is due to the fact that in reality the molecules are quantum systems and the observed behaviour of $C_V^{\text{m.}}$ comes from their properties on a quantum level.

To be more explicit: it is a known fact that in quantum mechanics the energy levels of a system are discrete (for example, for a harmonic oscillator $E_n = (n + 1/2)\hbar\omega$); in particular any system will have a non-null *zero-point energy*, i.e. the lowest possible energy level, and then all the possible excited states. If the temperature of the system is low enough it will occupy its lowest possible energy state; if we then increase T the thermal energy of the system ($k_B T$) at a certain point will be large enough to allow the system to pass to the first excited states, and then to the others.

For example the harmonic oscillator has equally spaced energy levels, with the spacing equal to $\hbar\omega$; if the temperature of the system is such that $k_B T \ll \hbar\omega$ the system cannot acquire the necessary energy to pass to the first excited state (this will be possible as soon as $k_B T \approx \hbar\omega$).

In the case of the biatomic gas the observed behaviour is due exactly to this mechanism: the first excited state of the vibrational spectrum has an energy higher to the first excited state of the rotational spectrum, so increasing the temperature at a certain point we will be able to give enough energy to the molecules to pass from their ground state to the first rotational excited state, and similarly when we further increase the temperature we will be able to make them pass to the first vibrational excited state.

3.4 The grand canonical ensemble

In 3.3 we have relaxed the constraint of having a fixed value of the energy, and thus defined the canonical ensemble. However, we still supposed that the number of particles of our system was fixed: we can remove also this constraint, and so consider systems that exchange not only energy but also particles.

The statistical ensemble of such systems is called *grand canonical ensemble*.

We therefore consider a system like that shown in figure 3.7: we again have a heat bath at temperature T and a small subsystem that this time can also exchange particles with the bath; let us call N_{b} the number of particles of the heat bath, N_{s} the particles of the subsystem (which
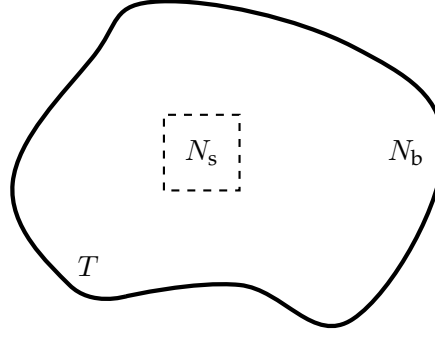
Figure 3.7: Grand canonical ensemble

can both vary from instant to instant), $N = N_{\mathrm{s}} + N_{\mathrm{b}}$ the total number of particles of the system (which is instead fixed) and V the total volume.

As we have done for the canonical ensemble, we rely on the tools of the microcanonical ensemble, namely we must determine the phase space volume of the system taking also into account that the number of particles is not fixed. This volume is:

$$\Omega(E, N, V) = \sum_{N_{\mathrm{s}}=0}^{N} \int d\Gamma_{\mathrm{s}} d\Gamma_{\mathrm{b}} \delta(\mathcal{H}_{\mathrm{s}} + \mathcal{H}_{\mathrm{b}} - E)$$

where E is the total energy of the system, and the sum is taken exactly because the number of particles in the system is not fixed and in principle can run from zero to the largest possible value, i.e. N; furthermore all the quantities with the subscript "s" are relative to the subsystem while those with "b" to the heat bath, and it is implied that the former are computed for N_{s} particles while the latter for $N - N_{\mathrm{s}}$.

Integrating over $d\Gamma_{\mathrm{b}}$ we can write Ω as:

$$\Omega(E, N, V) = \sum_{N_{\mathrm{s}}=0}^{N} \int d\Gamma_{\mathrm{s}} d\Gamma_{\mathrm{b}} \delta[\mathcal{H}_{\mathrm{b}} - (E - \mathcal{H}_{\mathrm{s}})] = \sum_{N_{\mathrm{s}}=0}^{N} \int d\Gamma_{\mathrm{s}} \Omega_{\mathrm{b}}(E - \mathcal{H}_{\mathrm{s}}, V_{\mathrm{b}}, N - N_{\mathrm{s}})$$

If we now divide both sides by $\Omega(E, V, N)$:

$$\sum_{N=0}^{N} \int d\Gamma_{\mathrm{s}} \frac{\Omega_{\mathrm{b}}(E - \mathcal{H}_{\mathrm{s}}, V_{\mathrm{b}}, N - N_{\mathrm{s}})}{\Omega(E, V, N)} = 1$$

we recognise, in analogy to what we have seen in 3.2, that now the phase space probability density is:

$$\rho_{N_{\mathrm{s}}}(\mathbb{Q}, \mathbb{P}) = \frac{\Omega_{\mathrm{b}}(E - \mathcal{H}_{\mathrm{s}}, V_{\mathrm{b}}, N - N_{\mathrm{s}})}{\Omega(E, V, N)}$$

where the subscript N_{s} has been added in order to remember that this probability density depends on the number of particles that are in the subsystem, and the normalization condition requires that we must not only integrate ρ in $(\mathbb{Q}, \mathbb{P})$, but also sum over all the possible numbers of particles.

Now, just like we have done for the canonical ensemble we can use the fundamental postulate of statistical mechanics to write:

$$\Omega_{\mathrm{b}}(E - \mathcal{H}_{\mathrm{s}}, V_{\mathrm{b}}, N - N_{\mathrm{s}}) = e^{\frac{1}{k_B} S_{\mathrm{b}}(E - \mathcal{H}_{\mathrm{s}}, V_{\mathrm{b}}, N - N_{\mathrm{s}})}$$

and since the subsystem we are considering is again macroscopic but much smaller than the heat bath we have $\mathcal{H}_{\mathrm{s}} \ll E$ and $N_{\mathrm{s}} \ll N$, so we can expand S_{b} in the exponential:

$$\Omega_{\mathrm{b}}(E - \mathcal{H}_{\mathrm{s}}, V_{\mathrm{b}}, N - N_{\mathrm{s}}) = e^{\frac{1}{k_B} \left(S_{\mathrm{b}}(E, V_{\mathrm{b}}, N) - \mathcal{H} \frac{\partial}{\partial E} S_{\mathrm{b}}(E, V_{\mathrm{b}}, N) - N_{\mathrm{s}} \frac{\partial}{\partial N} S_{\mathrm{b}}(E, V_{\mathrm{b}}, N) + \cdots \right)}$$

From thermodynamics we know that:

$$\frac{\partial S_b}{\partial E} = \frac{1}{T} \qquad\qquad \frac{\partial S_b}{\partial N} = -\frac{\mu}{T}$$

where μ is the chemical potential, and in the end we get (removing the subscript "s" to have a more general notation):

$$\rho_N(\mathbb{Q}, \mathbb{P}) = \frac{e^{-\beta(\mathcal{H}(\mathbb{Q}, \mathbb{P}) - \mu N)}}{\sum_N \int d\Gamma e^{-\beta(\mathcal{H}(\mathbb{Q}, \mathbb{P}) - \mu N)}}$$

where:

$$\mathcal{Z} = \sum_N \int d\Gamma e^{-\beta(\mathcal{H}(\mathbb{Q}, \mathbb{P}) - \mu N)} \tag{3.16}$$

is called *grand partition function*.

Just like we have done for the canonical ensemble, we could wonder what relations does this newly found ensemble have with the ones that we already know, in particular the canonical one; what we are now going to show is that they are equivalent when the system considered is macroscopic, since this time the (relative) fluctuations of the number of particles will be incredibly small.

3.4.1 Fluctuations in the grand canonical ensemble

As we have anticipated we expect that, similarly to what happens in the canonical ensemble, the fluctuations of the energy and number of particles around their mean values vanish in the thermodynamic limit so that the grand canonical ensemble is indeed equivalent to the canonical one.

The fluctuations of the energy around $\langle \mathcal{H} \rangle$ are computed exactly in the same way as we have done in 3.3.1, and we get to the same result.
We therefore now focus on the fluctuations of the number of particles around $\langle N \rangle$:

$$\sigma_N^2 = \langle N^2 \rangle - \langle N \rangle^2 = \frac{1}{\beta^2} \frac{\partial^2 \ln \mathcal{Z}}{\partial \mu^2} = \frac{1}{\beta^2} \frac{\partial^2}{\partial \mu^2}[-\beta \Phi(T, V, \mu)]$$

where we have used the fact that $\ln \mathcal{Z} = -\beta \Phi$ (with Φ the grand potential), which will be shown in 3.4.2.
Now, from the definition $\Phi = \overline{E} - TS(\overline{E}, V, \overline{N}) - \mu \overline{N}$ of grand potential we have (considering that $\partial S/\partial V = P/T$ and that, again, there are some derivatives that vanish because $\overline{E}$ and $\overline{N}$ define a minimum) that:

$$\frac{\partial \Phi}{\partial V} = -P$$

Furthermore, Φ is extensive:

$$\Phi(T, \lambda V, \mu) = \lambda \Phi(T, V, \mu)$$

and setting $\lambda = 1/V$ we get $\Phi(T, V, \mu) = V\Phi(T, 1, \mu)$: the grand potential is proportional to the volume.
Therefore, $\partial \Phi(T, V, \mu)/\partial V = \Phi(T, 1, \mu) = -P$ and so:

$$\Phi(T, V, \mu) = V\Phi(T, 1, \mu) = -VP$$

Thus:

$$\sigma_N^2 = -\frac{1}{\beta} \frac{\partial^2}{\partial \mu^2} \Phi(T, V, \mu) = \frac{V}{\beta} \frac{\partial^2}{\partial \mu^2} P(T, \mu)$$

We must therefore find a way to express $\partial^2 P / \partial \mu^2$.

We define $v(T, \mu) = V/\overline{N}$ and $f(T, v) = F/\overline{N}$, and then write:

$$\frac{\partial P}{\partial \mu} = \frac{\partial P / \partial v}{\partial \mu / \partial v} \tag{3.17}$$

Since $\Phi = -PV = F - \mu \overline{N}$, we have:

$$f \overline{N} - \mu \overline{N} = -PV \qquad \Rightarrow \qquad f(T, v) = \mu - v P(T, \mu)$$

and since:

$$P = -\frac{\partial F}{\partial V} = -\frac{\partial f}{\partial v}$$

then:

$$f(T, v) = \mu + v \frac{\partial f}{\partial v} \qquad \Rightarrow \qquad \mu = f - v \frac{\partial f}{\partial v}$$

We therefore have:

$$\frac{\partial P}{\partial v} = -\frac{\partial^2 f}{\partial v^2} \qquad\qquad \frac{\partial \mu}{\partial v} = -v \frac{\partial^2 f}{\partial v^2} \tag{3.18}$$

and substituting in (3.17), this leads to:

$$\frac{\partial P}{\partial \mu} = \frac{1}{v}$$

If we now further derive P with respect to μ:

$$\frac{\partial^2 P}{\partial \mu^2} = -\frac{1}{v^2} \frac{\partial v}{\partial \mu} = -\frac{1}{v^2} \frac{1}{\partial \mu / \partial v}$$

and from (3.18) we have:

$$\frac{\partial \mu}{\partial v} = -v \frac{\partial^2 f}{\partial v^2} = v \frac{\partial P}{\partial v}$$

so that:

$$\frac{\partial^2 P}{\partial \mu^2} = -\frac{1}{v^2} \cdot \frac{1}{v \cdot \partial P / \partial v} = -\left(v^3 \frac{\partial P}{\partial v} \right)^{-1}$$

We have done all these intricate computations because now we can use the definition of isothermal compressibility (see 1.1.5)[23]:

$$K_T = -\frac{1}{V} \frac{\partial V}{\partial P} = -\frac{1}{V} \frac{1}{\partial P / \partial V}$$

(which as we can see is an intensive quantity) to write:

$$\frac{\partial^2 P}{\partial \mu^2} = \frac{K_T}{v^2}$$

This way the variance of the number of particles can be written as:

$$\sigma_N^2 = \frac{V}{\beta} \cdot \frac{\partial^2 P}{\partial \mu^2} = \frac{V}{\beta} \cdot \frac{K_T}{v^2} = N \frac{K_T}{\beta v}$$

(which is of course positive since $K_T > 0$).

We therefore see that:

$$\sigma_N^2 \propto N \qquad \Rightarrow \qquad \sigma_N \propto \sqrt{N}$$

---

[23]This is of course equal to what we have found, because $v = V/\overline{N}$.

and so the relative fluctuation of the number of particles is:

$$\frac{\sigma_N}{N} \propto \frac{1}{\sqrt{N}}$$

which is a result analogous to what we have seen for the energy.

Therefore, as we expected, the fluctuations of the number of particles of a system in the grand canonical ensemble are negligible in the thermodynamic limit; this ultimately means that the grand canonical ensemble is equivalent to the canonical one.

3.4.2 Grand potential

Similarly to what we have done for the canonical ensemble in 3.3.2, we now want to show how we can derive the thermodynamics of the system from the grand canonical ensemble. In particular we want to show that $\mathcal{Z}$ can be expressed in terms of the *grand potential* Φ (see 1.1.3), i.e. $\mathcal{Z} = e^{-\beta\Phi}$.

From the definition of the grand partition function, using the "trick" of inserting a $\int dE\delta(\mathcal{H} - E)$, which is of course equal to one, we get:

$$\mathcal{Z} = \sum_N \int d\Gamma dE e^{-\beta(\mathcal{H}-\mu N)}\delta(\mathcal{H} - E) = \sum_N \int dE e^{-\beta(E-\mu N)}\Omega(E, V, N)$$

and using the fundamental postulate of statistical mechanics:

$$\mathcal{Z} = \sum_N \int dE e^{-\beta(E-TS-\mu N)} = \sum_N \int dE e^{-\beta(F-\mu N)}$$

where $F = E - TS$ is the free energy of the system.
Now, F is extensive and so we can use the saddle point approximation (see appendix B) in order to compute the integral; we therefore must find when the integrand is maximized, i.e. when $F - \mu N$ is minimized.
Since for macroscopic systems N is extremely large we can treat it as a continuous function, and so the minima of the exponent are determined by the conditions[24]:

$$\frac{\partial}{\partial E}(E - TS - \mu N)_{|\overline{E},\overline{N}} = 0 \qquad\qquad \frac{\partial}{\partial N}(E - TS - \mu N)_{|\overline{E},\overline{N}} = 0$$

where $\overline{E}$ and $\overline{N}$ are the values of E and N that extremize $F - \mu N$. We therefore have:

$$1 - T\frac{\partial S}{\partial E}_{|\overline{E},\overline{N}} = 0 \qquad\qquad \left(-T\frac{\partial S}{\partial N} - \mu\right)_{|\overline{E},\overline{N}} = 0$$

namely:

$$\frac{\partial S}{\partial E}_{|\overline{E},\overline{N}} = \frac{1}{T} \qquad\qquad \frac{\partial S}{\partial N}_{|\overline{E},\overline{N}} = -\frac{\mu}{T}$$

These two equations allow us to determine $\overline{E}$ and $\overline{N}$ once T and μ are known.
Therefore, we can approximate:

$$\mathcal{Z} = e^{-\beta(\overline{E}-TS(\overline{E},V,\overline{N})-\mu\overline{N})}$$

---

[24]We don't verify that the Hessian is definite positive in the extrema that we have found because the computations are long and tedious.

since, similarly to what seen in 3.3.2, all the other terms in the exponential vanish in the thermodynamic limit. Therefore we see that:

$$-k_B T \ln \mathcal{Z} = \overline{E} - TS(\overline{E}, V, \overline{N}) - \mu \overline{N}$$

namely the exponent of the grand partition function is a Legendre transformation of the same exponent of the partition function in the canonical ensemble with respect to the number of particles N; furthermore, we have that this exponent is really the grand potential Φ if $\overline{E} = \langle \mathcal{H} \rangle$ and $\overline{N} = \langle N \rangle$. We now want to show that this is indeed the case.

From the definition (3.16) of the grand partition function, we have:

$$\langle N \rangle = \frac{1}{\beta} \frac{\partial \ln \mathcal{Z}}{\partial \mu} = \frac{\partial}{\partial \mu} (k_B T \ln \mathcal{Z}) = -\frac{\partial}{\partial \mu} \left(\overline{E} - TS(\overline{E}, V, \overline{N}) - \mu \overline{N} \right) =$$

$$= \overline{N} + \frac{\partial \overline{E}}{\partial \mu} \underbrace{\frac{\partial}{\partial E} (E - TS(E, V, N) - \mu N)_{|\overline{E}, \overline{N}}}_{=0} + \frac{\partial \overline{N}}{\partial \mu} \underbrace{\frac{\partial}{\partial N} (E - TS(E, V, N) - \mu N)_{|\overline{E}, \overline{N}}}_{=0}$$

However, the two derivatives are null because $\overline{E}$ and $\overline{N}$ are by definition minima of $F - \mu N$. Therefore:

$$\langle N \rangle = \overline{N}$$

Now, again from the definition of $\mathcal{Z}$ we have:

$$\langle \mathcal{H} - \mu N \rangle = -\frac{\partial \ln \mathcal{Z}}{\partial \beta} = \frac{\partial}{\partial \beta} \left[\beta (\overline{E} - TS(\overline{E}, V, \overline{N}) - \mu \overline{N}) \right] =$$

$$= \overline{E} - \mu \overline{N} + \frac{\partial \overline{E}}{\partial \beta} \underbrace{\frac{\partial}{\partial E} (E - TS(E, V, N) - \mu N)_{|\overline{E}, \overline{N}}}_{=0} + \frac{\partial \overline{N}}{\partial \beta} \underbrace{\frac{\partial}{\partial N} (E - TS(E, V, N) - \mu N)_{|\overline{E}, \overline{N}}}_{=0}$$

where the derivatives vanish again for the same reason. Therefore:

$$\langle \mathcal{H} - \mu N \rangle = \overline{E} - \mu \overline{N}$$

and since $\overline{N} = \langle N \rangle$ and of course $\langle \mathcal{H} - \mu N \rangle = \langle \mathcal{H} \rangle - \mu \langle N \rangle$, we have:

$$\langle \mathcal{H} \rangle = \overline{E}$$

Therefore, we indeed have:

$$\mathcal{Z} = e^{-\beta \Phi}$$

and we see that also in the grand canonical ensemble the partition function is largely dominated by the configurations of the system where the energy is $\langle \mathcal{H} \rangle$ and the number of particles is $\langle N \rangle$.

Chapter 4

Entropy and its meanings

We now stop to analyse in more detail one of the most important and nonetheless mysterious concepts that emerge within thermodynamics and statistical mechanics, and the one that had the greatest impact in all the other fields of physics: *entropy*.

With the development of the ensemble theory entropy has gained a new meaning beyond the (not really intuitive) one that it used to have in thermodynamics, namely a "measure" of the irreversibility of a process; in particular with the introduction of the fundamental postulate of statistical mechanics we can interpret entropy as a "measure" of the possible states that a system can assume.
However, entropy goes much further than that and can be interpreted in many other ways. The aim of this chapter, after briefly recalling the first interpretation of entropy given within the field of thermodynamics and how it was achieved, is to introduce these other possible interpretations and to understand them.

4.1 Entropy as irreversibility

4.1.1 Heat engines and entropy

Historically the concept of entropy was introduced in 19th century within thermodynamics from the study of thermodynamic processes. In particular the properties of reversible and irreversible transformations were analysed with the aim of applying the results to the construction of steam engines. It was realized by Carnot that the way to create the most efficient heat engine (namely a machine that can turn heat into mechanical work) was to avoid irreversibility.
In particular Carnot considered an engine made of a piston with external pressure P containing
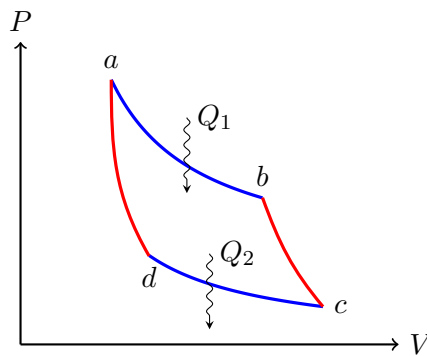
Figure 4.1: Carnot cycle
(isotherms are shown in blue while adiabatic transformations in red)

a monoatomic ideal gas, working between two heat reservoirs at temperatures $T_1 > T_2$. Since it must work between two different temperatures the engine must realize at least two isothermal transformations; since isotherms in (V, P) plane are represented by hyperbolae, in order to make a closed cycle we must also use two different transformations, for example adiabatic ones. Such a thermodynamic cycle is also called a *Carnot cycle* and is shown in figure 4.1. It works this way:

1. $a \to b$: first, the gas is put in thermal contact with the hot reservoir (the one at temperature T_1) and isothermally expanded so that it absorbs heat Q_1

2. $b \to c$: the thermal contact is then removed and the gas is further expanded (adiabatically) until it reaches temperature T_2

3. $c \to d$: the gas is then put in contact with the cold reservoir and isothermally compressed, releasing heat Q_2

4. $d \to a$ the thermal contact is then removed and the gas is adiabatically compressed until it reaches temperature T_1

Carnot then managed to show that a heat engine working with such a cycle is the most efficient possible one.

Let us also remember that according to possible different conventions the heat *absorbed* or *released* by a system or the work done *by* or *on* a system can be negative or positive; we will consider positive the heat released by a system (and therefore negative the heat absorbed).

Now, from the conservation of energy we can arrive to the thermodynamic definition of entropy. Consider in fact the first part of the cycle, the isotherm $a \to b$: in this process we add heat Q_1 to the system, which is equal to:

$$Q_1 = W_{a \to b} = \int_a^b P dV = Nk_B T_1 \ln \frac{V_b}{V_a}$$

Similarly, in $c \to d$ we release heat Q_2, which is:

$$Q_2 = -W_{c \to d} = Nk_B T_2 \ln \frac{V_c}{V_d}$$

Along the adiabatic curves, on the other hand, we have $dE = -dW = -PdV = -Nk_B T dV/V$ and also $dE = \frac{3}{2} Nk_B dT$, so that along $b \to c$ and $d \to a$ we have:

$$\frac{3}{2} \frac{dT}{T} = -\frac{dV}{V} \qquad \Rightarrow \qquad \left(\frac{T_1}{T_2} \right)^{3/2} = \frac{V_c}{V_b} = \frac{V_d}{V_a} \qquad \Rightarrow \qquad \frac{V_c}{V_d} = \frac{V_b}{V_a}$$

and thus we have:

$$\frac{Q_1}{T_1} = Nk_B \ln \frac{V_b}{V_a} = Nk_B \ln \frac{V_c}{V_d} = \frac{Q_2}{T_2}$$

Therefore, for Carnot engines we have:

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2} \tag{4.1}$$

or in other words the quantity Q/T is constant. This quantity is what has been defined as *entropy S*, or better since Q is a variation of energy Q/T is a variation of entropy:

$$\Delta S := \frac{Q}{T}$$

Let us note that if we consider a reversible engine, the *total* variation of entropy for the system composed of the engine and the two reservoirs is null. In fact, this total variation of entropy is:

$$\Delta S = \Delta S_{\text{eng.}} + \Delta S_1 + \Delta S_2$$

where $\Delta S_{\text{eng.}} = 0$ since the engine does a cyclic transformation (which means that final and initial states coincide), $\Delta S_1 = -Q_1/T_1$ and $\Delta S_2 = Q_2/T_2$; remember that Q_1 is the heat absorbed by the hot reservoir so it is negative in our conventions (because it releases heat to the engine). Since (4.1) holds, this means that $\Delta S_1 + \Delta S_2 = 0$: thus $\Delta S = 0$. Therefore the total entropy of a system is such that it remains constant in reversible transformations.

Now, what happens if we let the engine be non reversible? In this case the machine in one cycle absorbs heat $Q_1 + \delta$ and releases $Q_2 + \delta$, with $\delta > 0$ a small heat leak, and therefore we have:

$$\Delta S_1 + \Delta S_2 = -\frac{Q_1 + \delta}{T_1} + \frac{Q_2 + \delta}{T_2} = \delta \cdot \left(\frac{1}{T_2} - \frac{1}{T_1} \right) > 0$$

and so in irreversible processes the total entropy increases.

Therefore, entropy can be considered a "measure" of the irreversibility of a process.

This is what has led physicists to introduce the second principle of thermodynamics which states that the entropy of a system never decreases, namely remains constant only for reversible processes and always increases for irreversible ones.

4.1.2 Time and entropy

We conclude by noting that the irreversible increase in entropy is something that emerges only when considering macroscopic systems. In fact, if we look at a system from a microscopic point of view its properties are perfectly invariant under time reversal. More formally, if $(\mathbb{Q}, \mathbb{P})$ is the representative point of our system in phase space we know that solving Hamilton's equations for the Hamiltonian $\mathcal{H}$ of the system gives us the motion $(\mathbb{Q}(t), \mathbb{P}(t))$ of this point; however if $(\mathbb{Q}(t), \mathbb{P}(t))$ is a solution of Hamilton's equations so is $(\mathbb{Q}(-t), \mathbb{P}(-t))$[1]: in other words if we have a solution of Hamilton's equations and we reverse time we obtain another valid solution and so another physically possible evolution for the system[2].

However, this time-reversal invariance does not hold any more if we consider macroscopic systems. For example, consider the two following processes:

1. the equilibration of a cold and a hot body into two bodies with the same temperature

2. the adiabatic expansion of a gas: a box is divided by a wall into two halves of volume V, one completely empty and one containing a gas of N particles and energy E; opening a hole in the wall the gas will spontaneously fill all the volume of the box. This is clearly an irreversible process so the entropy of the system increases (while the energy of the gas, and thus its temperature, does not change)[3]

---

[1]This is true in all physical cases. One can always find "pathological" Hamiltonians for which this is not true, but generally they have no physical significance; of course we are thinking about classical systems, otherwise we can easily find systems which are not time-reversal invariant (as known, the laws of particle physics are invariant only under CPT).

[2]"Visually", we can think to watch a video representing the motion of the particles of our system: if we rewind it we obtain another possible microscopic evolution of the system.

[3]We can also see this from the microcanonical definition of entropy. In fact the phase space volume of the system in the two cases is:

$$\Omega(V) = \int_V d\mathbb{Q} \int_{\mathbb{R}^{3N}} d\mathbb{P} \delta(E - \mathcal{H}) \qquad \Omega(2V) = \int_{2V} d\mathbb{Q} \int_{\mathbb{R}^{3N}} d\mathbb{P} \delta(E - \mathcal{H})$$

and since the integrand is always positive and the domain of integration in configuration space is larger after the adiabatic expansion, we indeed have $\Omega(2V) > \Omega(V)$ and so the entropy increases in the process.

(a) Unmixed state   (b) Mixed state

Figure 4.2: Mixing of two gases

It is clear that they will never occur spontaneously in reverse, and that in order to do so we must do some kind of work on the system: two bodies in thermal contact at the same temperature will never spontaneously undergo a transformation at the end of which they will have two different temperatures (we must use some mechanical work to transfer heat), or a gas contained in a box will never fill on its own only one half of the volume (we must compress it)[4].

Therefore, entropy seen as a "measure" of irreversibility is what gives us the definition of *the arrow of time*, namely it allows us to distinguish *past* from *future*.
We will cover the relationship between entropy and time in more detail in 4.4.

4.2   Entropy as disorder

Entropy is often referred to as the "measure of the disorder" of a system; we now want to exploit this aspect.

4.2.1   Mixing of gases

The most simple example of a system where entropy can be regarded as a "measure of disorder" is the mixing of two gases, which we have already encountered in 3.2.3. Nonetheless, we now shortly re-analyse the problem from a slightly different perspective.
Consider a system divided into two subsystem with the same characteristics, namely both have volume $V/2$ and contain $N/2$ particles of total energy E (see figure 4.2). If the gases contained in the two subsystems are different then the total phase space volume of the whole system in the unmixed state will be $\Omega(E, V/2, N/2) \cdot \Omega(E, V/2, N/2)$, and thus its entropy:

$$S_{\text{un.}} = 2k_B \ln \Omega \left(E, \frac{V}{2}, \frac{N}{2} \right) = 2k_B \ln \left[\frac{(V/2)^{N/2}}{(N/2)!} \right] + 2k_B \ln \tilde{\Omega} \left(E, \frac{N}{2} \right)$$

where we have called $\tilde{\Omega}(E, N/2)$ the kinetic part of the phase space volume (which as we will shortly see is irrelevant for our purposes here).
If we now remove the wall that separates the two gases the system undergoes an irreversible transformation at the end of which the entropy will be:

$$S_{\text{mix.}} = 2k_B \ln \Omega \left(E, V, \frac{N}{2} \right) = 2k_B \ln \left[\frac{V^{N/2}}{(N/2)!} \right] + 2k_B \ln \tilde{\Omega} \left(E, \frac{N}{2} \right)$$

(where the kinetic part is left unchanged). Therefore, the variation of entropy in the process is:

$$\Delta S = S_{\text{mix.}} - S_{\text{un.}} = 2k_B \ln 2^{N/2} = N k_B \ln 2 \qquad (4.2)$$

---

[4]"Visually", as before, this means that if we are watching a video of the evolution of a macroscopic system we can always tell if the video is being rewinded or not.

We can also interpret this result as follows: every time we place an atom into one of the two halves without looking which one we choose we gain $k_B \ln 2$ in entropy. From this we could define a "counting entropy", for systems with a discrete number $N_{\text{conf.}}$ of equally-likely configurations[5]:

$$S_{\text{count.}} := k_B \ln N_{\text{conf.}}$$

Therefore, when the gases mix the entropy of the system increases: this reflects the fact that now the system has a larger accessible volume in phase space, which can be interpreted as the fact that the system is more "disordered".

Let us briefly note some facts that connect to the next section: in very intuitive words, by removing the wall we have increased our "ignorance" on the system. In fact in the initial state we had the certainty that (we still refer to figure 4.2) "black" atoms were in the left half of the system and "red" atoms in the right one, while removing the wall we don't have this certainty any more. We have therefore lost information and gained entropy; however, since $\Delta S \cdot T$ has the dimension of an energy this means that we have more available energy, with which we could do some work. Therefore, going from "low ignorance" to "high ignorance" states we can generate work[6].

If we now consider the case where the two gases are identical we know (as we have seen, even if from a different point of view, in 3.2.3) that the entropy must not increase, otherwise we would break the second principle of thermodynamics. In fact, if the entropy of the system increased if we removed the wall, inserting a new one would reduce the entropy of the whole system: this way by inserting an arbitrary amount of walls we can arbitrarily reduce the entropy of the system and therefore we would be able to generate an arbitrary amount of work from nothing; this clearly violates the principles of thermodynamics.

This time the entropy of the mixed state is:

$$S_{\text{mix.}} = k_B \ln \frac{V^N}{N!} + 2k_B \ln \tilde{\Omega}\left(E, \frac{N}{2}\right)$$

and therefore the variation of entropy is (using Stirling's approximation, see appendix C):

$$\frac{\Delta S}{k_B} = N \ln V - N \ln N + N + O(\ln N) - 2 \ln\left(\frac{V}{2}\right)^{N/2} + 2\frac{N}{2} \ln \frac{N}{2} - 2\frac{N}{2} + O(\ln N) = O(\ln N)$$

Therefore, the total variation of entropy is negligible but not properly zero[7] also in the thermodynamic limit (however, the variation of entropy per particle is null in the same limit). This is due to the fact that before removing the wall the number of particles in the two halves of the system is fixed and therefore doesn't fluctuate, while after removing it we know (see 3.2) that there are some little fluctuations; thus, if we reinsert the wall we will *never* recover the *exact* initial state but there will always be some particles in excess in one of the two halves. This is why the entropy of the system increases, even if by a perfectly negligible amount.

---

[5]Such systems arise often; a couple of examples are quantum systems (where energy states are discrete) and information theory (where every bit of a message can be found in one of two possible "states", 0 and 1).

[6]A tangible example of how this can be possible is given by semi-permeable membranes and ion pumps. The enzyme Na^+/K^+-ATPase (also known as *sodium-potassium pump*) is located on the membrane of almost every cell of our bodies, and burns ATP to maintain concentration gradients of potassium and sodium ions between a cell and its surroundings: this enzyme, therefore, does work in order to re-establish the "information" (the spatial concentration of ions) that otherwise would be lost. On the other hand semi-permeable membranes are permeable only by some particular kinds of particle and generally only in one direction; for example, suppose that the wall in figure 4.2a is a semi-permeable wall that allows only the red atoms to pass to the left: this way, there will be more particles in the left half of the system than in the right one, and therefore the right half of the system will be subjected to a pressure, the so called *osmotic pressure* (in fact, since V and T remain constant during the process, from the ideal gas law we see that the pressure must increase). In other words, the semi-permeable wall destroys the initial information and generates work.

[7]In 3.2.3 we neglected the term $O(\ln N)$ using Stirling's approximation, and that's why we found ΔS to be exactly zero.

4.3 Entropy as ignorance: information entropy

We now proceed to analyse the most general interpretation that can be given of entropy: a "measure" of our "ignorance" about a system.

In fact, all the previous interpretations of entropy can be connected to this one: a system in equilibrium maximizes its entropy because we have lost all the information about the initial conditions except for the conserved quantities; therefore maximizing the entropy means maximizing our ignorance about the details of the system. On the other hand the entropy of a mixture of different gases is a measure of the number of the possible configurations of the system, given our ignorance about it.

Therefore, entropy can be regarded as a property not of the system but of our ignorance about the system itself, represented by the ensemble of its possible configurations[8].

We have also always restricted ourselves to systems where our ignorance is maximal, namely where all the allowed configurations are equally probable; but what about systems where we have partial information and some configurations are more probable than others? As we will now see the definition of entropy can be generalized to general probability distributions and finds applications in many other fields of physics and science.

In the microcanonical ensemble we have seen that the number of allowed states for a system of energy E is the volume $\Omega(E)$ in phase space of the hypersurface of constant energy E, and that the phase space probability density is $\rho = 1/\Omega(E)$ on this surface. From this we have defined the entropy as $S = k_B \ln \Omega(E) = -k_B \ln \rho$; formally we can write[9]:

$$\frac{S}{k_B} = -\sum_{i=1}^{\Omega(E)} \rho \ln \rho = -\langle \ln \rho \rangle \tag{4.3}$$

We could therefore argue that if a system is described by a general phase space probability density $\rho(\mathbb{Q}, \mathbb{P})$ (which can also describe the system out of equilibrium, so in general ρ can explicitly depend on time) its entropy can be defined as:

$$S = -k_B \langle \ln \rho \rangle = -k_B \int \frac{d\Gamma}{h^{3N} N!} \rho(\mathbb{Q}, \mathbb{P}) \ln \rho(\mathbb{Q}, \mathbb{P}) \tag{4.4}$$

(where we are implicitly assuming that ρ has been adimensionalized so that $\ln \rho$ makes sense). We can immediately see that we recover the original definition of entropy in the microcanonical ensemble by substituting $\rho = \text{const.} = 1/\Omega(E)$:

$$S = k_B \int \frac{d\Gamma}{h^{3N} N!} \frac{\ln \Omega(E)}{\Omega(E)} = k_B \frac{\ln \Omega(E)}{\Omega(E)} \underbrace{\int \frac{d\Gamma}{h^{3N} N!}}_{\Omega(E)} = k_B \ln \Omega(E)$$

4.3.1 Shannon's entropy

From (4.3) and (4.4) we can introduce a more general definition of entropy for a generic system that can be found in Ω different discrete states[10], each with probability p_i (and $i = 1, \ldots, \Omega$).

---

[8]Note that entropy can be indeed regarded as a measure of ignorance or equivalently information, but it does not distinguish the *utility* of that information. In other words, having a lot of information from the point of view of entropy does not mean we have *useful* information about a system.

[9]As we will see in 4.5, this definition of entropy allows to derive in a systematic way the canonical and grand canonical ensembles from the microcanonical one.

[10]The Ω we are using here does not necessarily refer to the phase space volume, it's just a notation for the number of possible states.

This is known as *Shannon's entropy*, which is defined as:

$$S_S = -k_S \langle \ln p_i \rangle = -k_S \sum_{i=1}^{\Omega} p_i \ln p_i \tag{4.5}$$

the constant k_S is used instead of Boltzmann's constant because in general the connection to temperature can be irrelevant, and it is defined as $k_S = 1/\ln 2$ so that Shannon's entropy can be also rewritten as:

$$S_S = -\sum_{i=1}^{\Omega} p_i \log_2 p_i$$

The unit of measure of this entropy is the *bit*.

Shannon's entropy is particularly useful because it is the only one that satisfies three important properties, which we now show.

Theorem 4.1. *Let $p_1, \ldots, p_\Omega$ be a set of stochastic variables (they are probabilities, therefore such that $p_i \geq 0$ and $\sum_i p_i = 1$). Then Shannon's entropy S (which is a function of this set of variables) as defined in (4.5) satisfies the following properties:*

0. *It is a continuous function*

1. *$S(p_1, \ldots, p_\Omega)$ is maximized when all the p_i-s are uniform:*

$$S(p_i, \ldots, p_\Omega) < S\left(\frac{1}{\Omega}, \ldots, \frac{1}{\Omega}\right)$$

2. *S is not affected by extra states of zero probability:*

$$S(p_1, \ldots, p_\Omega, 0) = S(p_1, \ldots, p_\Omega)$$

(where on the left side S has $\Omega + 1$ arguments)

3. *S changes for conditional probabilities.*
Let $A = \{A_i\}$ and $B = \{B_k\}$ be two sets of events (with $i = 1, \ldots, \Omega$ and $k = 1, \ldots, M$), each with probability $p_i = p(A_i)$ and $q_k = p(B_k)$. If we define:

$$r_{ik} = p(A_i B_k) \qquad\qquad c_{ik} = p(A_i|B_k) = \frac{p(A_i B_k)}{p(B_k)} = \frac{r_{ik}}{q_k}$$

which, by definition, satisfy:

$$\sum_{i=1}^{\Omega} c_{ik} = 1 \qquad\qquad \sum_{i=1}^{\Omega} \sum_{k=1}^{M} r_{ik} = 1$$

then we can introduce:

$$S(AB) = -k_S \sum_{i,k} r_{ik} \ln r_{ik} \qquad\qquad S(B) = -k_S \sum_{k=1}^{M} q_k \ln q_k$$

$$S(A|B_\ell) = -k_S \sum_{i=1}^{\Omega} c_{i\ell} \ln c_{i\ell}$$

which are, respectively, the total entropy associated to the occurrence of events A and B, the entropy associated to the occurrence of events B and the entropy associated to the occurrence of

events A given B_ℓ.
Therefore, defining:

$$\langle S(A|B_\ell)\rangle = \sum_\ell q_\ell S(A|B_\ell) = -k_S \sum_{i,\ell} q_\ell c_{i\ell} \ln c_{i\ell}$$

we have that S satisfies:

$$\langle S(A|B_\ell)\rangle = S(AB) - S(B)$$

Proof.

0. It is immediate from its definition.

1. First of all, let us note that the function $f(x) := -x \ln x$ is concave since $f''(x) = -1/x$
   (and in our case $x \geq 0$ since it represents a probability).
   In general, from the definition of concave function we have:

   $$f(\lambda a + (1 - \lambda)b) \geq \lambda f(a) + (1 - \lambda)f(b) \qquad \lambda < 1 \qquad\qquad (4.6)$$

   From this we can prove by induction that:

   $$f\left(\frac{1}{\Omega}\sum_{i=1}^{\Omega} p_i\right) \geq \sum_{i=1}^{\Omega}\frac{1}{\Omega}f(p_i)$$

   This is surely true for $\Omega = 2$:

   $$f\left(\frac{p_1 + p_2}{2}\right) \geq \frac{1}{2}\left(f(p_1) + f(p_2)\right)$$

   since it follows directly from (4.6) with $p_1 = a$, $p_2 = b$ and $\lambda = 1/2$. If we now suppose
   that the inequality holds for $\Omega - 1$ probabilities:

   $$f\left(\frac{1}{\Omega - 1}\sum_{i=1}^{\Omega-1} p_i\right) \geq \frac{1}{\Omega - 1}\sum_{i=1}^{\Omega-1} f(p_i)$$

   it follows that it holds also for Ω. In fact:

   $$f\left(\frac{1}{\Omega}\sum_{i=1}^{\Omega} p_i\right) = f\left(\frac{\Omega - 1}{\Omega}\frac{1}{\Omega - 1}\sum_{i=1}^{\Omega-1} p_i + \frac{p_\Omega}{\Omega}\right)$$

   and choosing $(\Omega - 1)/\Omega = \lambda$:

   $$f\left(\frac{1}{\Omega}\sum_{i=1}^{\Omega} p_i\right) \geq \frac{\Omega - 1}{\Omega}f\left(\frac{1}{\Omega - 1}\sum_{i=1}^{\Omega-1} p_i\right) + \frac{1}{\Omega}f(p_\Omega) \geq \frac{\Omega - 1}{\Omega}\frac{1}{\Omega - 1}\sum_{i=1}^{\Omega-1} f(p_i) + \frac{1}{\Omega}f(p_\Omega)$$

   Therefore:

   $$f\left(\frac{1}{\Omega}\sum_{i=1}^{\Omega} p_i\right) \geq \frac{1}{\Omega}\sum_{i=1}^{\Omega} f(p_i)$$

   Now, considering the case $f(x) = -x \ln x$ we will have that:

   $$S(p_1, \ldots, p_\Omega) = k_S \sum_{i=1}^{\Omega} f(p_i) = k_S \Omega \frac{1}{\Omega}\sum_{i=1}^{\Omega} f(p_i) < k_S \Omega f\left(\frac{1}{\Omega}\sum_{i=1}^{\Omega} p_i\right) =$$

   $$= k_S \Omega f\left(\frac{1}{\Omega}\right) = -k_S \sum_{i=1}^{\Omega}\frac{1}{\Omega}\ln\frac{1}{\Omega} = S\left(\frac{1}{\Omega}, \ldots, \frac{1}{\Omega}\right)$$

   where the inequality is strict because in this case $f''(x) < 0$.

2. This follows immediately from the fact that $f(x)$ can be continuously extended in $x = 0$, so that $f(0) = 0$ (as we have previously stated). Therefore if we add $p_{\Omega+1} = 0$ to S, it will not contribute to S since by definition $p_{\Omega+1} \ln p_{\Omega+1} = 0$.

3. Since $r_{ik} = q_k c_{ik}$, we have:

$$S(AB) = -k_S \sum_{i,k} q_k c_{ik} \ln(q_k c_{ik}) = -k_S \sum_{i,k} q_k c_{ik}(\ln q_k + \ln c_{ik}) =$$

$$= -k_S \left(\sum_{i,k} q_k c_{ik} \ln q_k + \sum_{i,k} q_k c_{ik} \ln c_{ik} \right)$$

However:

$$\sum_{i,k} q_k c_{ik} \ln q_k = \sum_k q_k \ln q_k \sum_i c_{ik} = \sum_k q_k \ln q_k$$

and therefore:

$$S(AB) = -k_S \sum_k q_k \ln q_k - k_S \sum_{i,k} q_k c_{ik} \ln c_{ik} = S(B) + \langle S(A|B_k) \rangle$$

from which we have immediately:

$$\langle S(A|B_k) \rangle = S(AB) - S(B)$$

$\square$

Let us note an interesting consequence of the last property of the entropy: if we suppose A_i and B_k to be independent then $S(A|B_\ell) = S(A)$ and we get $S(AB) = S(A) + S(B)$: we have thus found that entropy is extensive! Therefore, the last property of S is a generalization of the extensivity of the entropy for correlated events.

Now, since these properties are a little bit obscure as we have stated them, let us make a simple example to illustrate their meaning. To make things more comprehensible, we use the same notation of the proof.
Suppose you have lost your house keys, and want to find them; you can measure your progress in finding them by measuring your ignorance with some function[11] S. Suppose there are Ω possible places A_i where you have might left the keys, each one with an estimated probability p_i; the three properties of the entropy S can thus be interpreted as follows:

1. without further information, your ignorance is maximal and the keys could be in any of the possible places, each with the same probability

2. if there is no possibility that the keys might be in a particular place A_Ω (your shoe, for example), then your ignorance is no larger than what it would have been if you had not included that place in the list of possible sites

3. to improve the research, you are likely to look where you have last seen the keys; let us call these M places B_k, each with probability q_k that the keys are indeed there.
   Before thinking where you have last seen the keys, your ignorance about their location is $S(A) = S(p_1, \ldots, p_\Omega)$ and the one about where you have last seen them is $S(B) = S(q_i, \ldots, q_M)$; therefore you also have the joint ignorance $S(AB)$. If the location where you last seen them is B_ℓ then your ignorance about the location of the keys is $S(A|B_\ell)$,

---

[11]Since it is the measure of your ignorance, this function is exactly the information entropy we have considered.

and so your combined ignorance has reduced from $S(AB)$ to $S(A|B_\ell)$.

You can now measure the usefulness of your guess by determining how much it will reduce your ignorance about where the keys are: the expected ignorance after you have guessed where the keys might have been is given by weighting the ignorance after each guess B_ℓ by the probability of that guess, namely it is $\langle S(A|B_\ell) \rangle$. The last property of the entropy thus states that after a guess your expected ignorance decreases exactly by the amount $S(B)$.

Now, what makes Shannon's entropy as defined in (4.3) so special is the fact that it is unique (of course up to a proportionality constant) and this follows only by the properties that we have just shown it satisfies.

Theorem. *Shannon's entropy, defined as:*

$$S_S = -k_S \sum_i p_i \ln p_i$$

and satisfying the properties shown in theorem 4.1, is unique up to a normalization constant. In other words, the properties shown in theorem 4.1 uniquely identify Shannon's entropy.

The idea of the proof is the following: we prove that from those properties we have that $S_S = S(\vec{p})$ has indeed the expression of Shannon's entropy when[12] $p_i \in \mathbb{Q}^+$. Then, since S is a continuous function (property 0 of theorem 4.1) we have that $S(\vec{p})$ has the same expression also for $p_i \in \mathbb{R}^+$.

Proof. Let us take $\vec{q} \in \mathbb{Q}^\Omega$, and we write its components as $q_\ell = g_\ell/g$ with $g, g_\ell \in \mathbb{N}$ (and g the least common multiple of the g_ℓ-s). We suppose $r_{k\ell} = 1/g$, so that $c_{k\ell} = r_{k\ell}/q_\ell = 1/g_\ell$ (and $k = 1, \ldots, g_\ell$). Defining $L(g) = S_S(1/g, \ldots, 1/g)$, we have:

$$S_S(AB) = S_S(r_{11}, r_{12}, \ldots, r_{\Omega M}) = L(g) \qquad S_S(A|B) = \sum_{\ell=1}^{M} q_\ell S_S(c_{1\ell}, \ldots, c_{\Omega\ell}) = \sum_\ell q_\ell L(g_\ell)$$

From the third property of S_S we have:

$$S_S(AB) = S_S(A|B) + S_S(B) \qquad \Rightarrow \qquad L(g) = \sum_\ell q_\ell L(g_\ell) + S_S(q_1, \ldots, q_M) \qquad \Rightarrow$$

$$\Rightarrow \qquad S_S(q_1, \ldots, q_M) = L(g) - \sum_\ell q_\ell L(g_\ell) \quad (4.7)$$

Therefore, if we know $L(g)$ we can express S_S (when its arguments are rational, for now). Let us see that from (4.7) we get the expression of Shannon's entropy if $L(g) = k \ln g$, with k a generic constant:

$$S_S(q_1, \ldots, q_M) = L(g) - \sum_\ell q_\ell L(g_\ell) = k \ln g - k \sum_\ell q_\ell \ln g_\ell =$$

$$= k \sum_\ell q_\ell (\ln g - \ln g_\ell) = -k \sum_\ell q_\ell \ln \frac{g_\ell}{g} = -k \sum_\ell q_\ell \ln q_\ell$$

which is exactly Shannon's entropy. We therefore must prove that $L(g) = k \ln g$, and in order to do that we will use the first and second properties of S_S.

Let us call a and b two integers such that $a < b$; then from the second property we have:

$$S_S \left(\underbrace{\frac{1}{a}, \ldots, \frac{1}{a}}_{a \text{ arg.}} \right) = S_S \left(\underbrace{\frac{1}{a}, \ldots, \frac{1}{a}}_{a \text{ arg.}}, \underbrace{0, \ldots, 0}_{b-a \text{ arg.}} \right)$$

---

[12]Note: only for this proof the symbol $\mathbb{Q}$ will be used in its usual meaning, namely the set of rational numbers.

and from the first:

$$S_S \left(\underbrace{\frac{1}{a}, \ldots, \frac{1}{a}}_{a \text{ arg.}}, \underbrace{0, \ldots, 0}_{b-a \text{ arg.}} \right) < S_S \left(\underbrace{\frac{1}{b}, \ldots, \frac{1}{b}}_{b \text{ arg.}} \right) \qquad \Rightarrow \qquad S_S(1/a, \ldots, 1/a) < S_S(1/b, \ldots, 1/b)$$

so if $a < b$ then $L(a) < L(b)$.

Let now $C^{(1)}, C^{(2)}, \ldots, C^{(n)}$ be n classes containing each g independent events with uniform probability; if we call A the set of g events in $C^{(1)}$ and B all the g^{n-1} remaining ones, from the third property of S_S we have:

$$S_S(AB) = S_S(A) + S_S(B) \qquad \Rightarrow \qquad L(g^n) = L(g) + L(g^{n-1})$$

we thus have found a recursive formula for $L(g^n)$; applying it n times we get:

$$L(g^n) = \underbrace{L(g) + L(g) + \cdots + L(g)}_{n \text{ times}} + L(1) = nL(g) + L(1)$$

and we set[13] $L(1) = 0$. Therefore, $L(g^n) = nL(g)$.

Let us now take $s \in \mathbb{N}$; there will surely be an $m \in \mathbb{N}$ for which:

$$2^m \leq s^n < 2^{m+1}$$

so (this can be done because $L(a) < L(b)$ if $a < b$):

$$L(2^m) \leq L(s^n) < L(2^{m+1}) \qquad \Rightarrow \qquad mL(2) \leq nL(s) < (m+1)L(2) \qquad \Rightarrow$$

$$\Rightarrow \qquad \frac{m}{n} \leq \frac{L(s)}{L(2)} < \frac{m+1}{n}$$

Now, the logarithm $\ln$ is a monotonically increasing function so if $a < b$ then $\ln a < \ln b$; therefore we find that a similar inequality holds also for the logarithm:

$$\frac{m}{n} \leq \frac{\ln s}{\ln 2} < \frac{m+1}{n}$$

This means that $L(s)/L(2)$ and $\ln s/\ln 2$ both belong to the interval $[m/n, (m+1)/n]$, whose width is $1/n$. Thus:

$$\left| \frac{\ln s}{\ln 2} - \frac{L(s)}{L(2)} \right| < \frac{1}{n}$$

and taking the limit $n \to \infty$ (in fact n is arbitrary, so the inequality must hold for all ns) we obtain:

$$L(s) = \frac{L(2)}{\ln 2} \ln s$$

and renaming $L(2)/\ln 2$ as k, we get:

$$L(g) = k \ln g$$

which is exactly what we wanted to prove.

Therefore, we see that the Shannon's entropy is uniquely determined by the properties shown in theorem 4.1. $\qquad \square$

---

[13]This is the entropy of a system with only one allowed configurations, so it is null; we could also have kept it but it cancels out in the computations so it is anyway irrelevant.

4.3.2   Conclusions: the most general definition of entropy

Therefore, from what we have seen the most general definition of entropy that can be given is the following: if the possible states of a system are described by a probability distribution ρ (note that we are not assuming anything on the nature of ρ, namely it can be both a continuous or discrete probability density) then the entropy S of the system is defined as:

$$S = -k \langle \ln \rho \rangle = -k \operatorname{Tr}(\rho \ln \rho) \tag{4.8}$$

where k can be k_B or k_S (depending on the kind of system considered) and the *trace* Tr is a general way, which we will often use in the future, to write the sum (intended as an effective sum or an integration, in case ρ is continuous) over all the degrees of freedom of the system. In other words Tr is a symbolic notation that includes both the cases where ρ is continuous or discrete:

$$\operatorname{Tr}(\rho \ln \rho) = \begin{cases} \sum_i \rho_i \ln \rho_i & \text{if } \rho \text{ is discrete} \\ \int d\Gamma \, \rho \ln \rho & \text{if } \rho \text{ is continuous} \end{cases}$$

4.4   Entropy and the arrow of time

Entropy as defined in (4.4) depends only on the microscopic laws of motion, which as we have seen in 4.1.2 are time-reversal invariant: this means that S in equation (4.4) is time independent and so strictly speaking the entropy of a system should never increase.
To see that explicitly let us show that in general:

$$\frac{d}{dt} \int f(\rho(\mathbb{Q}, \mathbb{P}, t)) d\Gamma = 0$$

where f is a generic function such that $f(0) = 0$. In our case $f(x) = -x \ln x$, which formally is not well defined for $x = 0$; however, since $\lim_{x \to 0} x \ln x = 0$ we can extend the function continuously and define $0 \cdot \ln 0$ to be zero.
Therefore:

$$\frac{d}{dt} \int f(\rho(\mathbb{Q}, \mathbb{P}, t)) d\Gamma = \int \frac{\partial f}{\partial \rho} \frac{\partial \rho}{\partial t} d\Gamma = -\int \frac{\partial f}{\partial \rho} \vec{\nabla} \cdot (\mathbb{V}\rho) d\Gamma$$

where in the last step we have used the fact that since ρ is a probability density it must satisfy the continuity equation $\dot{\rho} = -\vec{\nabla}(\mathbb{V}\rho)$ (see also the discussion of Liouville's theorem in D.2, appendix D), and $\mathbb{V} = (\dot{\mathbb{Q}}, \dot{\mathbb{P}})$; we now can easily see that $\vec{\nabla} \cdot (\mathbb{V}\rho) = \mathbb{V} \cdot \vec{\nabla}\rho$:

$$\vec{\nabla} \cdot (\mathbb{V}\rho) = \sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i}(\dot{q}_i \rho) + \frac{\partial}{\partial p_i}(\dot{p}_i \rho) \right] = \sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i} \left(\frac{\partial \mathcal{H}}{\partial p_i} \rho \right) - \frac{\partial}{\partial p_i} \left(\frac{\partial \mathcal{H}}{\partial q_i} \rho \right) \right] =$$

$$= \sum_{i=1}^{3N} \left[\frac{\partial \mathcal{H}}{\partial p_i} \frac{\partial \rho}{\partial q_i} - \frac{\partial \mathcal{H}}{\partial q_i} \frac{\partial \rho}{\partial p_i} \right] = \sum_{i=1}^{3N} \left(\dot{q}_i \frac{\partial \rho}{\partial q_i} + \dot{p}_i \frac{\partial \rho}{\partial p_i} \right) = \mathbb{V} \cdot \vec{\nabla}\rho$$

Therefore:

$$\frac{d}{dt} \int f(\rho(\mathbb{Q}, \mathbb{P}, t)) d\Gamma = -\int \frac{\partial f}{\partial \rho} \vec{\nabla}\rho \cdot \mathbb{V} d\Gamma = -\int \vec{\nabla} f \cdot \mathbb{V} d\Gamma$$

where by definition of gradient $\vec{\nabla} f = (\partial f / \partial \rho) \vec{\nabla}\rho$. Now, we have that $\vec{\nabla} f \cdot \mathbb{V} = \vec{\nabla}(f\mathbb{V})$, which can be shown exactly as we have done for ρ. Thus, using Gauss theorem:

$$\frac{d}{dt} \int f(\rho(\mathbb{Q}, \mathbb{P}, t)) d\Gamma = -\int \vec{\nabla}(f\mathbb{V}) d\Gamma = -\int_{\Sigma_\infty} f(\rho)\mathbb{V} \cdot d\vec{\Sigma}_\infty$$

where Σ_∞ is the surface that encloses the phase space volume[14]; since ρ is a probability density it will vanish on Σ_∞[15] and so will $f(\rho)$, since we are supposing $f(0) = 0$. Therefore the last integral is null:

$$\frac{d}{dt} \int f(\rho(\mathbb{Q}, \mathbb{P}, t)) d\Gamma = 0$$

From this we get:

$$\frac{dS}{dt} = 0$$

and note that as we have obtained it, this relation is *always* valid. This means that in principle if we consider a system undergoing an irreversible transformation, like the adiabatic expansion of a gas, its entropy remains constant; however we know that in such cases entropy always increases: where does this discrepancy come from?

What we want to show now is that this discrepancy comes from the fact that in reality *entropy is not a property of a given system, but of our knowledge about it*.
Let us first see this in a rather intuitive way: suppose we are computationally integrating the equations of motion of a system made of N particles closed in a fixed volume and to choose very unusual initial conditions, for example we set the initial positions of the particles in only one half of the system (we are thus simulating the adiabatic expansion of a gas). We let the system evolve for some time, then we stop and invert all the velocities of the particles and then restart the integration; what we are doing is essentially equivalent to letting the system evolve for some time and then "rewind" it. We would therefore expect that as the system evolves the particles will come back to their initial conditions since we are just "rewinding" the process; however this doesn't occur and the gas evolves as a normal ideal gas.
This happens because computers have *finite* precision: the position and momentum of every particle is stored with a *fixed* number of significant figures, and as time passes *we are loosing information about the system* because the computer will discard many significant figures that (mathematically) should be present. In order to actually see the gas go back to its original configuration we would thus need a computer with *infinite* precision, which would not loose information.

Let us now see this concept in a more formal way.
When we consider a system from a microscopic point of view[16] the probability density ρ has complete information on the system since it will be of the form:

$$\rho(\mathbb{Q}, \mathbb{P}, t) = \delta(\mathbb{Q} - \mathbb{Q}(t))\delta(\mathbb{P} - \mathbb{P}(t))$$

Consider now a situation where we have less information on the system: suppose for example we have a phase space probability density $\rho(\mathbb{Q}, t)$ that carries no information about the momenta of the particles[17]. Then as ρ evolves it will satisfy a diffusion-like equation in config-

---

[14]Note: in momentum space this is always a surface at infinity, while in configuration space it can also be a finite surface depending on the properties of the system (obviously, if the particles can occupy a limited space then Σ_∞ in configuration space will be finite).

[15]Intuitively, this can be justified from the fact that ρ must be normalized:

$$\int \rho(\mathbb{Q}, \mathbb{P}) d\Gamma = 1$$

and this can happen only if ρ tends to zero at infinity.

[16]Which is what we have done in the proof of the fact that $\dot{S} = 0$, since we have used Hamilton's equations.

[17]For example, this can me obtained from the previous probability density by integrating over the momenta and then renormalizing.

uration space[18]:

$$\dot{\rho}(\mathbb{Q}, \mathbb{P}) = D\nabla^2 \rho(\mathbb{Q}, \mathbb{P})$$

If the entropy of the system is:

$$S = -k_S \int \rho(\mathbb{Q}, t) \ln \rho(\mathbb{Q}, t) d\mathbb{Q}$$

then we have:

$$\frac{dS}{dt} = -k_S \int \frac{\partial \rho}{\partial t} \ln \rho d\mathbb{Q} - k_S \int \frac{\partial \rho}{\partial t} d\mathbb{Q}$$

The second integral is null:

$$\int \frac{\partial \rho}{\partial t} d\mathbb{Q} = \frac{\partial}{\partial t} \int \rho d\mathbb{Q} = \frac{\partial}{\partial t} 1 = 0$$

so integrating by parts the remaining term and using Gauss theorem:

$$\frac{dS}{dt} = -k_S \int \frac{\partial \rho}{\partial t} \ln \rho d\mathbb{Q} = -k_S D \int \nabla^2 \rho \ln \rho d\mathbb{Q} = -k_S D \int_{\Sigma_\infty} \ln \rho \vec{\nabla} \rho \cdot d\vec{\Sigma}_\infty + k_S D \int \vec{\nabla} \rho \cdot \vec{\nabla} \ln \rho d\mathbb{Q}$$

where Σ_∞ is the surface that encloses the system in configuration space. Assuming that $\rho, \vec{\nabla}\rho \to 0$ on Σ_∞, then since $\vec{\nabla} \ln \rho = \vec{\nabla}\rho/\rho$ we have:

$$\frac{dS}{dt} = k_S D \int \frac{|\vec{\nabla}\rho|^2}{\rho} \geq 0$$

Therefore, we have found that now the entropy of the system really increases, *and this follows only from our lack of knowledge about the momenta of the particles*.

We can thus see how entropy really emerges when we have not a perfect knowledge on the system, or in other words when we start ignoring or excluding some degrees of freedom.

4.5   A niftier framework for the statistical ensembles

The way we have obtained the canonical and grand canonical partition functions from the microcanonical ensemble in 3.3 and 3.4 is a rather "classic" one, and maybe also the most intuitive.

However this is not the only possible one: in fact, as we now want to show it is possible to obtain *all* the ensembles (including the microcanonical one) from the *principle of maximum entropy*, where the entropy is defined in the most general way, i.e. as in (4.8). In other words what we want to see is that maximizing the entropy of a system as defined in (4.8) with appropriately chosen constraints it is possible to determine both the canonical and grand canonical ensembles.

Let us consider a very different but simple and intuitive example to understand how this can be possible.
Suppose we have a normal six-sided die; if we know nothing about it (namely we don't know if it has been fixed or not) then all the possible rolls have the same probability, i.e. $p_i = 1/6$ for

---

[18]In fact, as we have seen in 2.1.2 the diffusion and continuity equations are equivalent if $\vec{J} = -D\vec{\nabla}\rho$: therefore since ρ satisfies a continuity equation (being a probability density) then it will also satisfy a diffusion equation with diffusion constant D.

$i = 1, \ldots, 6$. This fact can also be obtained from the maximization of Shannon's entropy (we remove any proportionality constant for simplicity):

$$S = -\sum_{i=1}^{6} p_i \ln p_i$$

with the constraint $\sum_i p_i = 1$. In fact (as it must be done for constrained optimization problems like this one) the maximization of S leads to:

$$\frac{\partial}{\partial p_j}\left[-\sum_i p_i \ln p_i - \lambda\left(\sum_i p_i - 1\right)\right] = -1 - \ln p_j - \lambda = 0 \qquad \Rightarrow \qquad p_j = e^{-\lambda - 1} := c$$

where we have simply relabelled the constant in the last step (note that p_j doesn't depend on j); therefore from $\sum_i p_i = 1$ we have exactly $p_i = 1/6$.

Now, suppose that the die has been fixed and that $p_1 = 2p_6$; in order to find the new probabilities p_i we now have to maximize S with the additional constraint $p_1 = 2p_6$. Therefore:

$$\frac{\partial}{\partial p_j}\left[-\sum_i p_i \ln p_i - \lambda\left(\sum_i p_i - 1\right) - \mu\left(p_1 - 2p_6\right)\right] = 0 \qquad \Rightarrow$$

$$\Rightarrow \quad \begin{cases} -1 - \ln p_j - \lambda = 0 & j \neq 1, 6 \\ -1 - \ln p_1 - \lambda - \mu = 0 & j = 1 \\ -1 - \ln p_6 - \lambda + 2\mu = 0 & j = 6 \end{cases} \quad \Rightarrow \quad \begin{cases} p_i = e^{-\lambda - 1} := 1/Z & i \neq 1, 6 \\ p_1 = e^{-\lambda - 1 - \mu} \\ p_6 = e^{-\lambda - 1 + 2\mu} \end{cases} \quad \Rightarrow$$

$$\Rightarrow \quad \begin{cases} p_i = 1/Z & i \neq 1, 6 \\ p_1 = e^{-\mu}/Z \\ p_6 = e^{2\mu}/Z \end{cases}$$

and requiring that $p_1 = 2p_6$ and $\sum_i p_i = 1$ we have $\mu = -\ln 2/3$ and $Z = 4 + e^{-\mu} + e^{2\mu}$. Explicitly:

$$p_i = 0.170 \quad i \neq 1, 6 \qquad\qquad p_1 = 0.214 \qquad\qquad p_6 = 0.107$$

So we see that we indeed managed to reconstruct all the probability distribution of the system only from the maximization of its entropy, with the appropriate constraints.

Let us now see this more in general: suppose we have a system which can be found in Ω different states (for simplicity we now consider the discrete case), each with probability p_i. Let us also suppose that we have put some contraints on the mean values of some observables $O^{(i)}$ defined on this system, i.e.:

$$\left\langle O^{(0)} \right\rangle = \sum_i p_i = 1 = \overline{O}^{(0)}$$

$$\left\langle O^{(1)} \right\rangle = \sum_i p_i O_i^{(1)} = \overline{O}^{(1)} \qquad \cdots \qquad \left\langle O^{(\alpha)} \right\rangle = \sum_i p_i O_i^{(\alpha)} = \overline{O}^{(\alpha)} \tag{4.9}$$

where $O_j^{(\delta)}$ are some functions depending on j (considering the previous example of the die, with our notation we have $O_i^{(1)} = 2\delta_{i1} - \delta_{i6}$) and $\overline{O}^{(\delta)}$ are some *given* values of the observables. We have put also the normalization condition in the same form as the other constraints (with $O_i^{(0)} = 1$) in order to have a more general notation.

As we know the entropy of the system will be given by (we again drop any constant in front of S):

$$S = -\sum_i p_i \ln p_i$$

Let us therefore see what happens if we maximize S with the constraints (4.9).
What we have to find is:

$$\max_{\vec{p}(\vec{\lambda})} \left[S - \lambda_0 \left(\left\langle O^{(0)} \right\rangle - \overline{O}^{(0)} \right) - \lambda_1 \left(\left\langle O^{(1)} \right\rangle - \overline{O}^{(1)} \right) - \cdots - \lambda_\alpha \left(\left\langle O^{(\alpha)} \right\rangle - \overline{O}^{(\alpha)} \right) \right]$$

where $\vec{p}(\vec{\lambda})$ is a short notation to indicate the set of the probabilities p_i seen as functions of λ_j. Therefore:

$$\frac{\partial}{\partial p_j} \left[S - \lambda_0 \left(\left\langle O^{(0)} \right\rangle - \overline{O}^{(0)} \right) - \lambda_1 \left(\left\langle O^{(1)} \right\rangle - \overline{O}^{(1)} \right) - \cdots - \lambda_\alpha \left(\left\langle O^{(\alpha)} \right\rangle - \overline{O}^{(\alpha)} \right) \right] =$$

$$= -1 - \ln p_j - \lambda_0 O_j^{(0)} - \lambda_1 O_j^{(1)} - \cdots - \lambda_\alpha O_j^{(\alpha)} = 0$$

$$\Rightarrow \qquad p_j = e^{-1 - \lambda_0 - \lambda_1 O_j^{(1)} - \cdots - \lambda_\alpha O_j^{(\alpha)}}$$

From the normalization condition we have:

$$1 = \sum_j p_j = e^{-1 - \lambda_0} \sum_j e^{-\lambda_1 O_j^{(1)} - \cdots - \lambda_\alpha O_j^{(\alpha)}}$$

If we define:

$$Z := \sum_j e^{-\lambda_1 O_j^{(1)} - \cdots - \lambda_\alpha O_j^{(\alpha)}}$$

then:

$$Z = e^{1 + \lambda_0} \qquad \Rightarrow \qquad p_j = \frac{1}{Z} e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}}$$

which has a very familiar form (the one of the canonical and grand canonical probability densities).

Now, in order to solve the problem we still have to impose all the other constraints: $\left\langle O^{(\gamma)} \right\rangle = \overline{O}^{(\gamma)}$. These can be written as:

$$\left\langle O^{(\gamma)} \right\rangle = \sum_j p_j O_j^{(\gamma)} = \frac{1}{Z} \sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}} O_j^{(\gamma)} = \frac{\sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}} O_j^{(\gamma)}}{\sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}}}$$

From the definition of Z we see that:

$$\left\langle O^{(\gamma)} \right\rangle = -\frac{1}{Z} \frac{\partial Z}{\partial \lambda_\gamma} = -\frac{\partial \ln Z}{\partial \lambda_\gamma}$$

which has exactly the same form as the equation that defines the mean value of the energy in the canonical ensemble, for example.
Therefore the values $\overline{\lambda}_\delta$ of the parameters λ_δ which maximize S with the constraints (4.9) are the solutions of the equations:

$$\overline{O}^{(1)} = -\frac{\partial \ln Z}{\partial \lambda_1}_{|\overline{\lambda}_1} \qquad \cdots \qquad \overline{O}^{(\alpha)} = -\frac{\partial \ln Z}{\partial \lambda_\alpha}_{|\overline{\lambda}_\alpha}$$

These equations are in general very difficult to solve analytically but there is a rather simple method, which we now briefly see, that allows one to determine the parameters $\overline{\lambda}_\delta$ numerically. Let us begin noting that:

$$\frac{\partial^2 \ln Z}{\partial \lambda_\alpha \partial \lambda_\gamma} = \frac{\partial}{\partial \lambda_\alpha} \left(\frac{1}{Z} \frac{\partial Z}{\partial \lambda_\gamma} \right) = \frac{\partial}{\partial \lambda_\alpha} \left\langle O^{(\gamma)} \right\rangle = \frac{\partial}{\partial \lambda_\alpha} \left(\frac{\sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}} O_j^{(\gamma)}}{\sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}}} \right) =$$

$$= \frac{\sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}} O_j^{(\gamma)} O_j^{(\alpha)}}{\sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}}} - \frac{\sum_j e^{-\sum_\delta \lambda_\delta O_j^{(\delta)}} O_j^{(\alpha)}}{Z^2} \sum_i e^{-\sum_\delta \lambda_\delta O_i^{(\delta)}} O_i^{(\gamma)} =$$

$$= \left\langle O^{(\alpha)} O^{(\gamma)} \right\rangle - \left\langle O^{(\alpha)} \right\rangle \left\langle O^{(\gamma)} \right\rangle = \left\langle \left(O^{(\alpha)} - \left\langle O^{(\alpha)} \right\rangle \right) \left(O^{(\gamma)} - \left\langle O^{(\gamma)} \right\rangle \right) \right\rangle$$

where the last term is the mean value of the product of two fluctuations: this is the *covariance* of $O^{(\alpha)}$ and $O^{(\gamma)}$. In general, the (α, γ)-th element of the *covariance matrix C* is exactly defined as the covariance between $O^{(\alpha)}$ and $O^{(\gamma)}$:

$$C^{\alpha\gamma} := \left\langle \left(O^{(\alpha)} - \left\langle O^{(\alpha)} \right\rangle \right) \left(O^{(\gamma)} - \left\langle O^{(\gamma)} \right\rangle \right) \right\rangle$$

Therefore, we have that:

$$\frac{\partial^2 \ln Z}{\partial \lambda_\alpha \partial \lambda_\gamma} = C^{\alpha\gamma}$$

The covariance matrix is positive (semi)definite; in fact if $\vec{x}$ is a generic vector, then:

$$\vec{x} \cdot C\vec{x} = \sum_{\alpha,\gamma} C^{\alpha\gamma} x_\alpha x_\gamma = \left\langle \left[\sum_\delta x_\delta \left(O^{(\delta)} - \left\langle O^{(\delta)} \right\rangle \right) \right]^2 \right\rangle \geq 0$$

We needed these observations because if we now define the function:

$$F = \ln Z + \sum_{\delta=1}^{\alpha} \lambda_\delta \overline{O}^{(\delta)}$$

then since $-\partial \ln Z / \partial \lambda_\delta{}_{|\overline{\lambda}_\delta} = \overline{O}^{(\delta)}$ we have that:

$$\frac{\partial F}{\partial \lambda_\delta}{}_{|\overline{\lambda}_\delta} = 0$$

i.e. F has an extremum in $\overline{\lambda}_\delta$. However:

$$\frac{\partial^2 F}{\partial \lambda_\alpha \partial \lambda_\gamma}{}_{|\overline{\lambda}_\alpha, \overline{\lambda}_\gamma} = C^{\alpha\gamma} \geq 0$$

and so this extremum is a minimum: F is minimized by the values $\overline{\lambda}_\delta$ of the parameters λ_δ which maximize the entropy S of the system.
Therefore, in this way we can simply determine the values $\overline{\lambda}_\delta$ by finding the minima of F, which is a rather straightforward computational problem.

Let us now briefly see what happens in the continuous case, so that we can use what we have seen in the framework of ensemble theory.
Since we are now dealing with continuous probability densities ρ, they will not depend on the

discrete index i but on the "continuous index" $(\mathbb{Q}, \mathbb{P})$, and of course the summations over i must be replaced with integrations in phase space. In other words, the entropy of the system will be:

$$S = -\int d\Gamma \rho(\mathbb{Q}, \mathbb{P}) \ln \rho(\mathbb{Q}, \mathbb{P})$$

and the contraints are:

$$\left\langle O^{(0)} \right\rangle = \int d\Gamma \rho(\mathbb{Q}, \mathbb{P}) = 1 = \overline{O}^{(0)}$$

$$\left\langle O^{(1)} \right\rangle = \int d\Gamma \rho(\mathbb{Q}, \mathbb{P}) O^{(1)}(\mathbb{Q}, \mathbb{P}) = \overline{O}^{(1)} \quad \cdots \quad \left\langle O^{(\alpha)} \right\rangle = \int d\Gamma \rho(\mathbb{Q}, \mathbb{P}) O^{(\alpha)}(\mathbb{Q}, \mathbb{P}) = \overline{O}^{(\alpha)}$$

The probability density will be of the form:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{1}{Z} e^{-\sum_\delta \lambda_\delta O^{(\delta)}(\mathbb{Q}, \mathbb{P})} \qquad\qquad Z = \int e^{-\sum_\delta \lambda_\delta O^{(\delta)}(\mathbb{Q}, \mathbb{P})} d\Gamma$$

and the values $\overline{\lambda}_\delta$ of the parameters λ_δ which maximize S will be again the solutions of the equations:

$$\overline{O}^{(1)} = -\frac{\partial \ln Z}{\partial \lambda_1}_{|\overline{\lambda}_1} \qquad \cdots \qquad \overline{O}^{(\alpha)} = -\frac{\partial \ln Z}{\partial \lambda_\alpha}_{|\overline{\lambda}_\alpha}$$

Let us now apply all this to the ensemble theory.

In the microcanonical ensemble we only have the normalization constraint:

$$\int \rho(\mathbb{Q}, \mathbb{P}) d\Gamma = 1$$

where the integration is done over the $(\mathbb{Q}, \mathbb{P})$ points that satisfy $\mathcal{H}(\mathbb{Q}, \mathbb{P}) = E$, with $\mathcal{H}$ the Hamiltonian of the system and E a *given* value of the energy. In this case, therefore, the only non-null "observable" is $O^{(0)}$, which as we have seen is a "fictious" one (defined so that also the normalization condition can be put in the form of a constraint on the mean value of a given observable). In other words, referring to our notation we have $\alpha = 0$ and the probability density has indeed the form:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{1}{\Omega(E)} = \text{const.}$$

where we have called $\Omega(E)$ the normalization factor.

The value of Z can be obtained intuitively as we have done in 3.2, i.e. since ρ must be zero everything but on the phase space hypersurface of constant energy E, whose volume is $\Omega(E)$, then:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{\delta(\mathcal{H}(\mathbb{Q}, \mathbb{P}) - E)}{\Omega(E)}$$

(see also the footnote on page 37 for a more "formal" way to derive this).

In the canonical ensemble we have a new constraint, i.e. we require the mean value of the energy to be fixed:

$$\langle \mathcal{H} \rangle = \int \rho(\mathbb{Q}, \mathbb{P}) \mathcal{H}(\mathbb{Q}, \mathbb{P}) d\Gamma = \overline{E}$$

With our previous notation we have $\alpha = 1$ and $O^{(1)}(\mathbb{Q}, \mathbb{P}) = \mathcal{H}(\mathbb{Q}, \mathbb{P})$, so that:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{1}{Z} e^{-\lambda_1 \mathcal{H}(\mathbb{Q}, \mathbb{P})} \qquad\qquad Z = \int e^{-\lambda_1 \mathcal{H}(\mathbb{Q}, \mathbb{P})} d\Gamma$$

where[19] $\lambda_1 = \beta$.

In the grand canonical ensemble, then, we have the additional constraint of having the mean value of the number of particles fixed, namely $O^{(2)} = N$. Explicitly we have that the entropy of the system is:

$$S = -\sum_N \int d\Gamma_N \rho_N(\mathbb{Q}, \mathbb{P}) \ln \rho_N(\mathbb{Q}, \mathbb{P})$$

and the constraints are:

$$\sum_N \int \rho_N(\mathbb{Q}, \mathbb{P}) d\Gamma_N = 1 \qquad \sum_N \int \rho_N(\mathbb{Q}, \mathbb{P}) \mathcal{H}_N(\mathbb{Q}, \mathbb{P}) d\Gamma_N = \overline{E}$$

$$\sum_N \int \rho_N(\mathbb{Q}, \mathbb{P}) N d\Gamma_N = \overline{N}$$

In this case, we will have:

$$\rho_N(\mathbb{Q}, \mathbb{P}) = \frac{1}{\mathcal{Z}} e^{-\lambda_1 \mathcal{H}_N(\mathbb{Q}, \mathbb{P}) - \lambda_2 N} \qquad \mathcal{Z} = \sum_N \int e^{-\lambda_1 \mathcal{H}_N(\mathbb{Q}, \mathbb{P}) - \lambda_2 N} d\Gamma_N$$

where $\lambda_1 = \beta$ and $\lambda_2 = -\mu\beta$.

We conclude with an observation.

We have determined the properties of the ensembles fixing the values of the first moments of the observables (i.e., $\mathcal{H}$ and N); we can ask: why haven't we fixed also other moments (namely $\mathcal{H}^2$, N^2 etc.)?

In general it can happen that those additional moments are redundant; let us see a simple example in order to understand that.

Suppose x is a stochastic variable distributed along a probability distribution $\varrho(x)$; imagine that we are given Ω values x_i of x without knowing $\varrho(x)$ and that we want to understand what $\varrho(x)$ is from the x_i-s. How can we proceed? We could try to guess ϱ with a procedure similar to what we have seen now. For example we could compute the n-th moments of ϱ with $n = 1, 2, 3$, namely $\langle x \rangle$, $\langle x^2 \rangle$ and $\langle x^3 \rangle$. Then, our guess for ϱ would be:

$$\varrho(x) = \frac{1}{Z} e^{-\lambda_1 x - \lambda_2 x^2 - \lambda_3 x^3} \qquad Z = \int e^{-\lambda_1 x - \lambda_2 x^2 - \lambda_3 x^3} dx$$

and the values $\overline{\lambda}_i$ of λ_i which give the correct expression of ϱ are given by the solutions of:

$$\frac{\partial \ln Z}{\partial \lambda_n}\bigg|_{\overline{\lambda}_n} = \overline{O}^{(n)}$$

where $\overline{O}^{(n)}$ is computed from the given set of x_i:

$$\overline{O}^{(n)} = \frac{1}{\Omega} \sum_i x_i^n$$

If we determine $\overline{\lambda}_i$ with and increasing number Ω of data, we expect (or at least hope) that the parameters λ_i will tend to some definite values; what happens is that they often tend to zero, when $i \geq 2$.

---

[19] At this point there is no way to understand that, and we are about to see something similar also in the grand canonical ensemble. This is the disadvantage of this way of deducing the statistical ensembles: it is elegant and mathematically consistent, but not really physically intuitive. The "classical" way we have used to derive the ensembles is surely less formal and rigorous, but it allows to understand *physically* what happens.

For example, if in reality $\varrho(x) = e^{-x/x_0}/x_0$ for $x \geq 0$ then repeating the computations with higher values of Ω we would find $\lambda_1 \to 1/x_0$ and $\lambda_2, \lambda_3 \to 0$: the second and third moments of x are useless if we want to determine $\varrho(x)$.

Let us note, however, that the use of the moments of x can be useless in some cases: if in fact $\varrho(x) = 1/x^2$ for $x \geq 1$ then we will never be able to express it as a product of exponentials, so the parameters λ_i will not tend to definite values. What can we do in this case? We can use the moments of $\ln x$ instead of x; in fact if we compute for example $\langle \ln x \rangle = \overline{O}^{(1)}$, then our guess for $\varrho(x)$ will be:

$$\varrho(x) = \frac{1}{Z}e^{-\lambda_1 \ln x} = \frac{1}{Z} \cdot \frac{1}{x^{\lambda_1}}$$

and so in this case we would expect $\lambda_1 \to 2$; we also see that if we included come higher moments of $\ln x$, their relative parameters would have all gone to zero, and so all the n-th moments with $n \geq 2$ are redundant.

Therefore, we see that depending on the kind probability distribution we use different "recipes"[20] in order to determine $\varrho(x)$.

However, in ensemble theory something slightly different happens. Suppose in fact that we have fixed the first two moments of $\mathcal{H}$ in the canonical ensemble; then we would have:

$$\rho(\mathbb{Q}, \mathbb{P}) = \frac{1}{Z}e^{-\lambda_1 \mathcal{H} - \lambda_2 \mathcal{H}^2}$$

Now the energy is extensive, so $\mathcal{H} \propto N$ and thus the leading term in the exponential should be $\mathcal{H}^2$; in general if we fixed an arbitrary number of moments of $\mathcal{H}$, the leading one is the last. This, however, doesn't really make sense from a physical point of view since it would imply that the only significant contribution is that of the n-th moment, with $n \to \infty$.

Therefore, in the case of statistical mechanics we are (although implicitly) *assuming* that the moments different from the first are actually insignificant, since strictly speaking there is nothing that would prevent us from fixing also their values. It is the incredible accuracy of the predictions made by statistical mechanics with the experimental results that confirms that this is a reasonable assumption.

---

[20]It could be asked then what can we do if we don't know *absolutely nothing* about ϱ. In this case there is nothing that can help, besides experience; in such cases, in fact, one tries to get some insights on the problems and then try different "recipes", from which something new can be learned about $\varrho(x)$.

Chapter 5

Statistical mechanics of phase transitions

5.1 Introduction

In chapter 3 we have seen how the tools of statistical mechanics allow us to interpret and explain in a new and elegant way the statistical "origin" of thermodynamics, and how to derive all the interesting properties of a thermodynamic system. Very briefly, in general terms the whole philosophy of what we have seen can be summarized as follows: given a *finite-sized* system, in general its Hamiltonian $\mathcal{H}$ can be written in the form:

$$\mathcal{H} = -k_B T \sum_n K_n O_n$$

where K_n are called *coupling constants* (which generally are external parameters that can be controlled experimentally) and O_n *local operators*, which are combinations (normally linear or quadratic, but in general they can be any function) of the degrees of freedom of the system considered (such as the positions and momenta of the particles in a gas, for example).
Then, we define the (canonical) *partition function* of a system as:

$$Z[K] = \operatorname{Tr} e^{-\beta \mathcal{H}}$$

where the *trace* Tr is a general way to express the sum (or integral, depending on the discrete or continuous nature of the system) over all the degrees of freedom, and then the free energy is defined as:

$$F[K] = -k_B T \ln Z[K]$$

The thermodynamic properties of the system can be obtained taking appropriate derivatives of $F[K]$ *once the thermodynamic limit has been taken*.

We now stop for a moment in order to study this concept, since we have not seen it explicitly before.
We know that the free energy of a system is an extensive quantity; if we call L a characteristic length of our system and d its dimensionality we will have that the volume V and the surface S of the system will be proportional to appropriate powers of L:

$$V \propto L^d \qquad\qquad S \propto L^{d-1}$$

Therefore we expect that for a finite system:

$$F[K] = V f_b[K] + S f_s[K] + O(L^{d-2})$$

where f_b is the *bulk free energy density* and f_s the *surface free energy density*. The *thermodynamic limit* of the bulk free energy density is defined as:

$$f_b[K] = \lim_{V \to \infty} \frac{F[K]}{V}$$

when this limit exists and is independent of the region where we have defined our system. Similarly, the thermodynamic limit of the surface free energy density is defined as:

$$f_s[K] = \lim_{S \to \infty} \frac{F[K] - V f_b[K]}{S}$$

when this limit exists and is independent of the region where the system is defined.

Of course, sometimes some other constraints must be put in order to take meaningful limits: for example, for a fluid system the limit $V \to \infty$ *per se* would be rather unreasonable unless we simultaneously take the limit $N \to \infty$ so that the density N/V of the system remains constant. The existence of the thermodynamic limit for a system is absolutely not trivial, and its proof can sometimes be really strenuous. In particular, it can be shown that in order for a thermodynamic limit to exist the forces acting between the degrees of freedom of the system must satisfy certain properties, for example being *short ranged*.

For example, if a d-dimensional system is made of particles that interact through a potential of the form:

$$\varphi(\vec{r}) = \frac{1}{r^\sigma}$$

where $r = |\vec{r}|$ is the distance between two particles, then it can be shown that the thermodynamic limit of the system exists if $d > \sigma$.

5.1.1 Statistical mechanics and phase transitions

Now, in all the cases considered in the previous chapters we have *never* taken into account the possibility for our system to undergo a phase transition. We can therefore wonder if all the machinery that we have built can actually be used also when a thermodynamic system changes phase.

In other more "philosphical" words the issue can be reformulated as follows: is statistical mechanics a tool that allows us to derive *the whole thermodynamics* of a system, *phase transitions included*, or it works only when we limit ourselves to single phases?

It would be rather unsatisfying if the latter was the case, because after all the Hamiltonian (and thus the partition function) of a system does *not* change during a phase transition, since the interaction potential between the particles does not change with the temperature.

As we have seen in 1.2, phase transitions are characterized by singularities (jump discontinuities or divergences) in the derivatives of thermodynamic potentials, so if we want the partition function of a system to include phase transitions it must exhibit such singularities. However, by definition the partition function is simply a sum of exponentials, so it should be itself an analytic function and thus incapable of reproducing phase transition-like phenomena.

This is true as soon as we consider *finite-sized* systems: in fact, in this case the partition function is a *finite* sum of exponential, so it is inevitably an analytic function. However, *infinite* sums of analytic functions can be non analytic, so we can suppose that the partition function can exhibit singularities (and thus phase transitions) only in the thermodynamic limit.

This is indeed what happens; in fact, it can be shown[1] that if one allows $e^{\beta\mu}$ (with μ the chemical potential) to assume complex values in the thermodynamic limit at least one of the poles of the partition function moves on the real axis.

---

[1] See C. N. Yang and T. D. Lee, *Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation*, Physical Review, vol. 87, 3.

Heuristic energy-entropy argument

We could now ask how can we understand if a system can undergo phase transitions at all; a simple but useful tool to do so is the so called *energy-entropy argument*: at high temperatures the entropy S will be the dominant term in the free energy $F = U - TS$ of the system, and the free energy is thus minimized by maximizing S; at low temperatures on the other hand the internal energy U can be the most important contribution to F and so the free energy is minimized by minimizing U. Therefore if maximizing S for high temperatures and minimizing U for low ones brings to two different macroscopic equilibrium configurations of the system, we can conclude that there must be at least one phase transition between $T = 0$ and $T = \infty$ (of course this requires that we know the exact expressions of U and S).

5.1.2 Long range correlations

We now want to show that one of the characteristic traits of critical transitions, that distinguishes them from other kinds of phase transitions, is the fact that the divergence of response functions in the proximity of the critical point is intimately bound to the existence of *long-ranged strongly correlated microscopic fluctuations*.
We shall do that in a very qualitative way, and to make things clearer we consider a magnetic system (but of course our considerations are absolutely general and extensible to other kinds of systems).

When a magnetic field H is present[2], the partition function of a magnet is:

$$Z = \text{Tr}\, e^{-\beta(\mathcal{H}-HM)}$$

where $\mathcal{H}$ is the Hamiltonian of the system. The magnetization of the system at equilibrium is:

$$\langle M \rangle = \frac{1}{Z} \text{Tr}\left[M e^{-\beta(\mathcal{H}-HM)}\right] = \frac{\partial \ln Z}{\partial(\beta H)}$$

while its susceptibility is:

$$\chi_T = \frac{\partial \langle M \rangle}{\partial H} = -\frac{1}{Z^2} \text{Tr}\left[\beta M e^{-\beta(\mathcal{H}-HM)}\right] \cdot \text{Tr}\left[M e^{-\beta(\mathcal{H}-HM)}\right] + \frac{1}{Z}\text{Tr}\left[\beta M^2 e^{-\beta(\mathcal{H}-HM)}\right] =$$

$$= \frac{\beta}{Z}\text{Tr}\left[M^2 e^{-\beta(\mathcal{H}-HM)}\right] - \frac{\beta}{Z^2}\left\{\text{Tr}\left[M e^{-\beta(\mathcal{H}-HM)}\right]\right\}^2 = \frac{1}{k_B T}\left(\langle M^2\rangle - \langle M\rangle^2\right)$$

However, from a microscopic point of view the magnetization of the system can be written as:

$$M = \int m(\vec{r})d\vec{r}$$

where $m(\vec{r})$ is the local magnetization in $\vec{r}$. If we substitute this expression in χ_T we get:

$$\chi_T = \beta \int \left(\langle m(\vec{r})m(\vec{s})\rangle - \langle m(\vec{r})\rangle \langle m(\vec{s})\rangle\right) d\vec{r}d\vec{s} \tag{5.1}$$

If our system is spatially homogeneous, then $\langle m(\vec{r})\rangle = m$ is constant and $\langle m(\vec{r})m(\vec{s})\rangle := G(\vec{r}-\vec{s})$, called *correlation function*, depends only on the relative distance between two points. Defining the *connected correlation function* as:

$$\langle m(\vec{r})m(\vec{s})\rangle_c := \langle m(\vec{r})m(\vec{s})\rangle - \langle m(\vec{r})\rangle \langle m(\vec{s})\rangle = G(\vec{r}-\vec{s}) - m^2$$

---

[2]We are treating H, just like the magnetization M, as scalars instead as vectors: in order to make things easier, we are supposing that the *real* fields $\vec{H}$ and $\vec{M}$ are both directed along the same direction, say the z axis of our reference frame, and we are only considering their magnitudes.

sometimes also called $G_c(\vec{r} - \vec{s})$, and changing variable to $\vec{x} = \vec{r} - \vec{s}$ we get:

$$\chi_T = \beta \int G_c(\vec{r} - \vec{s}) d\vec{r} d\vec{s} = \beta \int G_c(\vec{x}) d\vec{x} d\vec{s} = \frac{V}{k_B T} \int G_c(\vec{r}) d\vec{r} \qquad (5.2)$$

where in the last step we have renamed the variable $\vec{x}$ to $\vec{r}$.

The connected correlation function is essentially (as it can be understood from its definition) a measure of how the fluctuations of the magnetization from the mean value m in a part of the system influence those in another part of the system.

Now, we expect that in general the correlation function will decrease on long distances. We can therefore write[3]:

$$G_c(\vec{r}) \sim e^{-\frac{|\vec{r}|}{\xi}} \qquad \text{at least when } |\vec{r}| > \xi$$

where ξ is a characteristic length of the system called *correlation length*.

If we call g the characteristic value of the connected correlation function for $|\vec{r}| < \xi$ (namely we suppose that it is constant and equal to g for lengths smaller than ξ[4]), then from (5.2) we have (neglecting any proportionality constant):

$$\frac{\chi_T}{\beta V} = \int G_c(\vec{r}) d\vec{r} < \int_{|\vec{r}| < \xi} G_c(\vec{r}) d\vec{r} = \int_{|\vec{r}| < \xi} g d\vec{r} \propto g \xi^3$$

namely:

$$\frac{\chi_T}{\beta V} < g \xi^3$$

We thus immediately see that if $T \to T_c$, since $\chi_T \to \infty$ (as we know from the thermodynamics of critical phase transitions) we must also have $\xi \to \infty$: in the neighbourhood of a critical point the correlation length of the system diverges.

As we have seen in 1.5, we describe the divergence of ξ by means of the critical exponent ν, namely we set:

$$\xi \sim |t|^{-\nu}$$

Furthermore, experiments and exact solutions of simple models show that the correlation function decays near the critical point as a power of $|\vec{r}|$, with a critical exponent called η:

$$G(\vec{r}) \sim \frac{1}{|\vec{r}|^{d-2+\eta}}$$

where d is the dimensionality of the system (see 1.5).

5.1.3   Finite size effects on phase transitions

We have just seen that phase transitions can occur only in the thermodynamic limit. However, this is only a mathematical concept and can never be *really* reached: real systems, regardless of how much big they are, are *not* infinite! We can therefore ask if what we have just seen is useful at all, namely if and how much real systems differ from their thermodynamic limit.

Consider the correlation length of a thermodynamic system: we know that it depends on the temperature of the system, and that it diverges in the neighbourhood of a critical point. Let us suppose that our system is finite, and call L the length of its size; of course the correlation length ξ of the system cannot be greater that L itself and so for temperatures near enough to

---

[3]This form is also justified, as we will see, from the fact that in many cases G_c turns out to actually decay exponentially with $\vec{r}$.

[4]In fact, e^{-x} is of the order of 1 when $x < 1$.

the critical one the behaviour of the system will differ from that of its thermodynamic limit. Let us see with a realistic example how big this difference is; suppose that:

$$\xi \sim \xi_0 t^{-2/3}$$

where $t = (T - T_c)/T_c$ is the reduced temperature and $\xi_0 \approx 10$ Å is the correlation length far from the critical point. This form for ξ is realistic for fluids and magnets, and ξ_0 is an overestimate in many real cases.

If we suppose $L = 1$ cm, we have $\xi = L$ when $t \approx 10^{-11}$: this means that we should be able to measure temperatures with a precision of one part in 10^{11} in order to detect deviations from the thermodynamic limit!

We therefore see that even if in principle real systems behave differently from their thermodynamic limits, these differences are extremely small and negligible in all reasonable experimental conditions.

We are thus legitimated to use thermodynamic limits in statistical mechanics in order to study phase transitions of real macroscopic systems.

5.2 Models in statistical mechanics

Models are the main instrument of statistical mechanics that allow to "translate" the physical configuration of a system in a mathematical language, so that it can be thoroughly studied and its behaviour predicted.

The role of models in statistical mechanics, and more in general the role of models in the whole scientific method, is of course a very intriguing and philosophically demanding one; of course we will not cover such a vast and difficult issue, and we will only limit ourselves to consider the two diametrically opposing points of view about how models are used.

The "traditional" point of view has been to describe as much faithfully as possible a system, including all its details; in case the theory is unable to explain the experimental results, then some parameters of the system are fine-tuned, or additional parameters are included.

On the other hand a more modern point of view, born in the framework of statistical mechanics and motivated by the study of phase transitions and critical phenomena, is that of describing a system with the most minimal and simple possible model, introducing few parameters or properties eventually motivated by symmetry or very general arguments. In such systems in fact it is not necessary to introduce deep levels of details in order to understand their phenomenology, and the study of such minimal models has often led to answers that turned out to be even more universal that those that were looked for when the model was introduced.

The use of this more minimalistic approach has been ultimately justified with the introduction of the concepts of universality (see 1.5) and Renormalization Group (see chapter 8).

We are now going to introduce some very important models in statistical mechanics: their study will allow us to learn a lot about the statistical mechanics of phase transitions.

5.3 The Ising model

5.3.1 Introduction

The Ising model (in its one-dimensional version) was proposed by Ernst Ising in his PhD thesis in 1925 as a tool to describe the thermodynamic properties of magnetic systems from a microscopic point of view. Ising found (as we will shortly see) that in the case he considered the

(a) Ising model in one dimension

(b) Ising model in two dimensions

(c) Ising model in three dimensions
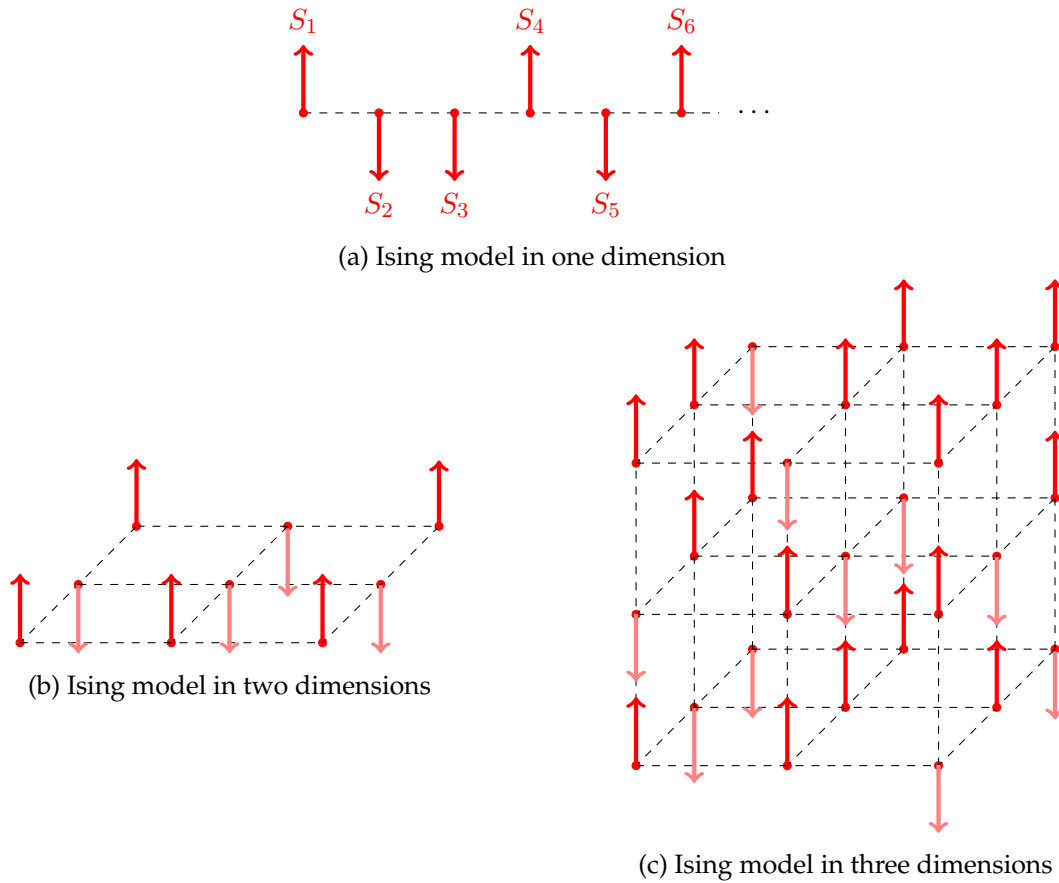
Figure 5.1: Representation of the Ising model in various dimensions (the spins pointing dow-nards in two and three dimensions have been coloured differently to be better identified)

system does not exhibit any phase transition for $T > 0$, and he incorrectly concluded that the whole model was not useful to describe such systems.

However, this model has been later studied again and in different configurations and many important properties have been discovered; historically Ising model has been one of the most (if not the most itself) heavily studied model in statistical mechanics and it is often used as a testing ground when new theories or methods are developed.

Another extremely important characteristic of the Ising model, that also justifies the effort we will make in order to study it, is that it does not only apply to magnetic systems: as we will see many other systems[5] can be shown to be equivalent to an appropriately defined Ising model.

The d-dimensional Ising model is defined as follows: consider a d-dimensional lattice with N sites, each labelled by the index $i = 1, \ldots, N$; in general the lattice is supposed to be hypercubic, but this is not necessary: in two dimensions, for example, we can consider triangular or "honeycomb" lattices, while in three dimensions we can have body-centered or face-centered cubic lattices[6]. What distinguishes one lattice from another is its *coordination number z*, defined as the number of the nearest neighbours of a site. In the case of hypercubic lattices it can be easily seen (case by case) that $z = 2d$, where d is the dimensionality of the system.

The degrees of freedom of the model are discrete variables S_i defined on each site that can only

---

[5]Examples of such systems are fluids (see 5.3.4), binary alloys (see 5.3.6) and neural networks (see 5.3.7).

[6]The number of possible lattices is predicted by group theory, and amazingly *all* these mathematically possible lattices are found in nature.

assume the values $+1$ and -1; therefore, the number of the possible configurations of the system is 2^N.

In the original purpose of the Ising model the lattice represents the atomic lattice of a metal and the variables S_i are the spins of the atoms, or rather their component along the vertical axis[7]; therefore, $S_i = +1$ corresponds to a spin pointing upwards while $S_i = -1$ downwards (see figure 5.1), and the study of this model should determine if and how all these spins can align so that the system can have a spontaneous net magnetization. However, since the Ising model can be used to describe completely different systems this interpretation is not the only possible one; since this model has always been associated to magnets (but also for historical reasons) we will in the following continue to use a terminology proper only to magnetic systems.

The usefulness of the Ising model (but in general of lattice theories) goes much further that what can now be imagined (considering also the fact that this model can be used to describe systems different from magnetic ones).
Lattice theories are in fact widely used in many areas of physics: just as an example, apart from "easily imaginable" applications of the Ising model in solid state physics, also quantum relativistic theories can be formulated in terms of lattices. QCD, for example, widely uses four-dimensional lattice models (with Minkowski spatio-temporal metric, of course) to describe quantum phenomena. We therefore see that also models with "strange" dimensionality (i.e. greater than $d = 3$) can be actually useful.

In order for this model to be interesting, the degrees of freedom S_i must not be independent[8]: we therefore assume that the spins interact with each other with exchange interactions that couple in general an arbitrary number of spins, and also with an external field H that can change from site to site.
Therefore, the most general form of the Hamiltonian of the Ising model for a given spin configuration is:

$$\mathcal{H} = -\sum_i H_i S_i - \sum_{i,j} J_{ij} S_i S_j + \sum_{i,j,k} K_{ijk} S_i S_j S_k + \cdots$$

The first two minus signs are present because in general two adjacent magnetic moments tend to align, so for the system it is energetically convenient to have as many aligned spins (and aligned along the external field) as possible.
For such systems the trace operator, being the sum over all the possible values of the degrees of freedom, takes the form:

$$\mathrm{Tr} = \sum_{\{S_i = \pm 1\}} = \sum_{S_1 = \pm 1} \sum_{S_2 = \pm 1} \cdots \sum_{S_N = \pm 1}$$

As usual, the partition function will be:

$$Z(T, \{H_i\}, \{J_{ij}\}, \dots) = \mathrm{Tr}\, e^{-\beta \mathcal{H}}$$

and the finite-size free energy:

$$F(T, \{H_i\}, \{J_{ij}\}, \dots) = -k_B T \ln Z(T, \{H_i\}, \{J_{ij}\}, \dots)$$

Then, the thermodynamic properties of the system can be obtained taking the thermodynamic limit:

$$f(T, \{H_i\}, \{J_{ij}\}, \dots) = \lim_{N \to \infty} \frac{1}{N} F(T, \{H_i\}, \{J_{ij}\}, \dots)$$

---

[7]Note that from quantum mechanics we know that in general the spin of an atom is proportional to a given fraction of $\hbar$: we are "ignoring" the magnitude of the spin, "reabsorbing" it in the definition of the coupling constants.

[8]This is not really true: as we will see in 5.3.5, we can learn something interesting also in the non-interacting case.

and then appropriately deriving f.

For the sake of simplicity, from now on we will always neglect interactions that couple more than two spins, and we will also consider the field H as constant over the system[9].

Now, what about the existence of the thermodynamic limit for the Ising model?
It has been shown that in general the thermodynamic limit exists if the two-spin interaction J_{ij} satisfies:

$$\sum_{i \neq j} |J_{ij}| < \infty$$

Therefore, we can see that what determines the existence of the thermodynamic limit is the dimensionality of the system and the range of the interactions. For example, if the interaction between two spins S_i and S_j at the positions $\vec{r}_i$ and $\vec{r}_j$ is of the form:

$$J_{ij} = A|\vec{r}_i - \vec{r}_j|^{-\sigma}$$

then it can be shown that we must have $\sigma > d$ in order for the thermodynamic limit to exist.

5.3.2   Analytic properties of the Ising model

Since we want to analyse the Ising model in order to determine if and how it exhibits phase transitions, and that phase transitions are characterized by singularities in the thermodynamic potentials, it is important to study the analytic properties of the Ising model's free energy so that we can find possible non-analytic behaviours and therefore understand when such phenomena can occur.

Let us therefore consider an Ising model where the external field H is constant and the interaction J_{ij} occurs only between nearest neighbouring spins, namely:

$$J_{ij} = \begin{cases} J & S_i \text{ and } S_j \text{ are nearest neighbours} \\ 0 & \text{otherwise} \end{cases}$$

This way the Hamiltonian of the system is:

$$-\mathcal{H} = J \sum_{\langle ij \rangle} S_i S_j + H \sum_i S_i$$

where the notation $\langle ij \rangle$ means that i and j are nearest neighbours[10].

We now state the most important analytic properties of the bulk free energy density f (which hold in general).

Theorem. *Let f be the free energy density of a system at temperature T subject to an external (generic) field H. Then:*

---

[9]In the study of this situation, including the one in which there is no external field, we will use the *method of sources*: this consists in keeping always the term containing H_i in the Hamiltonian, and *at the end of the calculations* (i.e. after the thermodynamic limit) set H_i to any desired constant value, including zero.

[10]With this notation we are also implicitly assuming that we are not counting twice terms that are equal; to be explicit, considering the unidmensional case if we sum over all the possible nearest neighbours we would have terms like $S_1 S_2 + S_2 S_1 + S_2 S_3 + S_3 S_2 + \cdots$: every term (except eventually for the boundaries, but only if we don't take periodic conditions) is counted twice ($S_i S_j = S_j S_i$), so we should multiply everything by $1/2$ to correct this exceeding number of terms. With our notation we are implicitly doing that, and so the sum can be explicitly written as $S_1 S_2 + S_2 S_3 + S_3 S_4 + \cdots$; however, sometimes in the future it will be convenient to use the other convention, so we will count every possible term and then divide by 2.

a) f is negative, $f < 0$

b) f is a continuous function of its arguments

c) The derivatives of f exist almost everywhere; furthermore, right and left derivatives of f exist everywhere and are equal almost everywhere

d) $-\partial f/\partial T \geq 0$, namely the entropy per site is non-negative

e) $\partial f/\partial T$ is monotonically non-increasing with T, namely $\partial^2 f/\partial T^2 \leq 0$; this implies, from its definition, that $C_H \geq 0$

f) $\partial f/\partial H$ is monotonically non-increasing with H, namely $\partial^2 f/\partial H^2 \leq 0$; this implies, from its definition, that $\chi_T \geq 0$

We shall only prove d); the general strategy is to start from a finite system and then take the thermodynamic limit.

***Proof.* d):** If we assume the existence of the derivatives, then:

$$-\frac{\partial F}{\partial T} = \frac{\partial}{\partial T}\left[k_B T \ln\left(\operatorname{Tr} e^{-\frac{\mathcal{H}}{k_B T}}\right)\right] =$$

$$= k_B \ln\left(\operatorname{Tr} e^{-\frac{\mathcal{H}}{k_B T}}\right) + k_B T \frac{1}{\operatorname{Tr} e^{-\frac{\mathcal{H}}{k_B T}}} \operatorname{Tr}\left(e^{-\frac{\mathcal{H}}{k_B T}}\frac{\mathcal{H}}{k_B T^2}\right) =$$

$$= k_B \ln\left(\operatorname{Tr} e^{-\beta\mathcal{H}}\right) + k_B \frac{\operatorname{Tr}\left(\beta\mathcal{H}e^{-\beta\mathcal{H}}\right)}{Z} = k_B \left[\ln Z + \frac{\operatorname{Tr}\left(\beta\mathcal{H}e^{-\beta\mathcal{H}}\right)}{Z}\right]$$

If we now define:

$$\rho = \frac{e^{-\beta\mathcal{H}}}{Z} = \frac{e^{-\beta\mathcal{H}}}{\operatorname{Tr} e^{-\beta\mathcal{H}}}$$

we have:

$$-\frac{\partial F}{\partial T} = -k_B \operatorname{Tr}(\rho \ln \rho)$$

Now, since by definition $0 < \rho < 1$, $\ln \rho < 0$ and thus $-\partial F/\partial T$ is a sum of positive terms; therefore $-\partial F/\partial T$ is positive. Dividing by N and taking the thermodynamic limit we get $-\partial f/\partial T \geq 0$. $\qquad \square$

We can even prove something more (we will do it in the particular case of the Ising model):

Theorem. *The free energy density f is a concave function of H.*

Proof. The proof is based upon the *Hölder inequality*, which we now recall.
If $\{g_k\}$ and $\{h_k\}$ are two sequences with $g_k, h_k \geq 0$ for all k and $\alpha_1, \alpha_2 \in \mathbb{R}$ are such that $\alpha_1 + \alpha_2 = 1$, then:

$$\sum_k g_k^{\alpha_1} h_k^{\alpha_2} \leq \left(\sum_k g_k\right)^{\alpha_1}\left(\sum_k h_k\right)^{\alpha_2}$$

Let us now consider $Z(H)$ (for simplicity we will write only H as its argument). By definition:

$$Z(H) = \operatorname{Tr} e^{\beta H \sum_i S_i + \beta J \sum_{\langle ij \rangle} S_i S_j} = \operatorname{Tr}\left(e^{\beta H \sum_i S_i} G[S]\right)$$

where we have defined $G[S] := e^{\beta J \sum_{\langle ij \rangle} S_i S_j}$. Then we have:

$$Z(\alpha_1 H_1 + \alpha_2 H_2) = \mathrm{Tr}\left(e^{\beta \alpha_1 H_1 \sum_i S_i + \beta \alpha_2 H_2 \sum_i S_i} G[S]\right) =$$

$$= \mathrm{Tr}\left[\left(e^{\beta H_1 \sum_i S_i} G[S]\right)^{\alpha_1}\left(e^{\beta H_2 \sum_i S_i} G[S]\right)^{\alpha_2}\right] \leq \left(\mathrm{Tr}\, e^{\beta H_1 \sum_i S_i}\right)^{\alpha_1}\left(\mathrm{Tr}\, e^{\beta H_1 2 \sum_i S_i}\right)^{\alpha_2}$$

where we have also used the fact that $\alpha_1 + \alpha_2 = 1$, and thus:

$$Z(\alpha_1 H_1 + \alpha_2 H_2) \leq Z(H_1)^{\alpha_1} Z(H_2)^{\alpha_2}$$

Therefore, taking the logarithm at both sides, multiplying for $-k_B T$ and taking the thermodynamic limit we get:

$$f(\alpha_1 H_1 + \alpha_2 H_2) \geq \alpha_1 f(H_1) + \alpha_2 f(H_2)$$

and so f is indeed a concave function of H. $\qquad\square$

5.3.3 Absence of phase transitions for finite systems

We now analyse some symmetry properties of the Ising model, which will allow us to show that for finite systems no phase transition can occur at all.

To begin with, let us note that for any function φ of the spin configurations $\{S_i\}$ we have:

$$\sum_{\{S_i = \pm 1\}} \varphi(\{S_i\}) = \sum_{\{S_i = \pm 1\}} \varphi(\{-S_i\}) \qquad \Rightarrow \qquad \mathrm{Tr}\, \varphi(\{S_i\}) = \mathrm{Tr}\, \varphi(\{-S_i\}) \qquad (5.3)$$

which can be "proved" by explicitly writing all the terms[11].
Then, we can also see that f is an even function of H. In fact, from the definition of the Hamiltonian:

$$-\mathcal{H} = J \sum_{\langle i,j \rangle} S_i S_j + H \sum_i S_i$$

it is immediate to see that:

$$\mathcal{H}(H, J, \{S_i\}) = \mathcal{H}(-H, J, \{-S_i\}) \qquad (5.4)$$

Therefore, using (5.3) and then (5.4):

$$Z(-H, J, T) = \mathrm{Tr}\, e^{-\beta \mathcal{H}(-H, J, \{S_i\})} = \mathrm{Tr}\, e^{-\beta \mathcal{H}(-H, J, \{-S_i\})} = \mathrm{Tr}\, e^{-\beta \mathcal{H}(H, J, \{S_i\})} = Z(H, J, T)$$

If we now take the logarithm on both sides and multiply by $-k_B T$, we have:

$$F(H, J, T) = F(-H, J, T)$$

namely, the free energy is an even function of H.
This implies that the system (note that we have never taken the thermodynamic limit, so its size is still finite) can never exhibit a spontaneous magnetization when $H = 0$, and so there are no phase transitions at all (the system will always remain in its paramagnetic phase). In fact, we have:

$$M(H) = -\frac{\partial}{\partial H} F(H) = -\frac{\partial}{\partial H} F(-H) = \frac{\partial}{\partial(-H)} F(-H) = -M(-H)$$

---

[11]Very simply, since $S_i = \pm 1$, when we sum over all possible values of S_i we will in both cases cover all the possibilities for the argument $\{S_i\}$ of φ.

and if $H = 0$:

$$M(0) = -M(0) \qquad \Rightarrow \qquad M(0) = 0$$

i.e. the spontaneous magnetization is *always* null.

Note that this result has been obtained only with the use of symmetry properties, and we have never resorted to the "traditional" approach of statistical mechanics, namely the explicit computation of the partition function and subsequently the derivation of the thermodynamics of the system. This will be done later on.

5.3.4 Ising model and fluids

As we have stated in 5.3.1, the Ising model can also be used to describe systems different from magnets. The most important example is the correspondence that can be established between an Ising model and a fluid through a *lattice gas* model. We will now show how this equivalence can be defined; in order to do so, we will proceed in two steps: we will first show how an Ising model is equivalent to a lattice gas, and then show (qualitatively) that this lattice gas model is equivalent to the classical model for a fluid.
Before doing so, we briefly review the formalism used in classical statistical mechanics in order to describe fluids (see 3.3.3, for example).

Let us then consider a system of N particles subjected to some generic potentials, so that its Hamiltonian can be written as:

$$\mathcal{H} = \sum_{i=1}^{N} \left(\frac{\vec{p_i}^2}{2m} + U_1(\vec{r_i}) \right) + \frac{1}{2} \sum_{i \neq j} U_2(\vec{r_i}, \vec{r_j}) + \frac{1}{3!} \sum_{i \neq j \neq k} U_3(\vec{r_i}, \vec{r_j}, \vec{r_k}) + \cdots$$

where U_n is an n-body potential, which generally depends only on the distances between the particles if it involves two or more of them.
Working in the grand canonical ensemble, the grand partition function will be:

$$\mathcal{Z} = \text{Tr} \, e^{-\beta(\mathcal{H} - \mu N)}$$

where[12]:

$$\text{Tr} = \sum_{N=0}^{\infty} \frac{1}{N!} \int \prod_{i=1}^{N} \frac{1}{h^{dN}} d^d \vec{r_i} d^d \vec{p_i}$$

(h is the Planck constant, and d the dimensionality of the system[13]). It is convenient to separate the contributes of to the kinetic and configurational terms:

$$\mathcal{Z} = \sum_{N=0}^{\infty} \frac{1}{N!} \left(\int \prod_{i=1}^{N} \frac{d^d \vec{p_i}}{h^{dN}} e^{-\beta \frac{\vec{p_i}^2}{2m}} \right) \left(\int \prod_{i=1}^{N} d^d \vec{r_i} e^{-\beta(\{U\} - \mu N)} \right)$$

where with $\{U\}$ we mean all the possible potentials acting on the system. However (see, again, 3.3.3) we have:

$$\int \frac{d^d \vec{p}}{h^d} e^{-\beta \frac{\vec{p}^2}{2m}} = \Lambda(T)^{-d} \qquad \qquad \Lambda(T) = \frac{h}{\sqrt{2\pi m k_B T}}$$

where $\Lambda(T)$ is the *thermal wavelength*. Therefore:

$$\mathcal{Z} = \sum_{N=0}^{\infty} \frac{1}{N!} \left(\frac{e^{\beta \mu}}{\Lambda(T)^d} \right)^N Q_N$$

---

[12]This of course is valid for classical mechanics, but the correspondence we will establish holds also in quantum statistical mechanics.
[13]The only big difference with 3.3.3 is that we are considering d generic instead of equal to 3.

where:

$$Q_N = \int \prod_{i=1}^{N} d^d \vec{r}_i e^{-\beta\{U\}}$$

The grand free energy is thus[14]:

$$\mathcal{F}(T, \mu, \{U\}) = -k_B T \ln \mathcal{Z}$$

and as usual the thermodynamic limit of F is defined as:

$$f(T, \mu, \{U\}) = \lim_{V \to \infty} \frac{\mathcal{F}}{V}$$

with the constraint:

$$\rho = \lim_{V \to \infty} \frac{\langle N \rangle}{V} = \text{const.}$$

Ising model and lattice gas

The basic idea of the lattice gas model is to describe a fluid where the particles are located on the sites of a lattice instead of continuously occupying any position in space; it is a sort of "discretization" of the classical description of fluids. The correspondence with the Ising model is established relating the local density of a fluid with the local magnetization density of an Ising model.

Let us therefore consider a d-dimensional lattice with coordination number z. Each site of the lattice can be occupied by a particle, so if we call n_i the occupation number of a site we will have either $n_i = 0$ or $n_i = 1$, and the total number of particles will be:

$$N = \sum_{i=1}^{N_s} n_i$$

where N_s is the number of sites. In analogy with the continuum case we can guess a Hamiltonian of the form:

$$\mathcal{H} = \sum_{i=1}^{N_s} U_1(i) n_i + \frac{1}{2} \sum_{i \neq j} U_2(i, j) n_i n_j + \cdots$$

where the interaction potential U_2 is symmetric, namely $U_2(i, j) = U_2(j, i)$.
We have written only the configurational term, because we know that the kinetic part always contributes to the grand partition function with the thermal wavelength.
This way, in the grand canonical ensemble we have:

$$\mathcal{H} - \mu N = \sum_{i=1}^{N_s} (U_1(i) - \mu) n_i + \frac{1}{2} \sum_{i \neq j} U_2(i, j) n_i n_j + \cdots \tag{5.5}$$

Considering now an Ising model defined on the same lattice, we can establish a correspondence with the lattice gas model defining:

$$n_i = \frac{1}{2}(1 + S_i) \tag{5.6}$$

---

[14]A small remark: for finite and "reasonable" systems the grand free energy is not singular even if it involves an infinite sum over N. The reason is that generally (in the "reasonable" cases we have just mentioned) the interaction potentials have a hard-core component that prevents the particles from overlapping: therefore, a finite system will be able to contain only a finite number of particles, so that the sum has in reality an upper limit and is not infinite.

where S_i is the spin variable of the Ising model defined on the same lattice: in fact, doing so we have $n_i = 0$ when $S_i = -1$ and $n_i = 1$ when $S_i = 1$.

What we now want to show is that substituting (5.6) into (5.5) we obtain a Hamiltonian that leads to the grand partition function of the Ising model, so that the equivalence between the two models is made explicit[15].

From now on we will neglect any potential that couples more than two particles; therefore, the first term of the right-hand side of (5.5) becomes:

$$\sum_{i=1}^{N_s} (U_1(i) - \mu) \frac{1}{2}(1 + S_i) = \frac{1}{2} \sum_{i=1}^{N_s} (U_1(i) - \mu) + \frac{1}{2} \sum_{i=1}^{N_s} (U_1(i) - \mu) S_i$$

while the second term:

$$\frac{1}{8} \sum_{i \neq j} U_2(i,j)(1 + S_i)(1 + S_j) = \frac{1}{8} \sum_{i \neq j} U_2(i,j) + \frac{1}{4} \sum_{i \neq j} U_2(i,j) S_i + \frac{1}{8} \sum_{i \neq j} U_2(i,j) S_i S_j$$

(where in the second summation has been obtained relabelling indices and using the symmetry of U_2). If we now suppose that U_2 is a nearest-neighbour interaction, i.e.:

$$U_2(i,j) = \begin{cases} U_2 & i \text{ and } j \text{ are nearest neighbours} \\ 0 & \text{otherwise} \end{cases}$$

then:

$$\frac{1}{2} \sum_{i \neq j} U_2(i,j) n_i n_j = \frac{1}{8} U_2 z N_s + \frac{1}{4} U_2 z \sum_{i=1}^{N_s} S_i + \frac{1}{8} U_2 \sum_{\langle ij \rangle} S_i S_j$$

If we now set $U_1 = 0$ (i.e. there are no external fields acting on the fluid) we have:

$$\frac{1}{2} \sum_{i=1}^{N_s} (U_1(i) - \mu) n_i = -\frac{1}{2} \mu N_s - \frac{1}{2} \mu \sum_{i=1}^{N_s} S_i$$

Therefore:

$$\mathcal{H} - \mu N = -\frac{\mu}{2} N_s + \frac{z}{8} U_2 N_s + \left(-\frac{\mu}{2} + \frac{z}{4} U_2\right) \sum_{i=1}^{N_s} S_i + \frac{U_2}{8} \sum_{\langle ij \rangle} S_i S_j$$

For the sake of simplicity, we set[16]:

$$E_0 = -\frac{\mu}{2} N_s + \frac{z}{8} U_2 N_s \qquad -H = -\frac{\mu}{2} + \frac{z}{4} U_2 \qquad -J = \frac{U_2}{8} \qquad (5.7)$$

so that:

$$\mathcal{H} - \mu N = E_0 - H \sum_{i=1}^{N_s} S_i - J \sum_{\langle ij \rangle} S_i S_j$$

We thus have that the grand partition function of the lattice gas can be written as:

$$\mathcal{Z}_{\text{l.g.}} = \operatorname{Tr} e^{-\beta(\mathcal{H} - \mu N)} = \operatorname{Tr} e^{-\beta\left(-E_0 + H \sum_i S_i + J \sum_{\langle ij \rangle} S_i S_j\right)} =$$

$$= e^{-\beta E_0} \operatorname{Tr} e^{-\beta\left(H \sum_i S_i + J \sum_{\langle ij \rangle} S_i S_j\right)} \qquad \Rightarrow$$

---

[15]Of course, if two systems have the same partition function, their thermodynamics will coincide, so they are at all effects equivalent.

[16]From these definition we see that in reality the precise values of the coefficients $1/4$, $1/2$, $1/8$ that we have encountered are absolutely irrelevant.

$$\Rightarrow \qquad \mathcal{Z}_{\text{l.g.}} = e^{-\beta E_0} \mathcal{Z}_{\text{I.m.}}$$

where $\mathcal{Z}_{\text{I.m.}}$ is the grand partition function of the Ising model.

Therefore, we see that the grand partition functions of both the Ising model and the lattice gas are equal, except for the irrelevant constant rescaling factor $e^{-\beta E_0}$.

This way we can see explicitly that the two models are perfectly equivalent.

Lattice gas and continuous fluids

We now want to show (although not really rigorously) that the lattice gas model can be derived from the "classical" model of a fluid.

Consider the configurational sum:

$$Q_N = \int \prod_{i=1}^{N} d^d \vec{r}_i \, e^{-\beta\{U\}}$$

We can approximate it dividing our system in N_c cells of linear dimension a, such that the probability to find more than one particle inside one cell is negligible; in other words, if our particles interact via a potential that has a hard core component we can take a of the order of the hard core radius. This way the integral in Q_N can be rewritten as:

$$\int \prod_{i=1}^{N} d^d \vec{r}_i \approx a^{dN} \sum_{\alpha=1}^{N_c}$$

where i labels the particles while α labels the cells.

Since we have focused our attention towards the cells, we can substitute the interaction $U_2(\vec{r}_i, \vec{r}_j)$ between particles with an interaction between occupied cells:

$$U_2(\alpha, \beta) = U_2(\vec{r}_i, \vec{r}_j) \qquad \text{if } \vec{r}_i \in \alpha \text{ and } \vec{r}_j \in \beta$$

As we have done before we consider only two-particle interactions and suppose that the relative potential is short-ranged, namely U_2 acts only between nearest neighbours and in that case is equal to a constant value U_2; therefore:

$$\sum_{\alpha \neq \beta} U_2(\alpha, \beta) n_\alpha n_\beta = U_2 \sum_{\langle \alpha\beta \rangle} n_\alpha n_\beta$$

where of course n_α is the occupation number of the α-th cell.

Now, for each specified configuration of the occupation numbers n_α there can be $N!$ possible configurations for the positions $\vec{r}_i$: in fact, the configuration $\{n_\alpha\}$ of the occupied cells only tells us *where* we can find a particle, but not *which* specific particle is in a given cell; furthermore, the system is left unchanged if we interchange the particles. We can thus write:

$$Q_N = N! a^{dN} \sum_{\{n_\alpha=0,1\}}' e^{-\beta U(\{n_\alpha\})}$$

where $\sum'$ means that the sum on n_α must be subject to the constraint that the total number of particle be fixed and equal to N:

$$\sum_{\{n_\alpha=0,1\}}' = \sum_{\{n_\alpha=0,1\}} \qquad \text{with } \sum_\alpha n_\alpha = N$$

Therefore:

$$\mathcal{Z}_{\text{c.f.}} = \sum_{N=0}^{\infty} \frac{1}{N!} \left(\frac{e^{\beta\mu}}{\Lambda(T)^d} \right)^N Q_N = \sum_{N=0}^{\infty} \left[e^{\beta\mu} \left(\frac{a}{\Lambda(T)} \right)^d \right]^N \sum_{\{n_\alpha=0,1\}}' e^{-\beta U(\{n_\alpha\})}$$

where "c.f." stands for "continuum fluid". The presence of the sum $\sum_{N=0}^{\infty}$ strongly simplifies the calculations: in fact, if $f(n_\alpha)$ is a generic function, then:

$$\sum_{N=0}^{\infty} \sum_{\{n_\alpha\}}' f(n_\alpha) = \underbrace{\sum_{\{n_\alpha\}} f(n_\alpha)}_{\sum_\alpha n_\alpha = 0} + \underbrace{\sum_{\{n_\alpha\}} f(n_\alpha)}_{\sum_\alpha n_\alpha = 1} + \cdots + \underbrace{\sum_{\{n_\alpha\}} f(n_\alpha)}_{\sum_\alpha n_\alpha = \infty} = \sum_{n_\alpha} f(n_\alpha)$$

where the last sum in unconstrained. Therefore we can write $\mathcal{Z}_{\text{c.f.}}$ as:

$$\mathcal{Z}_{\text{c.f.}} = \sum_{\{n_\alpha=0,1\}} \left[e^{\beta\mu} e^{d\log\frac{a}{\Lambda(T)}} \right]^N e^{-\beta U_2 \sum_{\langle\alpha\beta\rangle} n_\alpha n_\beta} = \sum_{\{n_\alpha=0,1\}} e^{-\beta\left[U_2 \sum_{\langle\alpha\beta\rangle} n_\alpha n_\beta - \left(\mu + \frac{d}{\beta}\log\frac{a}{\Lambda(T)}\right)N \right]}$$

namely:

$$\mathcal{Z}_{\text{c.f.}} = \text{Tr}\, e^{-\beta(\mathcal{H}-\overline{\mu}N)} = \mathcal{Z}_{\text{l.g.}}$$

where $\mathcal{Z}_{\text{l.g.}}$ is the grand partition function of the lattice gas model, while:

$$\overline{\mu} = \mu_{\text{cl.}} + \frac{d}{\beta} \log \frac{a}{\Lambda(T)}$$

and $\mu_{\text{cl.}}$ is the "classical" chemical potential. However, since the chemical potential has the meaning of an energy, any shift in μ by a constant quantity[17] is physically irrelevant, so $\overline{\mu}$ and $\mu_{\text{cl.}}$ are equivalent[18].
Therefore, the grand partition functions of the continuum fluid and the lattice gas model are the same: we thus see that these two models are equivalent.

5.3.5 Ising model and the ideal gas

When we have discussed the relation of the Ising model with lattice gases, we have included the interactions between spins (or equivalently particles), so in the end the Ising model could be used to describe non-ideal fluids.
What we now want to show is that the Ising model in the non interacting case (which could seem not really interesting at first) is equivalent to the ideal gas.

We start of course from what we have seen in the case of a generic lattice gas:

$$\mathcal{Z}_{\text{l.g.}} = e^{-\beta E_0} \mathcal{Z}_{\text{I.m.}}$$

where (in the case of a square lattice, for which we know that $z = 2d$):

$$-E_0 = \left(\frac{2}{\mu} - \frac{U_2 d}{4} \right) N$$

and we have renamed N_{s} with N. From (5.7) we can rewrite E_0 as:

$$E_0 = -N \left(H + \frac{U_2}{4}d \right)$$

---

[17]Remember that our system is at fixed temperature, so T is constant.
[18]Equivalently, we can also note that $e^{\beta\overline{\mu}N} = e^{\beta\mu N} e^{dN\log(a/\Lambda)}$, so $\mathcal{Z}_{\text{c.f.}} = e^{dN\log(a/\Lambda)} \mathcal{Z}_{\text{l.g.}}$: the grand partition functions of the fluid and the lattice gas differ for a constant rescaling factor, similarly to what we have seen before.

Since we are working in the grand canonical ensemble, we have[19] $\mathcal{Z}_{\text{l.g.}} = e^{-\beta\Phi} = e^{\beta PN}$ and so:

$$P = -\frac{E_0}{N} + \frac{1}{\beta N}\ln \mathcal{Z}_{\text{I.m.}} = H + \frac{U_2}{4}d + \frac{1}{\beta N}\ln \mathcal{Z}_{\text{I.m.}}$$

which, in the non interacting case ($U_2 = 0$), reduces to:

$$P = H + \frac{1}{\beta N}\ln \mathcal{Z}_{\text{I.m}}$$

The partition function of a non interacting Ising model is:

$$\mathcal{Z}_{\text{I.m.}} = \sum_{\{S_i=\pm 1\}} e^{h\sum_i S_i}$$

where as usual $h = \beta H$. This can be computed easily:

$$\mathcal{Z}_{\text{I.m.}} = \sum_{S_1=\pm 1}\cdots\sum_{S_N=\pm 1} e^{hS_1}\cdots e^{hS_N} = \prod_{i=1}^{N}\left(\sum_{S=\pm 1} e^{hS}\right) = (2\cosh h)^N$$

and thus:

$$\frac{\ln \mathcal{Z}_{\text{I.m.}}}{N} = \ln(2\cosh h)$$

If we now call $\langle N\rangle$ the mean number of spins of the system, what we want to do is to relate P with the density $\langle N\rangle/N$.

Similarly to what we have done for the lattice gas, we define $n_i = (1+S_i)/2$ to be the occupation number of the i-th cell of the system. This way, the mean number of particles is:

$$\langle N\rangle = \left\langle\sum_i \frac{1+S_i}{2}\right\rangle = \frac{N}{2} + \frac{1}{2}\left\langle\sum_i S_i\right\rangle$$

but:

$$\left\langle\sum_i S_i\right\rangle = \frac{\sum_{\{S\}} e^{h\sum_i S_i}\sum_j S_j}{\sum_{\{S\}} e^{h\sum_i S_i}} = \frac{1}{\mathcal{Z}_{\text{I.m}}}\frac{\partial \mathcal{Z}_{\text{I.m}}}{\partial h} = \frac{\partial \ln \mathcal{Z}_{\text{I.m}}}{\partial h}$$

Therefore:

$$v^{-1} := \frac{\langle N\rangle}{N} = \frac{1}{2}\left(1 + \frac{1}{N}\frac{\partial \ln \mathcal{Z}_{\text{I.m}}}{\partial h}\right) = \frac{1}{2}(1 + \tanh h)$$

However, we can also define the mean "magnetization" per spin of the system as:

$$m := \left\langle\frac{1}{N}\sum_i S_i\right\rangle = \frac{1}{N}\frac{\partial \ln \mathcal{Z}_{\text{I.m}}}{\partial h}$$

We therefore have $v^{-1} = (1+m)/2$, and we can also rewrite:

$$\frac{1}{N}\ln \mathcal{Z}_{\text{I.m}} = \ln 2 + \frac{1}{2}\ln\cosh^2 h = \ln 2 + \frac{1}{2}\ln\frac{\cosh^2 h}{\cosh^2 h - \sinh^2 h} = \ln 2 - \frac{1}{2}\ln(1-m^2)$$

where we have used two "tricks": first, we have written $\cosh h = \sqrt{\cosh^2 h}$ and then divided by $\cosh^2 h - \sinh^2 h = 1$. This way:

$$P = \frac{h}{\beta} + \frac{1}{\beta}\left[\ln 2 - \frac{1}{2}\ln(1-m^2)\right]$$

---

[19]The "volume" of our system is now the number of spins N.

and since $h = \tanh^{-1} m = \frac{1}{2} \ln \frac{1+m}{1-m}$, this can be rewritten as:

$$P = -\frac{1}{\beta} \ln \frac{1-m}{2} = -\frac{1}{\beta} \ln(1 - v^{-1})$$

If our gas is dilute, i.e. $v^{-1} \approx 0$, we can expand the logarithm and get:

$$P \sim \frac{v^{-1}}{\beta} = \frac{\langle N \rangle}{N} k_B T$$

which is the ideal gas law (since N is the volume of our system)!

5.3.6 Ising model and binary alloys

The Ising model can be also used to described systems composed of different kinds of particles, like binary alloys.
What we now want to see is that, similarly to what happened in 5.3.4 for the lattice gas, the Hamiltonian of a binary alloy can be mapped into the Hamiltonian of an Ising model.

Consider a lattice with coordination number z, and suppose that on every site of this lattice there can be an atom of two possible different elements, say A and B, and that they can "move" (i.e. exchange their positions) on the lattice[20].
Let us call $-E_{AA}$ and $-E_{BB}$ (writing explicitly the negative sign) the interaction energies between neighbouring atoms of the same elements, respectively A and B; similarly, we call $-E_{AB}$ the interaction energy of neighbouring atoms of different kinds.
Let us also call:

- N_{AA}, N_{BB} the number of A-A and B-B bonds

- N_{AB} the number of A-B bonds

- N_A, N_B the numbers of A and B atoms

- $N = N_A + N_B$ the total number of atoms

This way, the energy of the system will be:

$$E = -E_{AA} N_{AA} - E_{AB} N_{AB} - E_{BB} N_{BB}$$

However, N_{AA}, N_{BB} and N_{AB} are not independent.
Let us in fact consider all the A atoms of our system: we have that every A-A bond contributes to N_A with two atoms and every A-B bond with a single A atom. If we add these numbers we have z times the total number of A atoms, i.e.:

$$2N_{AA} + N_{AB} = zN_A$$

and similarly:

$$2N_{BB} + N_{AB} = zN_B$$

The fact that these sums are z times equal to N_A or N_B is better understood if explicitly verified in simple cases. Check for example that this is true in the following two-dimensional case

---

[20]This happens, for example, in β-brasses: at temperatures lower than approximately 733°C the atoms are arranged in a body-centered cubic lattice, with zinc atoms occupying the center of the copper cubes; if the temprature is raised then zincs and coppers freely exchange.

(assuming periodic boundary conditions):

$$
\begin{matrix}
A & B & A & A & B \\
B & A & A & B & B \\
A & B & A & A & B \\
B & A & B & B & B \\
B & A & A & B & A
\end{matrix}
$$

If we now solve these last two equations, expressing everything in terms of N_A and N_{AA} we get:

$$N_B = N - N_A \qquad N_{BB} = N_{AA} + (N - 2N_A)\frac{z}{2} \qquad N_{AB} = zN_A - 2N_{AA}$$

and thus the energy of the system can be rewritten as:

$$-E = N_{AA}(E_{AA} + E_{BB} - 2E_{AB}) + zN_A(E_{AB} - E_{BB}) + \frac{zN}{2}E_{BB}$$

Now, in order to establish a correspondence with the Ising model, similarly to what we have done for the lattice gas, we can define a site variable n_i which represents if an A or B atom is occupying the i-th site. We define n_i such that $n_i = 1$ when an A atom is present while $n_i = 0$ when a B atom is present in the site; this way, we can map this system into an Ising model setting (just like before[21]):

$$n_i = \frac{1 + S_i}{2}$$

Therefore:

$$N_A = \sum_i n_i = \sum_i \frac{1 + S_i}{2} \qquad N_{AA} = \sum_{\langle ij \rangle} \frac{1 + S_i}{2}\frac{1 + S_j}{2}$$

Computing the sums just like we have done for the lattice gas, we have:

$$N_A = \frac{N}{2} + \frac{1}{2}\sum_i S_i \qquad N_{AA} = \frac{1}{4}\sum_{\langle ij \rangle} S_i S_j + \frac{z}{2}\sum_i S_i + \frac{zN}{8}$$

Thus, the energy of the system can be rewritten as:

$$-E = J\sum_{\langle ij \rangle} S_i S_j + h\sum_i S_i + CN \tag{5.8}$$

where:

$$J = \frac{1}{4}(E_{AA} + E_{BB} - 2E_{AB}) \qquad h = \frac{z}{4}(E_{AA} - E_{BB}) \qquad C = \frac{z}{8}(E_{AA} + E_{BB} + 2E_{AB})$$

The energy (5.8) has again the same for of the Hamiltonian of the Ising model, apart from the irrelevant constant shift CN.

Therefore, the partition functions of the Ising model and the binary alloy are perfectly equivalent.

5.3.7 Ising model and neural networks

Another kind of systems that can be described in terms of the Ising model are neural networks: in this case every site of the lattice represents a neuron and the interaction bonds are synapses; the two possible states of the neurons are "firing" (i.e. transmitting an electric pulse) when $S_i = +1$ and "not firing" when $S_i = -1$. In this case, however, the system is dynamic and not static so the situation is much more complicated, and the tools we are developing are not really useful in order to understand something interesting about the system.

---

[21]From this definition we can guess that from now on we can do exactly what we have seen for the lattice gas.

5.3.8 Final remark on equivalence in statistical mechanics

The analysis that we have just carried out has shown an *equivalence* between physical models. In statistical mechanics, the term *equivalence* can be used with two different "nuances":

exact equivalence: this means that there is an exact mapping between the partition functions of two different models

approximate equivalence: this means that the partition functions of two models are *not* related by an exact mapping, but nevertheless they behave in exactly the same way near a critical point. Such models are said to belong to the same *universality class*.

Apart from the equivalence between the Ising model and neural networks, all the correspondences we have shown are exact.

5.4 Ising model in one dimension

We will now proceed to study the thermodynamic properties of a one-dimensional Ising model. In this case, as we will shortly see, we will be able to compute exactly the partition function of the system using different methods; therefore, we will be able to determine exactly all the thermodynamic properties of the system.

5.4.1 Bulk free energy, thermodynamic limit and absence of phase transitions

Let us therefore consider a one-dimensional Ising model with N spins and free boundary conditions, i.e. the first and the last spin can assume any value. Using the nearest neighbour interaction Hamiltonian we have:

$$-\mathcal{H} = H\sum_i S_i + J\sum_{\langle ij \rangle} S_i S_j$$

and defining $h = \beta H$ and $K = \beta J$ for the sake of simplicity, the partition function of the system will be:

$$Z_N = \text{Tr}\, e^{-\beta\mathcal{H}} = \sum_{\{S_i = \pm 1\}} e^{h\sum_i S_i + K\sum_i S_i S_{i+1}}$$

If we now set $h = 0$, namely if the system is not subjected to any external field, then:

$$Z_N = \sum_{\{S_i = \pm 1\}} e^{K\sum_i S_i S_{i+1}} = \sum_{S_1 = \pm 1}\cdots\sum_{S_N = \pm 1} e^{KS_1 S_2 + \cdots + KS_{N-1}S_N}$$

In order to compute Z_N we use the so called *recursion method*: from the expression of Z_N we can deduce the expression of the partition function Z_{N+1} of the same system with one additional spin added to the lattice:

$$Z_{N+1} = \sum_{S_1 = \pm 1}\cdots\sum_{S_N = \pm 1}\sum_{S_{N+1} = \pm 1} e^{K(S_1 S_2 + \cdots + S_{N-1}S_N)} e^{KS_N S_{N+1}}$$

However, the sum over S_{N+1} gives:

$$\sum_{S_{N+1} = \pm 1} e^{KS_N S_{N+1}} = e^{KS_N} + e^{-KS_N} = 2\cosh(KS_N) = 2\cosh K$$

where we have used the evenness of the hyperbolic cosine, namely the fact that $\cosh(\pm x) = \cosh x$. Therefore we have:

$$Z_{N+1} = Z_N \cdot 2\cosh K$$

and iterating the relation N times we get:

$$Z_{N+1} = Z_1(2\cosh K)^N \qquad \Rightarrow \qquad Z_N = Z_1(2\cosh K)^{N-1}$$

Since[22]:

$$Z_1 = \sum_{S_1=\pm 1} 1 = 1 + 1 = 2$$

we get the final result:

$$Z_N = 2(2\cosh K)^{N-1}$$

Now, following all the prescriptions we know, we get to:

$$F(T) = -k_B T \left[\ln 2 + (N-1)\ln\left(2\cosh\frac{J}{k_B T}\right) \right] \tag{5.9}$$

and in the thermodynamic limit:

$$f(T) = -k_B T \ln\left(2\cosh\frac{J}{k_B T}\right) \tag{5.10}$$

Let us note that f does indeed respect the properties we have seen in 5.3.2. Furthermore, since the logarithm and the hyperbolic cosine are analytic functions we see that $f(T)$ is itself an analytic function of T: it is therefore impossible that the system will exhibit any kind of phase transition (at least for $T \neq 0$[23]) even in the thermodynamic limit.

5.4.2   Irrelevance of boundary conditions

We can wonder what happens if we change the boundary conditions: since they affect only a small part of the system we expect that as long as the system is finite we can observe some differences if we choose different boundary conditions, but as soon as we take the thermodynamic limit these differences become irrelevant.

As an example let us compute again Z_N but with periodic boundary conditions, i.e. the first and the last spin are coupled (this can be visually interpreted as "closing" our system in a circle, connecting the first and last spins). In this case we have:

$$Z_N = \sum_{S_1=\pm 1} \cdots \sum_{S_N=\pm 1} e^{K(S_1 S_2 + \cdots S_{N-1} S_N)} e^{K S_N S_1}$$

In order to compute this partition function we use another "trick"; we define the variables:

$$\eta_i = \begin{cases} 1 & \text{if } S_i = S_{i+1} \\ -1 & \text{if } S_i = -S_{i+1} \end{cases}$$

and set $\eta_0 = S_1$. This way, we can substitute $S_i S_{i+1}$ with η_i and since $S_i^2 = 1$ we can also write:

$$S_N S_1 = S_N S_{N-1} S_{N-1} S_{N-2} \cdots S_2 S_1 = \eta_1 \eta_2 \cdots \eta_{N-1}$$

Therefore, the partition function in terms of η_i becomes:

$$Z_N = \sum_{\eta_0=\pm 1} \cdots \sum_{\eta_{N-1}=\pm 1} e^{K(\eta_1 + \cdots + \eta_{N-1})} e^{K \eta_1 \cdots \eta_{N-1}}$$

---

[22]This can be justified as follows: Z_1 is the partition function of a single-spin Ising model and we are considering two-spin interactions, so this single spin will not interact with anything. Therefore the Hamiltonian of the system in this case is null, and so $e^{-\beta \mathcal{H}} = e^0 = 1$.

[23]In this case, in fact, there can be problems. From this fact we can state that the only "phase transition" that can happen in a one-dimensional Ising model occurs at $T = 0$ (which is of course an unphysical case), where all the spins are aligned.

Now, summing over η_0 and rewriting $e^{K\eta_1\cdots\eta_{N-1}}$ with the definition of exponential series:

$$Z_N = 2 \sum_{\eta=\pm1} \cdots \sum_{\eta_{N-1}=\pm1} e^{K(\eta_1+\cdots+\eta_{N-1})} \sum_{\alpha=0}^{\infty} \frac{1}{\alpha!}(K\eta_1\cdots\eta_{N-1})^{\alpha} =$$

$$= 2\sum_{\alpha=0}^{\infty} \frac{K^{\alpha}}{\alpha!} \left(\sum_{\eta=\pm1} \eta^{\alpha} e^{K\eta}\right)^{N-1} = 2\sum_{\alpha=0}^{\infty} \frac{K^{\alpha}}{\alpha!} \left[e^K + (-1)^{\alpha} e^{-K}\right]^{N-1} =$$

$$= 2\left[(e^K + e^{-K})^{N-1} + K(e^K - e^{-K})^{N-1} + \frac{K^2}{2!}(e^K + e^{-K})^{N-1} + \frac{K^3}{3!}(e^K - e^{-K})^{N-1} + \cdots\right] =$$

$$= 2\left[2^{N-1}(\cosh K)^{N-1}\left(1 + \frac{K^2}{2!} + \frac{K^4}{4!} + \cdots\right) + 2^{N-1}(\sinh K)^{N-1}\left(K + \frac{K^3}{3!} + \frac{K^5}{5!} + \cdots\right)\right] =$$

$$= 2^N(\cosh K)^{N-1}\cosh K + 2^N(\sinh K)^{N-1}\sinh K$$

Therefore:

$$Z_N = (2\cosh K)^N + (2\sinh K)^N$$

and the finite-size free energy is:

$$F(T) = -k_B T N \left\{\ln\left(2\cosh\frac{J}{k_B T}\right) + \frac{1}{N}\ln\left[1 + \left(\tanh\frac{J}{k_B T}\right)^N\right]\right\}$$

which is different from (5.9), and as we have already stated this is due to the fact that the system we are considering is still finite-sized. If we now take the thermodynamic limit we will have:

$$f(T) = -k_B T \ln\left(2\cosh\frac{J}{k_B T}\right) - k_B T \lim_{N\to\infty} \frac{1}{N}\ln\left[1 + \left(\tanh\frac{J}{k_B T}\right)^N\right]$$

The last limit, however, is null: in fact $\tanh x < 1$ for any x and so as N grows $(\tanh x)^N$ goes to zero, so *a fortiori* $(\tanh x)^N/N$ vanishes as $N \to \infty$. Therefore:

$$f(T) = -k_B T \ln\left(2\cosh\frac{J}{k_B T}\right)$$

which is exactly (5.10).

Since, as expected, the boundary conditions do not affect the properties of macroscopic systems and their effect on finite systems are very small, when we are studying a model we can choose the one we prefer and which makes calculations easier.

5.4.3 Absence of spontaneous magnetization

Let us now come back to free boundary conditions and compute the mean magnetization $\langle S_j \rangle$ of a given site j. By definition we have[24]:

$$Z\langle S_j \rangle = \text{Tr}\left(S_j e^{-\beta\mathcal{H}}\right) = \sum_{\{S_i=\pm1\}} S_j e^{K\sum_{i=1}^{N-1} S_i S_{i+1}} = \sum_{\{S_i=\pm1\}} \prod_{i=1}^{N-1} S_j e^{K S_i S_{i+1}}$$

---

[24]Note that the sum on nearest neighbour is done without counting the same terms twice (as we have already stressed). In fact, in our case every spin S_i interacts with its nearest neighbours S_{i-1} and S_{i+1}, but the sum in the trace involves every two-spin interaction only once.

We now use the fact that $e^x = \cosh x + \sinh x$, so that in this case:

$$e^{KS_iS_{i+1}} = \cosh K + S_iS_{i+1}\sinh K$$

where we have used (like we have done previously) the evenness of $\cosh$, and now also the oddness of $\sinh$, i.e. $\sinh(\pm x) = \pm\sinh x$. This way:

$$Z\langle S_j\rangle = \sum_{\{S_i=\pm 1\}}\prod_{i=1}^{N-1} S_j\left(\cosh K + S_iS_{i+1}\sinh K\right) =$$

$$= (\cosh K)^{N-1}\sum_{\{S_i=\pm 1\}}\prod_{i=1}^{N-1} S_j\left(1 + S_iS_{i+1}\tanh K\right) =$$

$$= (\cosh K)^{N-1}\sum_{S_1=\pm 1}\cdots\sum_{S_N=\pm 1} S_j(1+S_1S_2\tanh K)(1+S_2S_3\tanh K)\cdots(1+S_{N-1}S_N\tanh K) =$$

$$= (\cosh K)^{N-1}\sum_{S_1=\pm 1}\cdots\sum_{S_N=\pm 1}\left[S_j + \sum_{M=1}^{N-1} S_jS_{i_1}S_{i_1+1}S_{i_2}S_{i_2+1}\cdots S_{i_M}S_{i_M+1}(\tanh K)^M\right]$$

where S_{i_1} etc. are the spin variables appropriately rearranged[25].

Let us now consider the two different contributions to $\langle S_j\rangle$. As of the first:

$$\sum_{S_1=\pm 1}\cdots\sum_{S_N=\pm 1} S_j = 2^{N-1}\sum_{S_j=\pm 1} S_j = 2^{N-1}(1-1) = 0$$

Considering now the second one:

$$\sum_{S_1=\pm 1}\cdots\sum_{S_N=\pm 1}\sum_{M=1}^{N-1} S_jS_{i_1}S_{i_1+1}S_{i_2}S_{i_2+1}\cdots S_{i_M}S_{i_M+1}(\tanh K)^M$$

we have that for every fixed M this term vanishes; for example, if we consider the contribution relative to a fixed value M^* of M and sum (for example) over S_{i_1} we have:

$$S_j\cdot(+1)\cdot S_{i_1+1}S_{i_2}S_{i_2+1}\cdots S_{i_{M^*}}S_{i_{M^*}+1}(\tanh K)^{M^*} +$$

$$+ S_j\cdot(-1)\cdot S_{i_1+1}S_{i_2}S_{i_2+1}\cdots S_{i_{M^*}}S_{i_{M^*}+1}(\tanh K)^{M^*} = 0$$

Therefore, also the second term vanishes and in the end:

$$\langle S_j\rangle = 0$$

This result perfectly agrees with what we have already seen in 5.3.3, but now it has been deduced from a direct computation.

5.4.4 The transfer matrix method

The method we have just seen is an *ad hoc* solution that works only in this case, and it would be rather difficult to extend it in the presence of an external field. A more general method that

---

[25]The form of this last term can be understood more easily doing an explicit computation with a simple example.

allows us to extend our considerations also when $h \neq 0$ and to compute other interesting properties is the so called *transfer matrix method*, which basically consists in defining an appropriate matrix related to the model such that all the thermodynamic properties of the system can be extracted from the eigenvalues of this matrix.

We are going to see this method applied to the one-dimensional Ising model, but its validity is completely general; we will stress every time if we are stating general properties of the transfer matrix method or restricting to particular cases.

The Hamiltonian of a one-dimensional Ising model with periodic boundary conditions when an external field is present is such that:

$$-\beta\mathcal{H} = K(S_1 S_2 + \cdots + S_{N-1} S_N + S_N S_1) + h \sum_{i=1}^{N} S_i$$

where $\beta\mathcal{H}$ is sometimes called *reduced Hamiltonian*.
We now rewrite the partition function in the following "symmetric" way:

$$Z_N = \sum_{S_1=\pm 1} \cdots \sum_{S_N=\pm 1} \left[e^{KS_1 S_2 + \frac{h}{2}(S_1+S_2)} \right] \left[e^{KS_2 S_3 + \frac{h}{2}(S_2+S_3)} \right] \cdots \left[e^{KS_N S_1 + \frac{h}{2}(S_N+S_1)} \right]$$

If we therefore define the *transfer matrix* $\boldsymbol{T}$ such that[26]:

$$\langle S | \boldsymbol{T} | S' \rangle = e^{KSS' + \frac{h}{2}(S+S')}$$

we can write Z_N as a product of the matrix elements of $\boldsymbol{T}$:

$$Z_N = \sum_{S_1=\pm 1} \cdots \sum_{S_N=\pm 1} \langle S_1 | \boldsymbol{T} | S_2 \rangle \langle S_2 | \boldsymbol{T} | S_3 \rangle \cdots \langle S_N | \boldsymbol{T} | S_1 \rangle \tag{5.11}$$

If we now choose $|S_i\rangle$ so that they are orthonormal, i.e.:

$$|+1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad\qquad |-1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

then an explicit representation of $\boldsymbol{T}$ is:

$$\boldsymbol{T} = \begin{pmatrix} e^{K+h} & e^{-K} \\ e^{-K} & e^{K-h} \end{pmatrix} \tag{5.12}$$

Note that the matrix elements of $\boldsymbol{T}$ are in one-to-one correspondence with the spin variables, and that the dimension of the transfer matrix depends on the number of possible values that they can assume.
Now, since the vectors $|S_i\rangle$ are orthonormal we have:

$$\sum_{S_i=\pm 1} |S_i\rangle\langle S_i| = \mathbb{I}$$

where $\mathbb{I}$ is the identity matrix, and (5.11) becomes:

$$Z_N = \sum_{S_1=\pm 1} \langle S_1 | \boldsymbol{T}^N | S_1 \rangle = \operatorname{Tr} \boldsymbol{T}^N$$

This is the general purpose of the transfer matrix method: being able to write the partition function of a system as the trace of the N-th power of an appropriately defined matrix (the

---

[26]We symbolically use Dirac's bra-ket notation.

transfer matrix).

Now, the trace can be easily computed if we diagonalize $\boldsymbol{T}$; if we call $\boldsymbol{T}_D$ the diagonalization of the transfer matrix we will have:

$$\boldsymbol{T}_D = \boldsymbol{P}^{-1}\boldsymbol{T}\boldsymbol{P}$$

where $\boldsymbol{P}$ is an invertible matrix whose columns are the eigenvectors of $\boldsymbol{T}$. Since $\boldsymbol{P}^{-1}\boldsymbol{P} = \mathbb{I}$, the partition function becomes:

$$Z_N = \operatorname{Tr}\boldsymbol{T}^N = \operatorname{Tr}(\boldsymbol{P}\underbrace{\boldsymbol{P}^{-1}\boldsymbol{T}\boldsymbol{P}\boldsymbol{P}^{-1}\cdots\boldsymbol{P}^{-1}\boldsymbol{T}\boldsymbol{P}}_{N \text{ times}}\boldsymbol{P}^{-1}) = \operatorname{Tr}\left(\boldsymbol{P}\boldsymbol{T}_D^N\boldsymbol{P}^{-1}\right)$$

and using the cyclic property of the trace[27] we get:

$$Z_N = \operatorname{Tr}\left(\boldsymbol{T}_D^N\boldsymbol{P}^{-1}\boldsymbol{P}\right) = \operatorname{Tr}\boldsymbol{T}_D^N$$

In the case of the one-dimensional Ising model we are considering $\boldsymbol{T}_D$ is a 2×2 matrix so it will have two eigenvalues which we call λ_+ and λ_-, with the convention $\lambda_+ > \lambda_-$ (we could in principle also consider the case $\lambda_+ = \lambda_-$, but we will shortly see why this is not necessary). We will therefore have:

$$Z_N = \lambda_+^N + \lambda_-^N$$

In general, if $\boldsymbol{T}$ is a $(n+2) \times (n+2)$ matrix whose eigenvalues are $\lambda_+ > \lambda_- > \lambda_1 > \cdots > \lambda_n$ we will have:

$$Z_N = \lambda_+^N + \lambda_-^N + \sum_{i=1}^{n}\lambda_i^N \tag{5.13}$$

Let us note that the dimension of the transfer matrix can increase if we consider interactions with longer ranges or if we allow the spin variables to assume more than one value[28], and clearly the larger the matrix the harder its eigenvalues are to be computed, but the principle is always the same.

We will now use the transfer matrix in order to compute some interesting properties of a generic system.

Free energy

Considering a general situation, the partition function of a model can be written with the use of the transfer matrix as in (5.13). Therefore, in the thermodynamic limit the free energy of the system will be:

$$f = -k_BT\lim_{N\to\infty}\frac{1}{N}\ln Z_N = -k_BT\lim_{N\to\infty}\frac{1}{N}\ln\left[\lambda_+^N\left(1 + \frac{\lambda_-^N}{\lambda_+^N} + \sum_{i=1}^{n}\frac{\lambda_i^N}{\lambda_+^N}\right)\right]$$

Since $\lambda_+ > \lambda_- > \lambda_i$ we have:

$$\left(\frac{\lambda_-}{\lambda_+}\right)^N \stackrel{N\to\infty}{\longrightarrow} 0 \qquad\qquad \left(\frac{\lambda_i}{\lambda_+}\right)^N \stackrel{N\to\infty}{\longrightarrow} 0$$

and therefore:

$$f = -k_BT\ln\lambda_+$$

---

[27]Namely, $\operatorname{Tr}(ABC) = \operatorname{Tr}(BCA) = \operatorname{Tr}(CAB)$.

[28]In this case (which will be studied later on, see 6.3.2) the model is called *Potts model*; if the spin variables can assume q different values, then the transfer matrix of a one-dimensional Potts model will be $q \times q$.

This is an extremely important result, since this means that the *entire* thermodynamics of the system can be obtained by only knowing the largest eigenvalue of the transfer matrix[29]. Furthermore, the fact that only λ_+ is involved in the expression of the free energy in the thermodynamic limit has a very important consequence on the possibility for phase transitions to occur. In fact, there exists a theorem called *Perron-Frobenius theorem* in linear algebra which states the following:

Theorem (Perron-Frobenius). *If A is an $n \times n$ square matrix (with n finite) such that all its elements are positive, namely $A_{ij} > 0 \quad \forall i, j$, then the eigenvalue λ_+ with largest magnitude is:*

1. *real and positive*

2. *non-degenerate*

3. *an analytic function of the elements A_{ij}*

We omit the proof of this theorem.

This means that if the transfer matrix of a model satisfies such properties, since λ_+ is an analytic function the system will *never* exhibit phase transitions because also f will be analytic.

For the one-dimensional Ising model with nearest neighbour interaction that we are considering, these properties are satisfied and so we have:

1. $\lambda_+ \neq 0$, so that f is well defined

2. $\lambda_+ \neq \lambda_-$ (this justifies *a posteriori* why we have considered $\lambda_+ \neq \lambda_-$ from the beginning)

3. λ_+ is analytic, and therefore so is f

From the last fact we deduce that no phase transition can occur for $T \neq 0$; if $T = 0$ some of the elements of $\boldsymbol{T}$ diverge and Perron-Frobenius theorem can't be applied[30].

In general, in higher dimensions or with different kinds of interactions the transfer matrix can become infinite-dimensional in the thermodynamic limit: in this case the assumptions of Perron-Frobenius theorem don't hold and so the system can actually exhibit phase transitions (since λ_+ is not necessarily an analytic function any more).

Correlation function and correlation length

The transfer matrix method can also be used to compute the correlation function and so the correlation length of a system.
As we know (see 5.1.2) in order to do that we first have to compute the two-point correlation function.

The connected correlation function of two spins which are at R sites of distance is defined as:

$$G_R = \langle S_1 S_R \rangle - \langle S_1 \rangle \langle S_R \rangle$$

where we have considered the first and the R-th spins because we are assuming periodic boundary conditions (so our choice is equivalent to considering two generic spins at sites i

---

[29]This is a blessing also from a computational point of view: it often happens, in fact, that the exact expression of the transfer matrix can be obtained but it is too big or complicated to diagonalize completely. There are however several algorithms that allows one to compute only the largest eigenvalue of the matrix in a rather efficient way.

[30]This agrees with what we have noted in 5.4.1 about the fact that a "phase transition" occurs in the one-dimensional Ising model for $T = 0$.

and $i + R$). For very large distances, we know that the correlation function decays exponentially, namely $G_R \sim e^{-R/\xi}$ for $R \to \infty$, where ξ is the correlation length. Therefore we can define the correlation length ξ of the system as:

$$\xi^{-1} = \lim_{R \to \infty} \left(-\frac{1}{R} \ln G_R \right)$$

We begin by computing $\langle S_1 S_R \rangle$; this is the thermodynamic limit of the quantity:

$$\langle S_1 S_R \rangle_N = \frac{1}{Z_N} \sum_{\{S_i = \pm 1\}} S_1 S_R e^{-\beta \mathcal{H}}$$

Using the same factorization of the Hamiltonian that we have previously seen and that led to equation (5.11), we have:

$$\langle S_1 S_R \rangle_N = \frac{1}{Z_N} \sum_{\{S_i = \pm 1\}} S_1 \langle S_1 | \boldsymbol{T} | S_2 \rangle \langle S_2 | \boldsymbol{T} | S_3 \rangle \cdots \langle S_{R-1} | \boldsymbol{T} | S_R \rangle S_R \langle S_R | \boldsymbol{T} | S_{R+1} \rangle \cdots \langle S_N | \boldsymbol{T} | S_1 \rangle =$$

$$= \frac{1}{Z_N} \sum_{S_1, S_R} S_1 \langle S_1 | \boldsymbol{T}^R | S_R \rangle S_R \langle S_R | \boldsymbol{T}^{N-R} | S_1 \rangle$$

Now, we can write:

$$\boldsymbol{T} = \sum_i |t_i\rangle \lambda_i \langle t_i|$$

where $|t_i\rangle$ are the eigenvectors of $\boldsymbol{T}$ and λ_i their relative eigenvalues (again ordered so that $\lambda_+ > \lambda_- > \lambda_1 > \cdots$); this way since these eigenvectors are orthonormal we also have:

$$\boldsymbol{T}^n = \sum_i |t_i\rangle \lambda_i^n \langle t_i|$$

Therefore:

$$\langle S_1 S_R \rangle_N = \frac{1}{Z_N} \sum_{S_1, S_R} \sum_{i,j} S_1 \langle S_1 | t_i \rangle \lambda_i^R \langle t_i | S_R \rangle S_R \langle S_R | t_j \rangle \lambda_j^{N-R} \langle t_j | S_1 \rangle \tag{5.14}$$

Now, we introduce the matrices:

$$\boldsymbol{S_i} = \sum_{S_i} |S_i\rangle S_i \langle S_i|$$

which are diagonal matrices such that on their diagonal there are all the possible spin values at the i-th site. This way, moving $\langle t_j | S_1 \rangle$ at the beginning of the sum in equation (5.14) (since it is simply a number) and summing over S_1 and S_R we get:

$$\langle S_1 S_R \rangle_N = \frac{1}{Z_N} \sum_{i,j} \langle t_j | \boldsymbol{S_1} | t_i \rangle \lambda_i^R \langle t_i | \boldsymbol{S_R} | t_j \rangle \lambda_j^{N-R}$$

and using the expression of Z_N given by equation (5.13):

$$\langle S_1 S_R \rangle_N = \frac{\sum_{i,j} \langle t_j | \boldsymbol{S_1} | t_i \rangle \lambda_i^R \langle t_i | \boldsymbol{S_R} | t_j \rangle \lambda_j^{N-R}}{\sum_k \lambda_k^N}$$

Multiplying and dividing by λ_+^N we get:

$$\langle S_1 S_R \rangle_N = \frac{\sum_{i,j} \langle t_j | \boldsymbol{S_1} | t_i \rangle (\lambda_i/\lambda_+)^R \langle t_i | \boldsymbol{S_R} | t_j \rangle (\lambda_j/\lambda_+)^{N-R}}{\sum_k (\lambda_k/\lambda_+)^N}$$

In the thermodynamic limit the surviving terms are those containing $\lambda_j = \lambda_+$ and $\lambda_k = \lambda_+$, and also all the terms with λ_i (because they are not affected by the limit in N), so:

$$\langle S_1 S_R \rangle = \lim_{N \to \infty} \langle S_1 S_R \rangle_N = \sum_i \left(\frac{\lambda_i}{\lambda_+} \right)^R \langle t_+|\boldsymbol{S_1}|t_i\rangle \langle t_i|\boldsymbol{S_R}|t_+\rangle =$$

$$\tag{5.15}$$

$$= \langle t_+|\boldsymbol{S_1}|t_+\rangle \langle t_+|\boldsymbol{S_R}|t_+\rangle + \sum_{i \neq +} \left(\frac{\lambda_i}{\lambda_+} \right)^R \langle t_+|\boldsymbol{S_1}|t_i\rangle \langle t_i|\boldsymbol{S_R}|t_+\rangle$$

where we have used the symbolic notation "$i \neq +$" in the sum to indicate that we are excluding the case $\lambda_i = \lambda_+$.

What we now want to show is that:

$$\langle S_i \rangle = \lim_{N \to \infty} \langle S_i \rangle_N = \langle t_+|\boldsymbol{S_i}|t_+\rangle$$

Considering $S_i = S_R$ and proceeding like we have done for $\langle S_1 S_R \rangle_N$, we get to:

$$\langle S_R \rangle_N = \frac{1}{Z_N} \sum_{\{S_i = \pm 1\}} S_R e^{-\beta \mathcal{H}} = \frac{1}{Z_N} \sum_{S_1, S_R} \langle S_1|\boldsymbol{T}^R|S_R\rangle S_R \langle S_R|\boldsymbol{T}^{N-R}|S_1\rangle =$$

$$= \frac{1}{Z_N} \sum_{S_1, S_R} \sum_{i,j} \langle t_j|S_1\rangle \langle S_1|t_i\rangle \lambda_i^R \langle t_i|S_R\rangle S_R \langle S_R|t_j\rangle \lambda_j^{N-R} =$$

$$\frac{1}{Z_N} \sum_{i,j} \underbrace{\langle t_j|t_i\rangle}_{\delta_{ij}} \lambda_i^R \langle t_i|\boldsymbol{S_R}|t_j\rangle \lambda_j^{N-R}$$

and again, using equation (5.13) and multiplying and dividing by λ_+:

$$\langle S_R \rangle_N = \frac{\sum_i (\lambda_i/\lambda_+)^R \langle t_i|\boldsymbol{S_R}|t_i\rangle (\lambda_i/\lambda_+)^{N-R}}{\sum_k (\lambda_k/\lambda_+)^N} = \frac{\sum_i (\lambda_i/\lambda_+)^N \langle t_i|\boldsymbol{S_R}|t_i\rangle}{\sum_k (\lambda_k/\lambda_+)^N}$$

Again, the only surviving term in the thermodynamic limit is that with $\lambda_i = \lambda_k = \lambda_+$, so indeed:

$$\langle S_R \rangle = \lim_{N \to \infty} \langle S_R \rangle_N = \langle t_+|\boldsymbol{S_R}|t_+\rangle$$

This way, equation (5.15) becomes:

$$\langle S_1 S_R \rangle = \langle S_1 \rangle \langle S_R \rangle + \sum_{i \neq +} \left(\frac{\lambda_i}{\lambda_+} \right)^R \langle t_+|\boldsymbol{S_1}|t_i\rangle \langle t_i|\boldsymbol{S_R}|t_+\rangle$$

and thus the connected correlation function is:

$$G_R = \sum_{i \neq +} \left(\frac{\lambda_i}{\lambda_+} \right)^R \langle t_+|\boldsymbol{S_1}|t_i\rangle \langle t_i|\boldsymbol{S_R}|t_+\rangle$$

If we now take the limit $R \to \infty$ the leading term will be that with the largest possible eigenvalue λ_i, i.e. λ_- and all the other will vanish. Therefore:

$$G_R \overset{R \to \infty}{\sim} \left(\frac{\lambda_-}{\lambda_+} \right)^R \langle t_+|\boldsymbol{S_1}|t_-\rangle \langle t_-|\boldsymbol{S_R}|t_+\rangle$$

and thus the correlation length will be such that:

$$\xi^{-1} = \lim_{R \to \infty} \left(-\frac{1}{R} \ln \left[\left(\frac{\lambda_-}{\lambda_+} \right)^R \langle t_+ | \boldsymbol{S_1} | t_- \rangle \langle t_- | \boldsymbol{S_R} | t_+ \rangle \right] \right) = \ln \frac{\lambda_+}{\lambda_-}$$

(since $\langle t_+ | \boldsymbol{S_1} | t_- \rangle$ and $\langle t_- | \boldsymbol{S_R} | t_+ \rangle$ are just numbers). Therefore:

$$\xi = \left(\ln \frac{\lambda_+}{\lambda_-} \right)^{-1}$$

Explicit computations for the one-dimensional Ising model

We now want to apply what we have just shown in order to do some explicit computations on the one-dimensional Ising model.
We have seen (equation (5.12)) that the explicit expression of the transfer matrix for the one-dimensional Ising model is:

$$\boldsymbol{T} = \begin{pmatrix} e^{K+h} & e^{-K} \\ e^{-K} & e^{K-h} \end{pmatrix}$$

Its eigenvalues can be determined, as usual, solving the equation $\det(\boldsymbol{T} - \lambda \mathbb{I}) = 0$, which yields:

$$\lambda_\pm = e^K \left(\cosh h \pm \sqrt{\sinh^2 h + e^{-4K}} \right)$$

Thus, the free energy of the system in the thermodynamic limit is:

$$f(K, h) = -J - k_B T \ln \left(\cosh h + \sqrt{\sinh^2 h + e^{-4K}} \right)$$

while its magnetization (remembering that $h = \beta H$) is:

$$m = -\frac{\partial f}{\partial H} = -\frac{1}{k_B T} \frac{\partial f}{\partial h} = \frac{\sinh h + \frac{\sinh h \cosh h}{\sqrt{\sinh^2 h + e^{-4K}}}}{\cosh h + \sqrt{\sinh^2 h + e^{-4K}}} = \frac{\sinh h}{\sqrt{\sinh^2 h + e^{-4K}}}$$

Let us note that for $H \to 0$ at fixed T, since $\sinh h = \sinh(\beta H) \to 0$ then m vanishes: this is again the expression that the Ising model in one dimension does not exhibit phase transitions. Furthermore we can see that in the limit $T \to 0$, namely $\beta \to \infty$, we have $h, K \to \infty$ and thus $m \to 1$: again, the only value of the temperature for which the unidimensional Ising model with nearest-neighbour interactions exhibits a spontaneous magnetization is $T = 0$.

The isothermal susceptibility of the system is:

$$\chi_T = \frac{\partial m}{\partial H} = \frac{1}{k_B T} \frac{\partial m}{\partial h}$$

Instead of explicitly calculating χ_T for h generic, since we are interested in the behaviour of m when there are no external fields let us see what happens for small values of h. Since in this case $\sinh h \sim h$ we have:

$$m \sim \frac{h}{\sqrt{h^2 + e^{-4K}}} \sim \frac{h}{e^{-2K}} = h e^{2K}$$

and therefore:

$$\chi_T \sim \frac{e^{2K}}{k_B T}$$

For high and low temperatures, we get:

$$\chi_T \overset{T \to \infty}{\sim} \frac{1}{k_B T} \qquad\qquad \chi_T \overset{T \to 0}{\sim} \frac{e^{\frac{2J}{k_B T}}}{k_B T}$$

and as we can see χ_T diverges exponentially for $T \to 0$; this agrees with the fact that (as already previously stated) for $T \to 0$ some elements of the transfer matrix diverge and thus the Perron-Frobenius theorem can't be applied.

Considering now the correlation length, we have:

$$\xi^{-1} = -\ln\left(\frac{\cosh h - \sqrt{\sinh^2 h + e^{-4K}}}{\cosh h + \sqrt{\sinh^2 h + e^{-4K}}}\right)$$

In the particular case $H = 0$:

$$\xi^{-1}_{H=0} = -\ln\frac{1 - e^{-2K}}{1 + e^{-2K}} = -\ln\frac{1}{\cotanh K} \qquad \Rightarrow \qquad \xi_{H=0} = \frac{1}{\ln(\cotanh K)}$$

Now, in the limit $T \to 0$, namely $K \to \infty$, the following asymptotic expansion holds:

$$\cotanh K \sim 1 + 2e^{-2K} + O(e^{-4K})$$

Therefore:

$$\xi_{H=0} \sim \frac{1}{\ln(1 + 2e^{-2K})} \sim \frac{1}{2e^{-2K}} = \frac{e^{\frac{2J}{k_B T}}}{2}$$

and so again we find an exponential divergence for $T \to 0$.
On the other hand, for $T \to \infty$, namely $K \to 0$:

$$\tanh K \sim \frac{1 + K^2}{K} \qquad \Rightarrow \qquad \ln(\tanh K) \approx -\ln K + \ln(1 + K^2) \overset{K\to 0}{\sim} -\ln K$$

Thus:

$$\xi_{H=0} \sim -\frac{1}{\ln K} \qquad \Rightarrow \qquad \xi_{H=0} \overset{K\to\infty}{\longrightarrow} 0$$

as expected[31].

5.5 A slightly trickier system: the Heisenberg model

We can now apply the transfer matrix method to a more complicated system, the so called *Heisenberg model*. This model is identical to the Ising one with the exception that this time the spin variables can assume *vectorial* values instead of being discrete scalars. In other words the degrees of freedom of the system will be N vectors $\vec{S}_i$, each residing on a site of the lattice and with unitary module, namely $|\vec{S}_i| = 1$.
If there is no external field[32], i.e. $H = 0$, then the reduced Hamiltonian of a one-dimensional Heisenberg model is:

$$-\beta\mathcal{H} = K\sum_{i=1} \vec{S}_i \cdot \vec{S}_{i+1}$$

and with periodic boundary conditions ($\vec{S}_{N+1} = \vec{S}_1$):

$$Z_N = \sum_{\{\vec{S}_i\}} e^{K\sum_{i=1}^{N} \vec{S}_i \cdot \vec{S}_{i+1}} = \sum_{\{\vec{S}_i\}} e^{K\vec{S}_1 \cdot \vec{S}_2} e^{K\vec{S}_2 \cdot \vec{S}_3} \cdots e^{K\vec{S}_N \cdot \vec{S}_1} = \operatorname{Tr} \boldsymbol{T}^N$$

---

[31]Remember that in general correlation lengths are negligible for large temperatures and become relevant near a critical point.
[32]We consider this particular case because otherwise we can't apply the transfer matrix method.

where the transfer matrix $\boldsymbol{T}$ is such that:

$$\langle \vec{S}_i|\boldsymbol{T}|\vec{S}_{i+1}\rangle = e^{K\vec{S}_i\cdot\vec{S}_{i+1}}$$

Also in this case the transfer matrix can be written in the form:

$$\boldsymbol{T} = \sum_i \lambda_i |t_i\rangle\langle t_i|$$

so that, for example:

$$\langle \vec{S}_1|\boldsymbol{T}|\vec{S}_2\rangle = e^{K\vec{S}_1\cdot\vec{S}_2} = \sum_i \lambda_i\langle \vec{S}_1|t_i\rangle\langle t_i|\vec{S}_2\rangle = \sum_i \lambda_i f_i(\vec{S}_1)f_i^*(\vec{S}_2)$$

where $f_i(\vec{S}_j) = \langle \vec{S}_j|t_i\rangle$ is in general a complex function.

Now, in order to determine the eigenvalues λ_i we must use the fact (which we will not prove) that plane waves $e^{-i\vec{r}\cdot\vec{k}}$ can be decomposed in spherical harmonics $Y_{\ell m}$ in this way:

$$e^{-i\vec{r}\cdot\vec{k}} = 4\pi\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} i^{\ell}j_{\ell}(|\vec{k}||\vec{r}|)Y_{\ell m}^*(\hat{k})Y_{\ell m}(\hat{r})$$

where j_ℓ are the *spherical Bessel functions*:

$$j_\ell(xy) = -\frac{i^\ell}{2}\int_0^\pi \sin\theta e^{ixy\cos\theta}P_\ell(\cos\theta)d\theta$$

and $P_\ell(\cos\theta)$ is the ℓ-th order Legendre polynomial.
In this case, setting $\vec{r} = \vec{S}_1$ and $\vec{k} = \vec{S}_2$ we get:

$$e^{k\vec{S}_1\cdot\vec{S}_2} = 4\pi\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} i^\ell j_\ell(-iK)Y_{\ell m}^*(\vec{S}_1)Y_{\ell m}(\vec{S}_2)$$

This suggests that the eigenvalues have the form:

$$\lambda_{\ell m}(K) = 4\pi i^\ell j_\ell(-iK)$$

(which actually doesn't depend on m). Therefore, the partition function is:

$$Z_N = \text{Tr}\,\boldsymbol{T}^N = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\lambda_{\ell m}(K)^N = \sum_{\ell=0}^{\infty}(2\ell+1)\lambda_\ell^N(K)$$

And from now on, from the explicit computation of the eigenvalues (which we are not going to do) we can determine all the thermodynamics of the system.

5.6 The Ising model in higher dimensions

If we now try to study some more complex configurations of the Ising model, we immediately encounter huge difficulties which make it impossible to exactly compute the partition function of the system. In particular, Onsager managed (with a huge effort) in 1944 to solve the problem for a two-dimensional Ising model in absence of external fields, but in all other cases (two-dimensional model with external field, or three-dimensional model) we still don't know an exact solution.

However, even if we don't know much in these cases there is still a lot to learn: in particular, from Onsager's solution[33] we can see that already in two dimensions an Ising model can exhibit phase transitions, showing a non null spontaneous magnetization for temperatures low enough.

Let us therefore consider a two-dimensional Ising model, defined on a lattice made of N rows and M columns. Applying periodic boundary conditions to the system in both directions (geometrically, this can be thought of as defining the model on a torus), and considering only nearest neighbour interactions, the reduced Hamiltonian of the system will be:

$$-\beta\mathcal{H} = K\sum_{\langle ij\rangle} S_i S_j + h\sum_i S_i$$

where we remember that $K = \beta J$ and $h = \beta H$. If we label each site of the lattice with the couple (m, n) where m is the number of the column and n of the row to which the site belongs, then we can rewrite[34]:

$$-\beta\mathcal{H} = K\sum_{n=1}^{N}\sum_{m=1}^{M}(S_{m,n}S_{m+1,n} + S_{m,n}S_{m,n+1}) + h\sum_{m,n} S_{m,n}$$

If we now call μ_m the set of spins belonging to the m-th column:

$$\mu_m = \{S_{m,1}, \ldots, S_{m,n}\}$$

and define:

$$E(\mu_m) = K\sum_{n=1}^{N} S_{m,n}S_{m,n+1} + h\sum_{n=1}^{N} S_{m,n}$$

$$E(\mu_m, \mu_{m+1}) = K\sum_{n=1}^{N} S_{m,n}S_{m+1,n}$$

we can write:

$$-\beta\mathcal{H} = \sum_{m=1}^{M}[E(\mu_m, \mu_{m+1}) + E(\mu_m)]$$

Therefore, defining the transfer matrix $\boldsymbol{T}$ so that:

$$\langle\mu_m|\boldsymbol{T}|\mu_k\rangle = e^{E(\mu_m,\mu_k)+E(\mu_m)}$$

the partition function will be:

$$Z_N = \operatorname{Tr}\boldsymbol{T}^N$$

and the thermodynamics of the system can be derived from the eigenvalue of $\boldsymbol{T}$ with largest magnitude. However, since $\boldsymbol{T}$ is a $2^N \times 2^N$ matrix, this is a rather difficult problem (the matrix becomes infinite in the thermodynamic limit!).
Onsager has shown that in the thermodynamic limit and for $H = 0$ the free energy of the system is:

$$f = -k_B T\ln[2\cosh(2\beta J)] - \frac{k_B T}{2\pi}\int_0^{\pi}\ln\left[\frac{1}{2}\left(1 + \sqrt{1 - g^2\sin^2\phi}\right)\right]d\phi$$

---

[33]We will not deduce it, and just limit ourselves to show it. However, in 5.7 we will use qualitative arguments to show that indeed the dimension of an Ising model must be at least two if we want phase transitions to occur.

[34]Note, again, that the sum over nearest neighbours is done so that we don't count twice the same terms.

(a) one-dimensional Ising model at $T = 0$ (b) one-dimensional Ising model with a spin flipped

(c) one-dimensional Ising model with a domain flipped
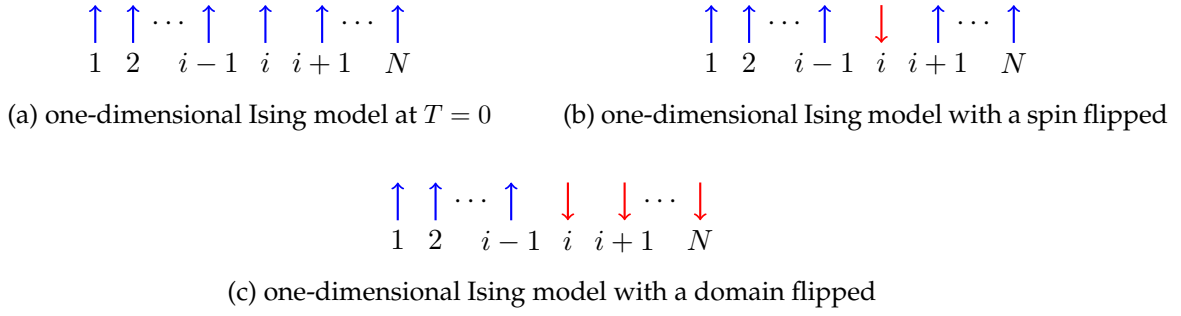
Figure 5.2: Spin flipping for a one-dimensional Ising model

where:

$$g = \frac{2}{\cosh(2\beta J) \coth(2\beta J)}$$

and also that the magnetization is:

$$m = \begin{cases} \left[1 - \sinh^{-4}(2\beta J)\right]^{1/8} & T < T_c \\ 0 & T > T_c \end{cases}$$

where T_c is the temperature given by the condition $2\tanh^2(2\beta J) = 1$, which yields the numeric result:

$$T_c \approx 2.27 \frac{J}{k_B} \tag{5.16}$$

This means that there is indeed a phase transition at $T = T_c$.
Onsager also showed that the critical exponents of this model are:

$$\alpha = 0 \qquad\qquad \beta = \frac{1}{8} \qquad\qquad \gamma = \frac{7}{4}$$

where $\alpha = 0$ because the specific heat diverges logarithmically for $T \sim T_c$.

5.7　Additional remarks on the Ising model

We now conclude the treatment of the Ising model with some additional remarks that allow us to understand a little more about its properties.

5.7.1　The role of dimensionality

We have seen (although not really directly) that the dimensionality of the Ising model is crucial for the existence of phase transitions; in particular we have seen that for $d = 1$ there are no phase transitions, while for $d > 1$ they can occur. Sometimes the dimension of a model above which phase transitions occur is called *lower critical dimension*, so in our case we can say that the lower critical dimension of the Ising model is one.

We shall now use a heuristic argument in order to show that this is indeed the case for the Ising model, and that for $d = 1$ there can be no *long range order*. This argument will also allow us to estimate (even if a bit roughly) the critical temperature of the two-dimensional Ising model.

One dimension

Let us consider a one-dimensional Ising model at $T = 0$; we know that in this case the system is completely ordered, namely all the spins point in the same direction, say upwards. If we now increase the temperature a little bit, then some spins will randomly flip due to thermal fluctuations; what we want to see is if the ordered state of the system is stable under this random spin flips[35].

Therefore, let us consider a one-dimensional Ising model of N spins at $T = 0$ (with all spins pointing upwards) with nearest neighbour interactions and without any external field, i.e. the Hamiltonian of the system is:

$$\mathcal{H} = -J \sum_{\langle ij \rangle} S_i S_j$$

and for the sake of convenience let us take periodic boundary conditions.

The entropy of such a state is null (the system has only one possible configuration) so its free energy will be equal to its internal energy, and since there are N parallel bonds we will have:

$$F_N^{\text{ord.}} = U_N^{\text{ord.}} = -JN$$

Let us now flip a single spin (see figure 5.2); in this case, since two previously parallel bonds become antiparallel, the internal energy of the system becomes:

$$U_N^{\text{flip}} = -J(N-2) + 2J$$

and as of the entropy of the system:

$$S_N^{\text{flip}} = k_B \ln N$$

since the flipped spin can be in any of the N sites of the lattice (namely, the system can have N different configurations). Therefore, from the thermodynamic definition of free energy ($F = U - TS$):

$$F_N^{\text{flip}} = -J(N-2) + 2J - k_B T \ln N$$

and the variation of free energy due to the spin flipping is:

$$\Delta F_N = F_N^{\text{flip}} - F_N^{\text{ord.}} = 4J - k_B T \ln N$$

We can therefore see that for fixed $T > 0$, $\Delta F \to -\infty$ if $N \to \infty$: in the thermodynamic limit it is energetically convenient for the system to flip spins. We can therefore continue with the operation of spin flipping until there are no other parallel spins left: the long range order of the system is unstable under thermal fluctuations, and so as soon as $T > 0$ the one-dimensional Ising model can't exhibit a spontaneous magnetization (namely, there are no phase transitions). We could have obtained the same result with a slightly different approach, namely considering the system divided into two different magnetic domains, positively and negatively magnetized, respectively (see again figure 5.2). In this case the internal energy of the system is:

$$U_N^{\text{flip}} = -J(N-2) + 2J = -JN + 4J$$

On the other hand the entropy of the system is again the same. Therefore, we should have found $\Delta F_N = F_N^{\text{flip}} - F_N^{\text{ord}} = 4J - k_B T \ln N$, deducing the same conclusions.
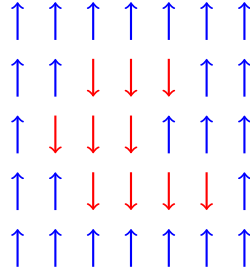
Figure 5.3: two-dimensional Ising model with a domain flipped

Two dimensions

We now consider a two-dimensional Ising model, with the same properties as before, defined on a lattice with coordination number z. Let us consider the system at $T = 0$ with all the spins pointing upwards, and flip a domain of spins; the boundary of this domain will be a closed path, and let us suppose that it is made of n bonds (see figure 5.3). The difference in internal energy between the ordered state and one with a flipped domain is therefore[36]:

$$\Delta U = 2Jn$$

The entropy is much trickier to compute exactly, so we will only give an estimate. If we consider the boundary of the domain as a random walk an upper limit to the number of possible configurations of this boundary is z^n; however, we are not taking into account the fact that the domain wall cannot intersect itself (otherwise there would be more than one domain), namely it must be a *self-avoiding* random walk. A first rough correction to this estimate could be supposing that at each step the domain wall can only go in $z - 1$ directions since it must not immediately go back on itself, so a slightly better upper limit to the number of possible configurations of the system is $(z - 1)^n$. Thus we can estimate the difference in entropy between the ordered and the domain-flipped state as:

$$\Delta S \approx k_B T n \ln(z - 1)$$

However, our assumption still allows the domain boundary to intersect itself so we are surely overestimating ΔS.

Therefore, the change in free energy due to the flipping of a domain of spins is:

$$\Delta F = [2J - k_B T \ln(z - 1)]\, n$$

We now see that the behaviour of ΔF is not well defined in the thermodynamic limit, namely when $n \to \infty$. In particular we will have $\Delta F \to -\infty$ only if:

$$T > \frac{2J}{k_B \ln(z - 1)} := T_c$$

We therefore find something very interesting: in two dimensions the long range order of the Ising model is unstable under thermal fluctuations only when $T > T_c$, while for $T < T_c$ it is stable, or in other words the system will always exhibit a spontaneous magnetization even in the absence of any external fields. We thus have found that for $d = 2$ the Ising model

---

[35]Of course, in the thermodynamic limit the flipping of a *finite* number of spins will not be sufficient, in general, in order to destroy the long range order; the only way to do so is flipping a non-zero *fraction* of spins.

[36]This can be understood thinking about the one-dimensional case. For $d = 1$ in fact we have $\Delta U = 4J$ when we flip a single spin (and therefore there are two antiparallel bonds), so we can argue that if we flip N spins the variation of internal energy is $2Jn$ (if not exactly, at least of the same order).

undergoes a phase transition at $T = T_c$; let us note that this critical temperature depends on the coordination number z, so it is a characteristic of the lattice considered and not a universal property of the system[37]. If for example we consider a square lattice, $z = 4$ and we get $T_c \approx 1.82 \frac{J}{k_B}$, while as we have seen in 5.6 (equation (5.16)) the exact result gives $T_c \approx 2.27 \frac{J}{k_B}$: this discrepancy is due to the fact that we have overestimated the change in entropy of the system.

5.7.2 The role of symmetry

We can also see that the symmetry of the Hamiltonian of the system plays an important role for the possibility of phase transitions, and also in determining the lower critical dimension of the system.

In particular the fact that for the Ising model phase transitions can occur for $d > 1$ relies also on the fact that the Hamiltonian of the system has a *discrete* symmetry, namely its symmetry group is $\mathbb{Z}_2$. This means that the boundary of magnetic domains has *finite* "thickness" (one lattice unit, since a boundary in this case simply separates two adjacent antiparallel spins).

However, if the degrees of freedom have a *continuous* symmetry things start getting different. For example, the Hamiltonian of the Heisenberg model (see 5.5) is:

$$\mathcal{H} = -J \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j - \vec{H} \cdot \sum_i \vec{S}_i$$

and in the case $\vec{H} = 0$ the system has a *rotational invariance*. In fact, if $R \in O(3)$ then we easily see that:

$$\mathcal{H} \left(\{R\vec{S}_i\}, H = 0 \right) = \mathcal{H} \left(\{\vec{S}_i\}, H = 0 \right)$$

(of course if $\vec{H} \neq 0$ then the Hamiltonian will be invariant only under $O(2)$ transformations). In this case since the system has a wider symmetry its entropy will be much larger[38], and so we need more energy if we want long range order in the Heisenberg model at $T > 0$ (in other words, the dimension of the domain wall will be comparable to the size of the system).

5.7.3 The role of interaction range

In order to see how the range of the interactions between the degrees of freedom of the system affects its properties, let us consider a one-dimensional Ising model with infinite-ranged interactions:

$$-\mathcal{H} = \frac{J_0}{2} \sum_{i,j} S_i S_j + H \sum_i S_i$$

(note that the sum in the interaction term is not restricted, and the $1/2$ factor has been introduced for later convenience). This model can be solved with the technique of *Hubbard transformation*, also called *auxiliary field method*.

First, we must note that J_0 can't be a constant independent of the dimension of the system because the sum $\sum_{i,j} S_i S_j$ contains a number of terms of the order of N^2 and so in the thermodynamic limit it would diverge; we must therefore use the so called *Kac prescription*, setting $J_0 = J/N$ so that the thermodynamic limit exists. Under these assumptions the partition function of the system is:

$$Z_N = \sum_{\{S_i = \pm 1\}} e^{\frac{\beta J}{2N} \sum_{i,j} S_i S_j + \beta H \sum_i S_i}$$

---

[37]But, as we have stressed many times, the behaviour of the thermodynamic properties of the system in the neighbourhood of T_c is universal.

[38]Very intuitively, since there are "more directions" for the spins, there will be much more possible configurations for the system.

Since the double sum is not restricted, we have:

$$\sum_{i,j} S_i S_j = \left(\sum_i S_i\right)^2$$

If we now call $x = \sum_i S_i$ and $a = \beta J$, we can use the *Hubbard-Stratonovich identity*[39]:

$$e^{\frac{ax^2}{2N}} = \sqrt{\frac{aN}{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{aN}{2}y^2 + axy} dy \tag{5.17}$$

where $\operatorname{Re} a > 0$. The advantage of this approach is that the variable x, which contains all the degrees of freedom of the system, is linear and not quadratic in the exponential; however we have "paid" the price of having introduced another field, y (the *auxiliary field* from which this method takes its name).

The partition function then becomes:

$$Z_N = \sqrt{\frac{\beta JN}{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{\beta JN}{2}y^2} \sum_{\{S_i = \pm 1\}} e^{\beta(H+Jy)\sum_i S_i} dy$$

Physically this can be interpreted as the "mean value" of the partition functions of non interacting Ising models subjected to an external field $H - Jy$ whose component y is distributed along a Gaussian.

We can therefore write:

$$Z_N = \sqrt{\frac{\beta JN}{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{\beta JN}{2}y^2} Q_y dy \qquad\qquad Q_y = \sum_{\{S_i = \pm 1\}} e^{\beta(H+Jy)\sum_i S_i}$$

and Q_y can be easily computed factorizing the sum, similarly to what we have done for the Ising model in 5.4.1:

$$Q_y = \prod_i \sum_{S=\pm 1} e^{\beta(H+Jy)S} = 2^N \cosh^N\left[\beta(H+Jy)\right]$$

Therefore:

$$Z_N = \sqrt{\frac{\beta JN}{2\pi}} \int_{-\infty}^{+\infty} e^{N\mathcal{L}} dy \qquad\qquad \mathcal{L} = \ln\left(2\cosh\left[\beta(H+Jy)\right]\right) - \frac{\beta J}{2}y^2$$

Now, since the exponent in the integral that defines Z_N is extensive ($\mathcal{L}$ doesn't depend on N) and N is large, we can compute it using the *saddle point approximation* (see appendix B). This consists in approximating the integral with the largest value of the integrand, namely:

$$Z_N \sim \sqrt{\frac{\beta JN}{2\pi}} e^{N\mathcal{L}(y_0)} \sim e^{N\mathcal{L}(y_0)}$$

where y_0 is the maximum of $\mathcal{L}$, thus given by the condition:

$$\frac{\partial \mathcal{L}}{\partial y}\bigg|_{y_0} = 0$$

---

[39] It can be easily verified completing the square in the exponential:

$$-\frac{aN}{2}y^2 + axy = -\frac{aN}{2}\left(y - \frac{x}{2}\right)^2 + \frac{ax^2}{2N}$$

and computing the Gaussian integral.

which yields:
$$\tanh(h + Ky_0) = y_0$$

Since it must be a maximum, we also must have:

$$\frac{\partial^2 \mathcal{L}}{\partial y^2}\Big|_{y_0} < 0$$

Now, we can see that the physical meaning of y_0 is the magnetization of the system in the thermodynamic limit. In fact:

$$m = \lim_{N \to \infty} -\frac{1}{N}\frac{\partial F}{\partial H} = \lim_{N \to \infty} \frac{1}{N}\frac{\partial \ln Z}{\partial h} = \lim_{N \to \infty} \frac{1}{N}\frac{\partial}{\partial h}\left[N\ln\left(2\cosh(h + Ky_0)\right) - \frac{K}{2}y_0^2\right] =$$

$$= \lim_{N \to \infty} \frac{1}{N}N\tanh(h + Ky_0) = \tanh(h + Ky_0) = y_0$$

Since we are interested in determining if the system can exhibit a *spontaneous* magnetization, we consider the case $h = 0$; therefore we will have:

$$m = \tanh(Km) \tag{5.18}$$

which is a transcendent equation, so it can't be solved analytically.
However, we can solve it graphically. From figure 5.4 we can see that there are three possible cases (remembering that by definition $K = \beta J$):

- for $\beta J > 1$ there are three solutions, one for $m = 0$ and two at $\pm \overline{m}$

- for $\beta J = 1$ these three solutions coincide

- for $\beta J < 1$ there is only one solution: $m = 0$

This means that two possible non null solutions appear when:

$$T < T_c := \frac{J}{k_B}$$

Let us see which of these solutions are acceptable, i.e. which of these solutions are maxima of $\mathcal{L}$. Still in the case $h = 0$, we have:

$$\frac{\partial^2 \mathcal{L}}{\partial y^2} = K\left[K\operatorname{sech}^2(Ky) - 1\right]$$

and so:

$K \leq 1$: in this case $y_0 = 0$, so:
$$\frac{\partial^2 \mathcal{L}}{\partial y^2}\Big|_{y_0 = 0} = K(K - 1) < 0$$

so y_0 is indeed a maximum for $\mathcal{L}$

$K > 1$: if $y_0 = 0$ then:
$$\frac{\partial^2 \mathcal{L}}{\partial y^2}\Big|_{y_0 = 0} = K(K - 1) > 0$$

so this is not a maximum for $\mathcal{L}$, and thus is not an acceptable solution of (5.18). On the other hand, if $y_0 = \pm \overline{m}$ then (also using the evenness of sech):

$$\frac{\partial^2 \mathcal{L}}{\partial y^2}\Big|_{y_0 = \pm \overline{m}} = K\left[K\operatorname{sech}^2(K\overline{m}) - 1\right]$$
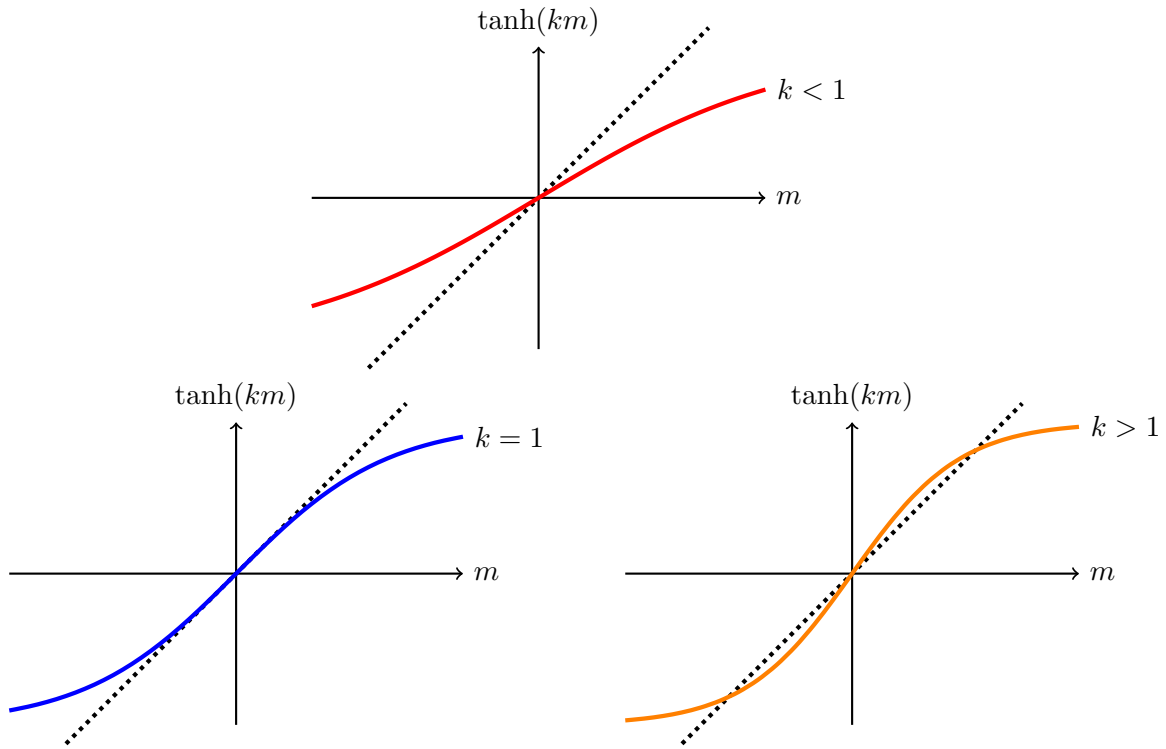
Figure 5.4: Graphical solution of equation (5.18)

and this time this derivative is negative, because:

$$K \operatorname{sech}^2(K\overline{m}) - 1 < 0 \qquad \Longleftrightarrow \qquad \operatorname{sech}(K\overline{m}) < \frac{1}{\sqrt{K}} \qquad \Longleftrightarrow \qquad \cosh(K\overline{m}) > \sqrt{K} > 1$$

which is always true.

Therefore, if $T > T_c = J/k_B$ the only acceptable value for the magnetization of the system is $m = \overline{m} = 0$, while if $T < T_c$ then $m = \pm\overline{m} \neq 0$: a phase transition has occurred, since now the system can exhibit a net spontaneous magnetization.

We thus see explicitly that if we let the interactions to be long-ranged the Ising model can undergo phase transitions already in the one-dimensional case.

Chapter 6

Mean field theories

It is a very rare fact that a model of interacting degrees of freedom can be solved *exactly*[1], and generally a model is not solvable in any dimension (e.g. the Ising model) so we must find other ways to study such systems in order to understand the possible occurrence of phase transitions and their behaviour near possible critical points. The most simple method, and the first to which one usually resorts to, is the so called *mean field approximation*.

One common feature of mean field theories is the identification of an appropriate order parameter; then, in very general and abstract words, there are two different approaches that can be taken (we shall see both):

– approximating an interacting system by a non-interacting one in a self-consistent external field expressed in terms of the order parameter

– expressing an approximate free energy in terms of this parameter and minimize the free energy with respect to the order parameter

In other (and maybe clearer) words, the first approach consists in substituting a system of interacting degrees of freedom with another system where these degrees of freedom do not interact but are subject to the action of an external *mean field*, which approximates the action of all the degrees of freedom on a single one; the second approach on the other hand is an "extension" in statistical mechanics of variational principles.

In order to be a little bit more explicit, let us see how the first approach applies to the Ising model[2]. In the case of nearest-neighbour interactions, the reduced Hamiltonian of the system is:

$$-\beta\mathcal{H} = K\sum_{\langle ij \rangle} S_i S_j + h\sum_i S_i$$

Consider now the single spin S_i; we can note that the first term of the reduced Hamiltonian can be written as:

$$K\sum_{\langle ij \rangle} S_i S_j = K\sum_i S_i \hat{h}_i(\{S\}) \tag{6.1}$$

where:

$$\hat{h}_i(\{S\}) = \sum_{j \in \text{n.n.}(i)} S_j$$

where with the notation $j \in \text{n.n.}(i)$ we mean that j is a nearest neighbour of i. We thus see that every spin S_i is subjected to an internal field $\hat{h}_i$ due to the presence of the nearest neighbouring

---

[1]An exhaustive review of exactly solvable models is given in [1].
[2]We now just make some qualitative observations; the mean field theory for the Ising model will be treated in detail shortly.

spins.

Now, if the number z of nearest neighbours is large (which happens in high dimensions, or equivalently we could assume interactions with a larger number of spins, for example including also the so called *next nearest neighbours*) this internal field can be approximated with the *mean field* generated by *all* the other spins in the lattice:

$$\frac{1}{z} \sum_{j \in \text{n.n.}(i)} S_j \approx \frac{1}{N} \sum_j S_j \tag{6.2}$$

This suggests that (6.1) can be rewritten as:

$$K \sum_i S_i \sum_j S_j \frac{z}{N}$$

If we consider an Ising model with an hypercubic lattice and simple nearest-neighbour interactions, we know that $z = 2d$ so the mean field term can be written as:

$$\hat{h}_{\text{m.f.}} = 2 \frac{dK}{N} \left(\sum_i S_i \right)^2$$

since the two sums are independent.

We therefore have that the approximation (6.2) (with $z = 2d$) is better the larger the dimensionality d of the system. From this very simple observation we can argue something that will become explicit and clear only later on (see 6.6.1): mean field theories are good approximations *only if the dimensionality of the system is large enough*.

Let us now come back to the general properties of mean field theories.

One of the main features of mean field theories is that they neglect the effects of fluctuations in the order parameter (in other words, within mean field theories the order parameter is supposed to be constant over all the system): on one hand we will see that this will make it possible to study a lot of systems, and to obtain loads of interesting and useful information about them, but on the other one we will see that this will be fatal for the reliability of mean field theories in the proximity of critical points, since they are characterised by the divergence of long-ranged fluctuations (see 5.1.2). This also means that mean field theories are in any case rather efficient far from critical points.

There are many possible mean field theories, and we are mainly going to study them applied to the Ising model (since it's the one we know better until now) or similar ones.

6.1   Weiss mean field theory for the Ising model

Let us now study the mean field theory of an Ising model, called *Weiss mean field theory*.

As usual, we start from the Hamiltonian of a nearest-neighbour interaction Ising model, so that the partition function of the system is[3]:

$$Z_N = \text{Tr}\, e^{-\beta \mathcal{H}} = \sum_{S_1 = \pm 1} \cdots \sum_{S_N = \pm 1} e^{\beta \left(\frac{1}{2} \sum_{\langle ij \rangle} J_{ij} S_i S_j + H \sum_i S_i \right)}$$

We define our order parameter as the single-spin magnetization:

$$m = \langle S_i \rangle = \langle S \rangle$$

---

[3]The $1/2$ factor has been introduced for later convenience (it can be reabsorbed in the definition of J_{ij}).

and we introduce it in the partition function through the identity:

$$S_i S_j = (S_i - m + m)(S_j - m + m) = -m^2 + m(S_i + S_j) + (S_i - m)(S_j - m)$$

Note that the last term is $(S_i - m)(S_j - m) = (S_i - \langle S \rangle)(S_j - \langle S \rangle)$, which exactly measures the fluctuations of the spin variables from their mean value. We now neglect it, so that in the partition function we can approximate:

$$\frac{1}{2} \sum_{\langle ij \rangle} J_{ij} S_i S_j \approx \frac{1}{2} \sum_{\langle ij \rangle} J_{ij} \left[-m^2 + m(S_i + S_j) \right]$$

If our system is isotropic then $J_{ij} = J$, so that we can write $\sum_{\langle ij \rangle} J_{ij} = Jz \sum_i$ where z is the coordination number of the lattice. This means that:

$$\frac{1}{2} \sum_{\langle ij \rangle} J_{ij} m(S_i + S_j) = Jzm \sum_i S_i$$

and the partition function reduces to:

$$Z_N = \sum_{S_1 = \pm 1} \cdots \sum_{S_N = \pm 1} e^{\beta \left(-Jzm^2 \frac{1}{2} \sum_i 1 + Jzm \sum_i S_i + H \sum_i S_i \right)} =$$

$$= e^{-\beta \frac{Jzm^2}{2} N} \sum_{S_1 = \pm 1} \cdots \sum_{S_N = \pm 1} e^{\beta(Jzm + H) \sum_i S_i} = e^{-\beta \frac{Jzm^2}{2} N} \sum_{S_1 = \pm 1} \cdots \sum_{S_N = \pm 1} \prod_{i=1}^{N} e^{\beta(Jzm + H)S_i} =$$

$$= e^{-\beta \frac{Jzm^2}{2} N} \left(\sum_{S = \pm 1} e^{\beta(Jzm + H)S} \right)^N = e^{-\beta \frac{Jzm^2}{2} N} \left[2 \cosh \left(\beta(Jzm + H) \right) \right]^N$$

Therefore the free energy of the system in the thermodynamic limit is:

$$f = - \lim_{N \to \infty} \frac{k_B T}{N} \ln Z_N = \frac{1}{2} Jzm^2 - k_B T \ln \left[\cosh \left(\beta(Jzm + H) \right) \right] - k_B T \ln 2 \qquad (6.3)$$

and the magnetization is:

$$m = -\frac{\partial f}{\partial H}_{|T} = \tanh \left[\beta(Jzm + H) \right]$$

We have thus found the so called *self-consistency equation* for the magnetization:

$$m = \tanh \left[\beta(Jzm + H) \right] \qquad (6.4)$$

which is almost the same as the one we have encountered in 5.7.3, i.e. equation (5.18).
We now proceed exactly as before. Since we are interested in studying the possible occurrence of phase transitions and in particular the existence of spontaneous magnetization we focus on the case $H = 0$:

$$m = \tanh(\beta Jzm) \qquad (6.5)$$

which is just like equation (5.18), but taking into account also the coordination number of the lattice.
In this case we will have that the phase transition occurs when $\beta Jz = 1$, i.e. at the temperature:

$$\beta_c Jz = 1 \qquad \Rightarrow \qquad T_c = \frac{z}{k_B} J$$
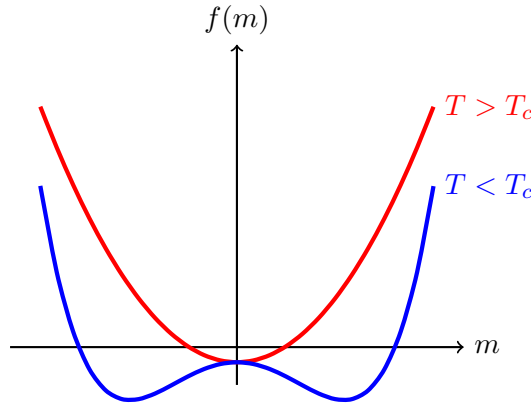
Figure 6.1: Trend of f for $m \approx 0$ in the two cases considered

Now, in order to better understand the physical properties of our system at different temperatures let us expand the free energy (6.3) around $m = 0$, still in the case $H = 0$. Using the Taylor series for $\cosh$ and $\ln$ we have $\ln(\cosh x) \sim x^2/2 - x^4/12 + \cdots$, so:

$$f(m, H = 0) \overset{m \approx 0}{\sim} -k_B T \ln 2 + \frac{Jz}{2}(1 - \beta Jz)m^2 + \frac{\beta^3}{12}J^4 z^4 m^4 \tag{6.6}$$

Therefore, the behaviour of f near $m = 0$ depends strongly on the sign of the coefficient of m^2; in particular:

– for $\beta Jz < 1$, namely for $T > T_c$ the free energy f has only one minimum for $m = 0$: this is the paramagnetic (disordered) phase

– for $\beta Jz > 1$, namely $T < T_c$ the free energy h has two stable minima at:

$$m = \pm \overline{m} := \pm\sqrt{\frac{3(\beta Jz - 1)}{(\beta Jz)^3}}$$

This is the magnetized (ordered) phase

Note that in the computations we have just made we have never imposed a particular value for the dimensionality of the system. This means that the results of this approximation should be valid also for $d = 1$, but we know that in one dimension the Ising model does not exhibit a phase transition.
This is an expression of the fact that *in the one-dimensional case mean field theory is not a good approximation* (again, the dimensionality of the system is still too low).

Let us finally note a fact. As we have previously stated mean field theories are characterized by the fact that the fluctuations of the order parameter are neglected, and we can see that the self-consistency equation (6.5) can be indeed obtained by neglecting the fluctuations of the single spins. In other words we can find it if we consider all the spins "frozen" to the value $\overline{m}$; in fact if we consider a single spin S we will have:

$$\langle S_0 \rangle = \overline{m} = \frac{\sum_{S=\pm 1} e^{\beta Jz\overline{m}S} S}{\sum_{S=\pm 1} e^{\beta Jz\overline{m}S}} = \frac{e^{\beta Jz\overline{m}} - e^{-\beta Jz\overline{m}}}{e^{\beta Jz\overline{m}} + e^{-\beta Jz\overline{m}}} = \tanh(\beta Jz\overline{m})$$

(where $\langle S \rangle = \overline{m}$ because we are requiring that the mean values of *all* the spins are equal).

6.1.1 Critical exponents of Weiss mean field theory for the Ising model

We can also try to understand something more about our system, for example computing its critical exponents[4].

Exponent β Let us begin with the exponent β, which (see 1.5 and table 1.2) is the one relative to the order parameter, namely the magnetization m. In order to find it we must see how the magnetization varies as a function of $T_c - T$; from the expansion of f we have just seen, imposing the extremal condition we get:

$$0 = \frac{\partial f}{\partial m}_{|\overline{m}} \sim Jz(1 - \beta Jz)\overline{m} + \frac{\beta^3}{3}J^4 z^4 \overline{m}^3 = \overline{m}\left[\frac{Jz}{T}(T - T_c) + \frac{\beta^3}{3}J^4 z^4 \overline{m}^2\right]$$

where we have used the definition of critical temperature $T_c = Jz/k_B$. Therefore (excluding of course the case $\overline{m} = 0$):

$$\overline{m}^2 \sim \text{const.} \cdot (T_c - T) \qquad \Rightarrow \qquad \overline{m} \sim (T_c - T)^{1/2}$$

We therefore have:

$$\beta = \frac{1}{2}$$

Exponent δ The exponent δ is the one that describes the behaviour of H as a function of m when $T = T_c$. To compute that let us begin from the self-consistency equation for m in the case $H \neq 0$, and invert it:

$$m = \tanh(\beta(Jzm + H)) \qquad \Rightarrow \qquad \beta(Jzm + H) = \tanh^{-1} m$$

Now expanding for small m, since $\tanh^{-1} x \sim x + x^3/3 + x^5/5$ we have:

$$Jzm + H \sim k_B Tm + \frac{k_B T}{3}m^3 \qquad \Rightarrow \qquad H \sim k_B m(T - T_c) + \frac{k_B T}{3}m^3$$

and if we set $T = T_c$:

$$H \sim m^3$$

Therefore:

$$\delta = 3$$

Exponent α This exponent is the one that describes the behaviour of the specific heat $C_H = -T\partial^2 f/\partial T^2_{|H}$ when $H = 0$. If $T > T_c$ then $\overline{m} = 0$, and since $f \sim -k_B T \ln 2$ we have $C_H = 0$. Therefore:

$$\alpha = 0$$

For $T < T_c$ the specific heat has a different dependence on the temperature, but in the end it turns out that C_H has a jump discontinuity at $T = T_c$, so from the definition of critical exponent (see 1.5) we indeed have $\alpha = 0$ (see 6.5.3 for a slightly more detailed computation).

---

[4]In fact, we have introduced mean field theories in order to be able to do something more than just see if phase transitions are possible. In particular, we would like to study the behaviour of a system near a critical point.

Exponent γ This is the exponent that regulates the isothermal susceptibility:

$$\chi_T = \frac{\partial m}{\partial H} = \frac{1}{\partial H / \partial m}$$

From the computation of the exponent δ we have seen that:

$$H \sim k_B m (T - T_c) + \frac{k_B T}{3} m^3$$

so:

$$\frac{\partial H}{\partial m} \sim k_B (T - T_c) + k_B T m^2$$

Therefore for small m, neglecting the quadratic term:

$$\chi_T \sim \frac{1}{k_B} (T - T_c)^{-1}$$

and thus:

$$\gamma = 1$$

To recap, the Weiss mean field theory for magnetic systems predicts the following values for the critical exponents:

$$\alpha = 0 \qquad\qquad \beta = \frac{1}{2} \qquad\qquad \gamma = 1 \qquad\qquad \delta = 3 \qquad\qquad (6.7)$$

We can immediately note that these exponents are different from those found by Onsager for the Ising model in two dimensions, so the mean field theory is giving us wrong predictions. As we have already stated (but this will be treated in much more detail later on) this is because mean field theories are good approximations only if the system has a high enough dimensionality (and $d = 2$ is still too low for the Ising model, see 6.6.3).

6.2   Hubbard-Stratonovich mean field theory for the Ising model

We have already encountered (see 5.7.3) the Hubbard-Stratonovich identity (5.17) with the saddle point approximation to compute the partition function of a one-dimensional Ising model with long-range interactions. The same technique can also be used to formulate mean field theories for systems with short-ranged interactions, and it is one of the most useful ones.

We therefore start from the most general Hamiltonian of a two-dimensional Ising model with nearest-neighbour interactions:

$$\mathcal{H} = -\frac{1}{2} \sum_{\langle ij \rangle} J_{ij} S_i S_j - \sum_i H_i S_i$$

whose partition function is:

$$Z_N = \mathrm{Tr}\, e^{\beta \left(\frac{1}{2} \sum_{\langle ij \rangle} J_{ij} S_i S_j + \sum_i H_i S_i \right)}$$

The idea of Hubbard-Stratonovich mean field theory is the same as in 5.7.3, namely substituting the quadratic term $S_i S_j$ in the Hamiltonian with a linear one, introducing some auxiliary field over which we integrate.
In order to do so we must generalize the Hubbard-Stratonovich identity (5.17).

Theorem. *Let $\boldsymbol{A}$ be a real symmetric matrix and $\vec{b}$ an arbitrary vector. Then:*

$$\frac{1}{\sqrt{\det \boldsymbol{A}}} e^{\frac{1}{2} b_i (\boldsymbol{A}^{-1})_{ij} b_j} = \int_{-\infty}^{+\infty} \prod_{i=1}^{N} \frac{e^{-\frac{1}{2} x_i \boldsymbol{A}_{ij} x_j + x_i b_i}}{\sqrt{2\pi}} dx_i$$

where for the sake of simplicity we have used Einstein's notation on repeated indices.

We omit the proof of this result.

Therefore, defining:

$$b_i := S_i \qquad\qquad (\boldsymbol{A}^{-1})_{ij} := J_{ij} \qquad\qquad x_i := \varphi_i$$

we get:

$$e^{\frac{1}{2} J_{ij} S_i S_j} = \frac{1}{\sqrt{\det J}} \int_{-\infty}^{+\infty} \prod_{i=1}^{N} \frac{e^{-\frac{1}{2} \varphi_i (J^{-1})_{ij} \varphi_j + \varphi_i S_i}}{\sqrt{2\pi}} d\varphi_i$$

and the partition function becomes:

$$Z_N = \frac{1}{\sqrt{(2\pi)^N \det J}} \int_{-\infty}^{+\infty} \prod_{i=1}^{N} e^{-\frac{\beta}{2} \varphi_i (J^{-1})_{ij} \varphi_j} \operatorname{Tr}_{\{S_i\}} e^{\beta(\varphi_i + H_i) S_i} d\varphi_i$$

where we have explicitly written that the trace is performed over the spin variables S_i.
Now that the terms in the exponential are linear in S_i the partition function is easier to compute; let us change variable defining:

$$\psi_i = \varphi_i + H_i$$

so that:

$$Z_N = \frac{1}{\sqrt{(2\pi)^N \det J}} \int_{-\infty}^{+\infty} \prod_{i=1}^{N} e^{-\frac{\beta}{2} (\psi_i - H_i)(J^{-1})_{ij}(\psi_j - H_j)} \operatorname{Tr}_{\{S_i\}} e^{\beta \psi_i S_i} d\psi_i$$

However:

$$\operatorname{Tr}_{\{S_i\}} e^{\beta \psi_i S_i} = \sum_{\{S_i = \pm 1\}} e^{\beta \psi_i S_i} = \prod_i \sum_{S = \pm 1} e^{\beta \psi_i S} = \prod_i 2\cosh(\beta \psi_i)$$

and since $x = e^{\ln x}$ we have:

$$\operatorname{Tr}_{\{S_i\}} e^{\beta \psi_i S_i} = e^{\sum_i \ln[2\cosh(\beta \psi_i)]}$$

Therefore:

$$Z_N = \frac{1}{\sqrt{(2\pi)^N \det J^{-1}}} \int_{-\infty}^{+\infty} \prod_{i=1}^{N} e^{-\beta \mathcal{L}(\psi_i, H_i, J_{ij})} d\psi_i$$

where:

$$\mathcal{L}(\psi_i, H_i, J_{ij}) = \frac{1}{2} \sum_{i,j} (\psi_i - H_i)(J^{-1})_{ij}(\psi_j - H_j) - \frac{1}{\beta} \sum_i \ln\left[2\cosh(\beta \psi_i)\right]$$

Let us note that there is no trace left of the original degrees of freedom S_i in the partition function, and that it has the form of a *functional integral*, since in the continuum limit ($N \to \infty$ and the distance between lattice sites tending to zero) the auxiliary fields become functions of the position, $\psi(\vec{r})$.
Let us also note that until now we still have not done any approximation (we have just rewritten the partition function); this comes into play right now: since the partition function has the form of a functional integral we must find some approximate ways to compute it. The simplest of all possible approximations is the *saddle point approximation* (see appendix B), which essentially

consists in approximating the integral with the maximum value of the integrand. In other words we approximate:

$$\int_{-\infty}^{+\infty} \prod_{i=1}^{N} e^{-\beta\mathcal{L}(\psi_i, H_i, J_{ij})} d\psi_i \approx e^{-\beta\mathcal{L}(\overline{\psi}_i, H_i, J_{ij})} \qquad \Rightarrow \qquad Z_N \approx \frac{1}{\sqrt{(2\pi)^N \det J}} e^{-\beta\mathcal{L}(\overline{\psi}_i, H_i, J_{ij})}$$

where the stationary solutions are those which satisfy:

$$\frac{\delta\mathcal{L}}{\delta\psi_k}\bigg|_{\overline{\psi}_k} = 0 \qquad \forall k$$

In this case we will have:

$$\sum_j (J^{-1})_{kj}(\overline{\psi}_j - H_j) - \tanh(\beta\overline{\psi}_k) = 0 \qquad \forall k \tag{6.8}$$

multiplying both sides by J_{ik} and summing over k, taking advantage of the fact that $J_{ij} = J_{ji}$ and in the end renaming k with j we get:

$$\overline{\psi}_i = H_i + \sum_j J_{ij}\tanh(\beta\overline{\psi}_j) \qquad \forall i \tag{6.9}$$

Let us see the relation between the auxiliary fields $\overline{\psi}_i$ and the order parameter of the system m_i. We know that:

$$m_k = -\frac{\partial F}{\partial H_k}$$

and from the saddle point approximation we have:

$$e^{-\beta F} = Z \approx e^{-\beta\mathcal{L}(\overline{\psi}_i, H_i, J_{ij})} \qquad \Rightarrow \qquad F = \mathcal{L}(\overline{\psi}_i, H_i, J_{ij})$$

Therefore:

$$m_k = -\frac{\partial F}{\partial H_k} = -\frac{\partial}{\partial H_k}\mathcal{L}(\overline{\psi}_i, H_i, J_{ij}) =$$

$$= -\frac{\partial}{\partial H_k}\left[\frac{1}{2}\sum_{i,j}(\overline{\psi}_i - H_i)(J^{-1})_{ij}(\overline{\psi}_j - H_j) - \frac{1}{\beta}\sum_i \ln\left(2\cosh(\beta\overline{\psi}_i)\right)\right] =$$

$$= \sum_j (J^{-1})_{kj}(\overline{\psi}_j - H_j) = \tanh(\beta\overline{\psi}_k)$$

where we have used (6.8). Thus:

$$\overline{\psi}_i = k_B T \tanh^{-1} m_i \qquad \forall i \tag{6.10}$$

and plugging this into (6.9):

$$k_B T \tanh^{-1} m_i = H_i + \sum_{i,j} J_{ij} m_j \qquad \Rightarrow \qquad m_i = \tanh\left[\beta\left(H_i + \sum_{i,j} J_{ij} m_j\right)\right]$$

We can note that this is a more general form of the self-consistency equation (6.4) for the magnetization that we have found in Weiss mean field theory. In fact, if we set $H_i = H = \text{const.}$ (so that also $m_i = m = \text{const.}$) and choose a nearest-neighbour interaction (i.e. J_{ij} is equal to a constant J if the i-th and j-th sites are nearest neighbours, otherwise is null), calling z the coordination number of the lattice we get exactly (6.4).

This means that in this approximation we deduce the same conclusions we have seen in 6.1 that come from this self-consistency equation; in particular we will have that the temperature of the phase transition when $H = 0$ is again $T_c = zJ/k_B$ and that the critical exponents δ and γ are, respectively, 3 and 1.

We can thus already see that all the mean field theories are equivalent when applied to the same system (which is something reasonable since the approximation we make is always the same). In the following sections we will see many other mean field theories, applied also to fluids (see 6.4), and we will always find the same values for the critical exponents. This is another expression of the fact that different systems, like an Ising model or a classical fluid, belong to the same universality class (even more, as we have shown in 5.3.4 they are actually equivalent) and thus behave similarly near a critical point.

6.3 Variational methods

Variational methods in statistical mechanics are very important since they provide a tool to formulate mean field theories which are valid for any temperature range and with order parameters of essentially arbitrary complexity.

Their central idea is what one would expect: if $\mathcal{H}$ is the Hamiltonian of a physical system and ψ_α is a set of arbitrary trial states, then we can obtain the energy of the ground state of the system by minimizing $\langle \mathcal{H} \rangle_{\psi_\alpha}$ with respect to ψ_α; since every ψ_α will be in general a function or an even more complex object, $\langle \mathcal{H} \rangle_{\psi_\alpha}$ in general is a functional so its minimization must be intended in the sense of functional analysis.

We will see that however the only mean value of the Hamiltonian won't be sufficient since we know that the equilibrium configurations of the system are given by the minima of the free energy. In other words, we will compute the free energy using some trial states ψ_α and then minimize it in the ways we will explain.

Such variational methods are also used in quantum mechanics when a system is too complex and its Schrödinger equation can't be solved exactly: in this case one introduces a set $|\psi_\alpha\rangle$ of trial wave functions and minimizes the functional $E_\alpha = \langle \psi_\alpha | \mathcal{H} | \psi_\alpha \rangle$ with respect to $|\psi_\alpha\rangle$, so that both the wave functions and the energy of the ground state of the system can be found.

In statistical mechanics variational methods are performed using the phase space equilibrium probability density of the system.

In particular, the approach of variational methods in statistical mechanics is based upon two inequalities which we now show.

Theorem. *Let φ be a random variable (it can be either discrete or continuous), and call its probability density ρ; clearly, for any function f of φ the mean value of f is defined as:*

$$\langle f(\varphi) \rangle_\rho := \mathrm{Tr}(\rho(\varphi) f(\varphi))$$

1. *If f is the exponential function then the following inequality holds:*

$$\left\langle e^{-\lambda \varphi} \right\rangle_\rho \geq e^{-\lambda \langle \varphi \rangle_\rho} \quad \forall \rho, \quad \forall \lambda \in \mathbb{R}$$

2. *If $\mathcal{H}(\varphi)$ is the Hamiltonian of a system and F its free energy, then:*

$$F \leq \mathrm{Tr}(\rho \mathcal{H}) + k_B T \, \mathrm{Tr}(\rho \ln \rho) \qquad \forall \rho$$

Proof.

1. Supposing φ a real number, from the Taylor expansion of the exponential we have $e^{-\varphi} \geq 1 - \varphi$ and so (we omit the subscript ρ on mean values for simplicity):

$$e^{-\lambda\varphi} = e^{-\lambda(\varphi + \langle\varphi\rangle - \langle\varphi\rangle)} = e^{-\lambda\langle\varphi\rangle}e^{-\lambda(\varphi - \langle\varphi\rangle)} \geq e^{-\lambda\langle\varphi\rangle}\left[1 - \lambda(\varphi - \langle\varphi\rangle)\right]$$

and therefore taking the mean value of both sides:

$$\left\langle e^{-\lambda\varphi}\right\rangle \geq \left\langle e^{-\lambda\langle\varphi\rangle}\right\rangle - \lambda e^{-\lambda\langle\varphi\rangle}\left\langle(\varphi - \langle\varphi\rangle)\right\rangle = e^{-\lambda\langle\varphi\rangle}$$

2. The canonical partition function of the system can be written as:

$$Z = \operatorname{Tr} e^{-\beta\mathcal{H}} = \operatorname{Tr}\left(\rho e^{-\beta\mathcal{H} - \ln\rho}\right) = \left\langle e^{-\beta\mathcal{H} - \ln\rho}\right\rangle \geq e^{-\beta\langle\mathcal{H}\rangle - \langle\ln\rho\rangle}$$

where the last step comes from the first inequality. Since $Z = e^{-\beta F}$ with F the free energy of the system, we have:

$$e^{-\beta F} \geq e^{-\beta\langle\mathcal{H}\rangle - \langle\ln\rho\rangle}$$

and taking the logarithm:

$$F \leq \langle\mathcal{H}\rangle + k_B T \langle\ln\rho\rangle = \operatorname{Tr}(\rho\mathcal{H}) + k_B T \operatorname{Tr}(\rho\ln\rho)$$

□

Remember also that since ρ is a probability distribution it must satisfy:

$$\rho(\varphi) \geq 0 \qquad\qquad \operatorname{Tr}\rho = 1$$

Now, the free energy F of the system is a functional of the probability density ρ and from what we have just seen we can set an upper bound to F:

$$F \leq \langle\mathcal{H}\rangle + k_B T \langle\ln\rho\rangle = \operatorname{Tr}(\rho\mathcal{H}) + k_B T \operatorname{Tr}(\rho\ln\rho) := F_{\text{m.f.}}$$

where "m.f." stands for "mean field". In other words we can estimate the free energy of the system in mean field theories as $\langle\mathcal{H}\rangle + k_B T \langle\ln\rho\rangle$, and the best approximation of the *real* free energy will be given by the minimization of $F_{\text{m.f.}}$. In particular, the ground state configuration of the system will be given by the form of ρ that minimizes F, which can be easily determined in general:

$$\frac{\delta F}{\delta\rho}_{|\rho_{\text{eq}}} = 0 \qquad \Rightarrow \qquad \rho_{\text{eq}} = \frac{A}{e}e^{-\beta\mathcal{H}}$$

with A a generic constant, and since ρ_{eq} must be subjected to the constraint $\operatorname{Tr}\rho_{\text{eq}} = 1$ we find $A = e/Z$, so that:

$$\rho_{\text{eq}} = \frac{1}{Z}e^{-\beta\mathcal{H}}$$

This is extremely reasonable: since we have only required the minimization of the free energy, the probability density we obtain is the one we would expect from ensemble theory.

However, until now the computation is still exact: if we want to determine ρ_{eq} we must compute Z, which in general is not feasible.

Within this variational approach the mean field approximation comes into play by choosing the following form of the trial probability density:

$$\rho_{\text{m.f.}} = \prod_{\alpha}\rho_\alpha$$

where again "m.f." stands for "mean field", α labels the degrees of freedom of the system and ρ_α is the probability distribution of the *sole* α-th degree of freedom. In other words we are approximating the probability distribution so that the degrees of freedom are *statistically independent*[5], namely:

$$\langle f_1(\varphi_1) f_2(\varphi_2) \rangle_{\rho_{\text{m.f.}}} = \langle f_1(\varphi_1) \rangle_{\rho_{\text{m.f.}}} \langle f_2(\varphi_2) \rangle_{\rho_{\text{m.f.}}} \quad \forall f_1, f_2$$

This way the free energy of the system has the form:

$$F_{\rho_{\text{m.f.}}} = \langle \mathcal{H} \rangle_{\rho_{\text{m.f.}}} + k_B T \sum_\alpha \langle \ln \rho_\alpha \rangle_{\rho_{\text{m.f.}}} = \langle \mathcal{H} \rangle_{\rho_{\text{m.f.}}} + k_B T \sum_\alpha \text{Tr}(\rho_\alpha \ln \rho_\alpha)$$

and must be minimized with respect to ρ_α.
This can be done with two different approaches:

1. The most used one consists in parametrizing ρ_α with an appropriately defined order parameter $\langle \varphi_\alpha \rangle$ that can describe an eventual phase transition; in this way F becomes a (real) function of $\langle \varphi_\alpha \rangle$, and the minimization becomes simpler since it reduces to minimizing a simple function.
   The parametrization must of course satisfy the constraints:

$$\text{Tr}\,\rho_\alpha = 1 \qquad\qquad \text{Tr}(\rho_\alpha \varphi_\alpha) = \langle \varphi_\alpha \rangle$$

   The advantage of such an approach is that the variational parameter φ_α coincides with the order parameter.

2. Another possible approach consists in considering ρ_α itself as a variational parameter, and minimizing F with respect to it. This is a more general approach, but this time it's harder to establish a connection between F as a functional of ρ and F as a function of the order parameter that describes a phase transition.

We will analyse these two different approaches applying them to two different models.

6.3.1 Bragg-Williams approximation for the Ising model

Again, we consider a system with the following Hamiltonian:

$$\mathcal{H} = -J \sum_{\langle ij \rangle} S_i S_j - \sum_i H_i S_i$$

In this case our random variables φ are the spins: $\varphi_\alpha = S_i$, and our order parameter is $\langle \varphi_\alpha \rangle = m_i$ (we will later see that this is exactly the local magnetization also in the mean field approximation). For what we have previously stated we must build a single-particle probability density $\rho_i = f(m_i)$ in terms of this order parameter, such that $\text{Tr}\,\rho_i = 1$ and $\text{Tr}(\rho_i S_i) = m_i$. For example, since we have two constraints on ρ_i we could use a linear expression for the probability density involving two parameters[6]:

$$\rho_i = a \delta_{S_i,1} + b(1 - \delta_{S_i,1})$$

---

[5]This is physically equivalent to what we have done in the Weiss mean field theory for the Ising model (see 6.1), where we neglected the correlations between spins.

[6]This form of ρ_i is very general, and does not depend on the fact that the degrees of freedom of the system can only assume two values: if there is a different number of possible states, say n, then ρ_i can be written in the same form, but a will be the probability of one state while b the probability of the remaining $n-1$ ones. We will shortly see this when we will apply the Bragg-Williams approximation to the Potts model.

where a and b are, respectively, the probability that $S_i = +1$ and $S_i = -1$. Inserting this expression of ρ_i in the two constraints we get:

$$\operatorname{Tr}\rho_i = a + b = 1 \qquad\qquad \operatorname{Tr}(\rho_i S_i) = \langle S_i\rangle_{\rho_i} = a - b = m_i$$

and therefore:

$$a = \frac{1 + m_i}{2} \qquad\qquad b = \frac{1 - m_i}{2}$$

Thus:

$$\rho_i = \frac{1 + m_i}{2}\delta_{S_i,1} + \frac{1 - m_i}{2}(1 - \delta_{S_i,1})$$

We are now able to compute the two terms that contribute to the free energy $F_{\rho_{\text{m.f.}}} = \langle \mathcal{H}\rangle_{\rho_{\text{m.f.}}} + k_B T \sum_\alpha \langle \ln \rho_\alpha\rangle_{\rho_{\text{m.f.}}}$ of the system.

First of all:

$$\langle \mathcal{H}\rangle_{\rho_{\text{m.f.}}} = -J\sum_{\langle ij\rangle}\langle S_i S_j\rangle_{\rho_{\text{m.f.}}} - \sum_i H_i \langle S_i\rangle_{\rho_{\text{m.f.}}}$$

However, since $\rho_{\text{m.f.}} = \prod_i \rho_i$ (namely the degrees of freedom are independent, as we have already seen), then $\langle S_i S_j\rangle_{\rho_{\text{m.f.}}} = \langle S_i\rangle_{\rho_i}\langle S_j\rangle_{\rho_j}$. Furthermore, if $g(S_i)$ is a generic function of S_i then:

$$\langle g(S_i)\rangle_{\rho_{\text{m.f.}}} = \operatorname{Tr}(g(S_i)\rho_i) = \sum_{S_i = \pm 1} g(S_i)\rho_i =$$

$$= \sum_{S_i = \pm 1} g(S_i)\left(\frac{1 + m_i}{2}\delta_{S_i,1} + \frac{1 - m_i}{2}(1 - \delta_{S_i,1})\right) = \frac{1 + m_i}{2}g(1) + \frac{1 - m_i}{2}g(-1)$$

and so if $g(S_i) = S_i$ we have $\langle S_i\rangle_{\rho_{\text{m.f.}}} = m_i$: our order parameter is (as we expected) the local magnetization of the system.
Therefore:

$$\langle \mathcal{H}\rangle_{\rho_{\text{m.f.}}} = -J\sum_{\langle ij\rangle} m_i m_j - \sum_i H_i m_i$$

Then, for the other term of F we have:

$$\sum_i \langle \ln \rho_i\rangle_{\rho_{\text{m.f.}}} = \sum_i \operatorname{Tr}(\rho_i \ln \rho_i) = \sum_i \left(\frac{1 + m_i}{2}\ln\frac{1 + m_i}{2} + \frac{1 - m_i}{2}\ln\frac{1 - m_i}{2}\right)$$

The total energy of the system will be:

$$F_{\rho_{\text{m.f.}}} = \langle \mathcal{H}\rangle_{\rho_{\text{m.f.}}} + k_B T \sum_i \langle \ln \rho_i\rangle_{\rho_{\text{m.f.}}} =$$

$$= -J\sum_{\langle ij\rangle} m_i m_j - \sum_i H_i m_i + k_B T \sum_i \left(\frac{1 + m_i}{2}\ln\frac{1 + m_i}{2} + \frac{1 - m_i}{2}\ln\frac{1 - m_i}{2}\right) \quad (6.11)$$

We now have to minimize $F_{\rho_{\text{m.f.}}}$ with respect to m_i:

$$0 = \frac{\partial F_{\rho_{\text{m.f.}}}}{\partial m_i} = -2J\sum_{j\in\text{n.n.}(i)} m_j - H_i + k_B T\left(\frac{1}{2}\ln\frac{1 + m_i}{2} + \frac{1}{2} - \frac{1}{2}\ln\frac{1 - m_i}{2} - \frac{1}{2}\right) =$$

$$= -2J\sum_{j\in\text{n.n.}(i)} m_j - H_i + \frac{k_B T}{2}\ln\frac{1 + m_i}{1 - m_i}$$

where $j \in$ n.n.(i) means that the j-th site is a nearest neighbour of the i-th one.
Now, recalling that:

$$\tanh^{-1} x = \frac{1}{2} \ln \frac{1+x}{1-x} \qquad |x| < 1$$

(and surely $|m_i| < 1$) we can write:

$$k_B T \tanh^{-1} m_i = 2J \sum_{j \in \text{n.n.}(i)} m_j + H_i$$

and inverting the hyperbolic tangent:

$$m_i = \tanh \left[\beta \left(2J \sum_{j \in \text{n.n.}(i)} m_j + H_i \right) \right]$$

We have again found the self-consistency equation (6.4) for the magnetization that we have already encountered in the Weiss mean field approximation! This is again a confirmation that all mean field theories are equivalent.

We can thus see that in the Bragg-Williams approximation also the β and α exponents are the same of the Weiss mean field theory; in fact, we have seen in 6.1.1 that they come from the expansion of the free energy density for small values of the magnetization when $H = 0$. If we now set $H = 0$ into (6.11), so that $m_i = m \quad \forall i$, we get:

$$\frac{F_{\rho_{\text{m.f.}}}}{N} = -Jzm^2 + k_B T \left(\frac{1+m}{2} \ln \frac{1+m}{2} + \frac{1-m}{2} \ln \frac{1-m}{2} \right)$$

If we now expand the logarithm for small m, we get:

$$f(m) = \lim_{N \to \infty} \frac{F_{\rho_{\text{m.f.}}}}{N} = -Jzm^2 + k_B T \left(-\ln 2 + \frac{m^2}{2} + \frac{m^4}{12} + \cdots \right) \sim$$
$$\sim -k_B T \ln 2 + \left(\frac{k_B T}{2} - Jz \right) m^2 + \frac{k_B T}{2} m^4$$

Therefore we see that the behaviour of f near $m = 0$ is the same of equation (6.6), and so both the exponents β and α turn out to be equal to what we have determined[7] in 6.1.1.

We can also say something[8] in the case $H \neq 0$. Supposing that our system is uniform, i.e. $m_i = m \quad \forall i$, we can rewrite (6.11) as:

$$\tilde{f} := \frac{F_{\rho_{\text{m.f.}}}}{N} = -Hm + f_0$$

where:

$$f_0 = -Jzm^2 + k_B T \left(\frac{1+m}{2} \ln \frac{1+m}{2} + \frac{1-m}{2} \ln \frac{1-m}{2} \right)$$

and since we are looking for the absolute minimum of $\tilde{f}$:

$$0 = \frac{\partial \tilde{f}}{\partial m} = -H - f_0' = -H - 2Jzm + \frac{k_B T}{2} \ln \frac{1+m}{1-m}$$

---

[7]Note that also the temperature of the transition is still the same, considering also the $1/2$ factor we have already mentioned.

[8]These considerations apply in general also in the other mean field theories considered, but we show them now.

(a) $T > T_c$

(b) $T < T_c$
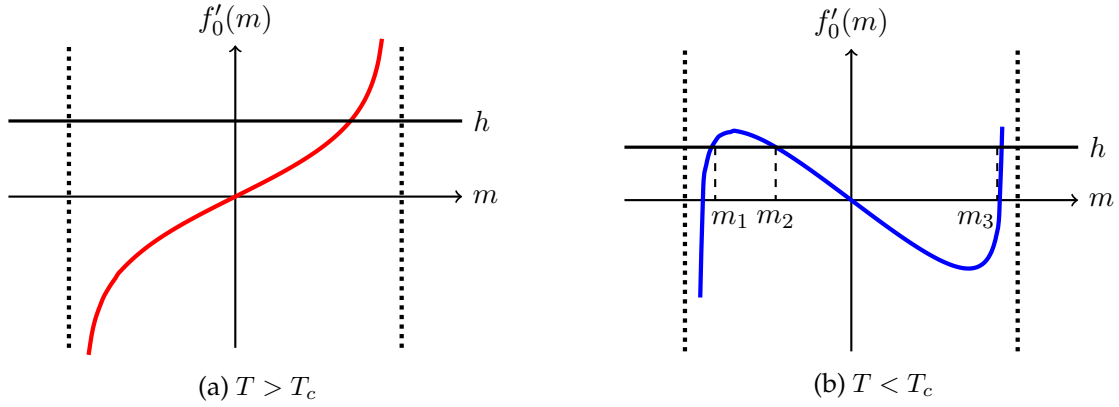
Figure 6.2: Solutions of (6.12)

which gives the self-consistency equation we already know:

$$h = -2\beta J z m + \tanh^{-1} m \tag{6.12}$$

As we can see from the plots in figure 6.2 the number of the possible solutions depends on the temperature of the system. In particular if $2\beta J z < 1$ (i.e. $T > T_c$), expanding the right hand side of (6.12) we get a positive linear term ($2\beta J z m > 0$), so that the behaviour of $-\beta J z m + \tanh^{-1} m$ is as shown in figure 6.2a, and (6.12) has only one solution.

On the other hand if $T < T_c$ then the linear term changes sign and $-\beta J z m + \tanh^{-1} m$ behaves as in figure 6.2b: in this case if $|h|$ is small enough there are three possible solutions, which we have called m_1, m_2 and m_3. These are all extrema of $\tilde{f}$, but how can we understand which is a minimum or a maximum? And above all, which of them is the global minimum?

If we suppose $h > 0$ to be large there will be only one solution, m_3, and as h decreases also m_1 and m_2 will appear; we can therefore argue that for the continuity of $\tilde{f}$ the solution m_3 is still a minimum also when m_1 and m_2 are present. Similarly, if we take $h < 0$ we can conclude that also m_1 is a minimum; therefore m_2 will necessarily be a maximum.

Now, in order to see which between m_1 and m_3 is the global minimum of $\tilde{f}$ let us take $h > 0$ as in figure 6.2b and compute:

$$\tilde{f}(m_3) - \tilde{f}(m_1) = \int_{m_1}^{m_3} \tilde{f}'(m)dm = \int_{m_1}^{m_3} (f_0' - h)dm$$

From figure 6.2b we see that this is equal to the area enclosed by the the straight line $h = \text{const.}$ and the graph of f_0', which is clearly negative if $h > 0$. Therefore $\tilde{f}(m_3) < \tilde{f}(m_1)$ and so we conclude that m_3 is always the global minimum of $\tilde{f}$.

Similarly, when $h < 0$ the global minimum of $\tilde{f}$ is m_1.

This means that as soon as h changes sign the global minimum of $\tilde{f}$ changes abruptly from m_3 to m_1. We are thus obtaining the phenomenology that is indeed observed for a magnet when we change the external field H. In other words the sets of points $(h(m), m)$ are exactly the graphs represented in figure 1.3b, i.e. the graphs of the magnetization seen as a function of the external field.

6.3.2   Bragg-Williams approximation for the Potts model

We now apply the same approximation that we have just seen to a slightly more complex situation: the *Potts model*.

This is defined exactly as the Ising model, but with an essential difference: the degrees of freedom of the system, which we now call σ_i, instead of just two values can assume q integer values: $\sigma_i \in \{1, \ldots, q\}$.

We can therefore write the Hamiltonian of such a system as:

$$\mathcal{H} = -J \sum_{\langle ij \rangle} \delta_{\sigma_i, \sigma_j} - H \sum_i \delta_{\sigma_i, 1}$$

(where we have supposed that the external magnetic field tends to favour the situation where the degrees of freedom assume the value 1; of course we could have done otherwise).

As can be expected the Potts model with $q = 2$ is equivalent to an Ising model, as can be seen from the following equivalence:

$$\delta_{\sigma, \sigma'} = \frac{1}{2}(1 + \sigma\sigma')$$

where σ and σ' are the degrees of freedom of the Ising model ($\sigma, \sigma' = \pm 1$).

However, a Potts model with $q = 3$ is *not* equivalent to an Ising model where $S_i = -1, 0, +1$, as one could have expected. In fact it is not possible in this case to write a δ of the three-state variable σ with "simple" terms involving σ-s (namely, only quadratic terms); in particular it turns out that:

$$\delta_{\sigma, \sigma'} = 1 + \frac{1}{2}\sigma\sigma' + \frac{3}{2}\sigma^2\sigma'^2 - (\sigma^2 + \sigma'^2)$$

We therefore want to apply the Bragg-Williams approximation to a q-state Potts model.

First of all, we call $x = \langle \delta_{\sigma, 1} \rangle$ our order parameter, and write the probability distribution of a single degree of freedom as[9]:

$$\rho_\sigma = a\delta_{\sigma, 1} + b(1 - \delta_{\sigma, 1})$$

Therefore, from:

$$\text{Tr}\, \rho_\sigma = a + (q-1)b = 1 \qquad\qquad \text{Tr}(\rho_\sigma \delta_{\sigma, 1}) = \sum_{\sigma=1}^{q} \left[\delta_{\sigma, 1} \cdot a\delta_{\sigma, 1} + \delta_{\sigma, 1} \cdot b(1 - \delta_{\sigma, 1}) \right] = x$$

we get:

$$a = x \qquad\qquad b = \frac{1 - x}{q - 1}$$

From now on one can proceed like we have previously seen.

6.3.3 Mean field theory for the Blume-Emery-Griffiths model

The Blume-Emery-Griffiths model (often shortened in "BEG" model) is a lattice gas model used to describe the properties of the *superfluid* transition of ^{4}He, which when it is cooled under approximately $T_c = 2.7$ K undergoes a continuous phase transition[10] from fluid to superfluid; superfluids exhibit several interesting properties, like having zero viscosity. The BEG model is used to describe what happens when we add some ^{3}He to the system; it does not consider quantum effects, but only the "messing up" due to the ^{3}He impurities.

Experimentally when ^{3}He is added to ^{4}He the temperature of the fluid-superfluid transition decreases. For small concentration of ^{3}He the mixture remains homogeneous, and the only effect is the change of T_c; however, when the concentration x of ^{3}He reaches the critical value $\overline{x} \approx 0.67$, ^{3}He and ^{4}He separate into two phases (just like oil separates from water) and the λ

---

[9]As we have stated in the footnote on page 139, we write ρ in general as the sum of the probability that the degree of freedom assume a particular value and of that of all the remaining values; in this case we have chosen $\sigma = 1$ as this particular value, but of course we could have done otherwise.

[10]This is generally called λ *transition*, because the plot of the specific heat as a function of the temperature has a shape that resembles a λ.

transition becomes first-order (namely, discontinuous). The transition point (x_t, T_t) where the system shifts from a continuous λ transition to a discontinuous one is that where the phase separation starts and is called *tricritical point*.
The BEG model was introduced to describe such a situation.

As we have anticipated, it is a lattice gas model and so it is based on an Ising-like Hamiltonian (see 5.3.4). On the sites of this lattice we define a variable S_i which can assume the values -1, 0 and $+1$: we decide that when an ^{4}He atom is present in a lattice site then $S_i = \pm 1$, while when $S_i = 0$ it means that the site is occupied by an ^{3}He atom. We then define our order parameter to be $m = \langle S_i \rangle$; in the Ising model $\langle S_i^2 \rangle$ can only be equal to 1, while in this case it can be either 0 or 1: we can thus interpret $\langle S_i^2 \rangle$ as the concentration of ^{4}He atoms, and $x := 1 - \langle S_i^2 \rangle$ as the fraction of ^{3}He. We also define $\Delta := \mu_{^3\text{He}} - \mu_{^4\text{He}}$ to be the difference of the chemical potentials of ^{3}He and ^{4}He; since this parameter is related to the number of ^{3}He and ^{4}He atoms, we expect that when $x \to 0$ (namely, there is only ^{4}He) $\Delta \to -\infty$, while if $x \to 1$ then $\Delta \to +\infty$.

We consider the following Hamiltonian for the system[11]:

$$\mathcal{H} = -J \sum_{\langle ij \rangle} S_i S_j + \Delta \sum_i S_i^2 - \Delta N$$

where N is the total number of lattice sites.
Since we want to apply the second variational method that we have seen, we write the mean field probability density as:

$$\rho_{\text{m.f.}} = \prod_i \rho_i = \prod_i \rho(S_i)$$

and the free energy:

$$F_{\rho_{\text{m.f.}}} = \langle \mathcal{H} \rangle_{\rho_{\text{m.f.}}} + k_B T \sum_i \text{Tr}(\rho_i \ln \rho_i)$$

The mean value of the Hamiltonian is:

$$\langle \mathcal{H} \rangle_{\rho_{\text{m.f.}}} = -J \sum_{\langle ij \rangle} \langle S_i S_j \rangle_{\rho_{\text{m.f.}}} + \Delta \sum_i \langle S_i^2 \rangle_{\rho_{\text{m.f.}}} - N\Delta$$

and since $\langle S_i S_j \rangle_{\rho_{\text{m.f.}}} = \langle S_i \rangle_{\rho_{\text{m.f.}}} \langle S_j \rangle_{\rho_{\text{m.f.}}}$ (it's the fundamental hypothesis of mean field theories) we get[12]:

$$\langle \mathcal{H} \rangle_{\rho_{\text{m.f.}}} = -\frac{1}{2} JNz \langle S_i \rangle_{\rho_{\text{m.f.}}}^2 + N\Delta \langle S_i^2 \rangle_{\rho_{\text{m.f.}}} - N\Delta$$

where z is the coordination number of the lattice. Therefore, the free energy of the system is:

$$F_{\rho_{\text{m.f.}}} = -\frac{1}{2} JNz \left(\text{Tr}(\rho_i S_i) \right)^2 + N\Delta \, \text{Tr}(\rho_i S_i^2) - N\Delta + k_B T N \, \text{Tr}(\rho_i \ln \rho_i)$$

We now must minimize this expression with respect to ρ_i, with the constraint $\text{Tr}\,\rho_i = 1$. Since we have:

$$\frac{\delta}{\delta \rho_j} \left(\text{Tr}(\rho_i S_i) \right)^2 = 2 \, \text{Tr}(\rho_i S_i) \cdot S_j = 2 \langle S_i \rangle S_j = 2m S_j$$

$$\frac{\delta}{\delta \rho_j} \text{Tr}(\rho_i S_i^2) = S_j^2 \qquad\qquad \frac{\delta}{\delta \rho_j} \text{Tr}(\rho_i \ln \rho_i) = \ln \rho_j + 1$$

---

[11]We do not justify why it has this form; furthermore, the Hamiltonian we are considering is in reality a simplification of the original one.

[12]We are introducing the 1/2 factor for later convenience. As we have stated in the footnote on page 98, this is perfectly possible if we change the convention of the sum on nearest neighbours.

then:

$$\frac{\delta}{\delta \rho_i} F_{\rho_{\text{m.f.}}} = -JNzmS_i + N\Delta S_i^2 + Nk_BT \ln \rho_i + Nk_BT = 0$$

which leads to:

$$\ln \rho_i = \beta JzmS_i - \beta \Delta S_i^2 - 1 \qquad \Rightarrow \qquad \rho(S_i) = \frac{1}{A} e^{\beta(zJmS_i - \Delta S_i^2)}$$

where we have reabsorbed e^{-1} into the normalization constant A. From the constraint $\text{Tr}\,\rho_i = 1$ we find:

$$A = 1 + 2e^{-\beta\Delta}\cosh(\beta zJm)$$

Substituting this expression of ρ_i into $F_{\rho_{\text{m.f.}}}$, after some mathematical rearrangement we get:

$$\frac{F_{\rho_{\text{m.f.}}}}{N} = \frac{z}{2}Jm^2 - k_BT \ln\left[1 + 2e^{-\beta\Delta}\cosh(\beta zJm)\right] - \Delta$$

In order to find the equilibrium state for any T and Δ, we must minimize this expression with respect to m.

If we expand $F_{\rho_{\text{m.f.}}}$ for small m, keeping in mind the Taylor expansions $\cosh x \sim x + x^2/2 + O(x^4)$ and $\ln(1+x) \sim 1 + x - x^2/2 + O(x^3)$ we get:

$$\frac{F_{\rho_{\text{m.f.}}}}{N} \overset{m \approx 0}{\sim} a + \frac{b}{2}m^2 + \frac{c}{4}m^4 + \frac{d}{6}m^6 + \cdots$$

where:

$$a = -k_BT \ln(1 + 2e^{-\beta\Delta}) - \Delta \qquad b = zJ\left(1 - \frac{zJ}{\delta k_BT}\right) \qquad c = \frac{zJ}{2\delta^2}(\beta zJ)^3\left(1 - \frac{\delta}{3}\right)$$

and $\delta = 1 + e^{\beta\Delta}/2$; d turns out to be always positive. Note that unlike the Ising model in the Weiss approximation (see 6.1) in this case both the quadratic and the quartic terms, b and c, can change sign when the parameters assume particular values.

Let us also note that the order parameter of the system, namely the concentration of ^{3}He, is:

$$x = 1 - \langle S_i^2 \rangle = \frac{1}{1 + 2e^{-\beta\Delta}\cosh(\beta zJm)}$$

Therefore, in the disordered phase (both ^{3}He and ^{4}He are present) we have $m = 0$ and the concentration of ^{3}He becomes:

$$x = \frac{1}{1 + 2e^{-\beta\Delta}} = 1 - \frac{1}{\delta}$$

This way we can determine how the temperature of the λ transition depends on x; in fact, the critical temperature will be the one that makes b change sign, so we can determine it from the condition $b = 0$:

$$T_c = \frac{zJ}{k_B\delta}$$

Since as we have just seen $1/\delta = 1 - x$, we have:

$$T_c(x) = T_c(0)(1 - x)$$

where $T_c(0) = zJ/k_B$. The other transition (from the continuous λ to the discontinuous one) will occur when the quartic term c changes sign, and so we can determine the critical value of x at which it occurs from the condition $c = 0$:

$$1 - \frac{\delta_t}{3} = 0 \qquad \Rightarrow \qquad \delta_t = 3 \qquad \Rightarrow \qquad x_t = 1 - \frac{1}{3} = \frac{2}{3} = 0.\overline{6}$$

which is in astonishingly good agreement with the experimental result of $x_t \approx 0.67$.

6.4   Mean field theories for fluids

Ideal gases are exceedingly idealised systems and are not suited to describe the behaviour of real systems: they always obey the same state equation and never undergo phase transitions (for example they never condense).

We must therefore step a little further: using the "philosophy" of mean field theories we can make the description of fluids a little bit more realistic. As we will see this will also lead to the derivation of the *Van der Waals equation*, which better describes the behaviour of real fluids (even if, as we will shortly see, it still has some problems).

In general, in a real gas all the atoms or molecules interact through a certain potential $\Phi(\{\vec{r}_i\})$ that will depend on the positions of all the particles. For a system of N particles the configurational contribution to the partition function will therefore be:

$$Q_N = \int \prod_{i=1}^{N} d\vec{r}_i e^{-\beta\left(\sum_{i=1}^{N} \varphi_{\text{ext}}(\{\vec{r}_i\}) + \Phi(\{\vec{r}_i\})\right)}$$

where φ_{ext} is an external potential and in general:

$$\Phi(\{\vec{r}_i\}) = \sum_{i \neq j} U_2(\vec{r}_i, \vec{r}_j) + \sum_{i \neq j \neq k} U_3(\vec{r}_i, \vec{r}_j, \vec{r}_k) + \cdots$$

(where U_n can be a generic n-body interaction potential).

Generally φ_{ext} does not pose great problems while it is Φ that makes Q_N impossible to compute exactly, forcing us to resort to approximations. In the framework of mean field theories we substitute the interaction potential Φ with an *effective single-particle potential* $\varphi(\vec{r}_i)$ that acts on every particle in the same way: $\Phi(\{\vec{r}_i\}) \approx \sum_i \varphi(\vec{r}_i)$.

Therefore, neglecting the external term for the sake of simplicity, mean field theories allow us to compute Q_N as:

$$Q_N = \left(\int d\vec{r} e^{-\beta\varphi(\vec{r})} \right)^N$$

Of course, every particular mean field theory will provide a different form of φ, which will lead to different results.

6.4.1   Van der Waals equation

The Van der Waals equation can be obtained considering the atoms of a gas as hard spheres. In this case, in fact, the mean field φ has the form:

$$\varphi(\vec{r}) = \begin{cases} +\infty & |\vec{r}| < r_0 \\ u & |\vec{r}| \geq r_0 \end{cases} \tag{6.13}$$

with $u < 0$ a parameter that in general can depend on N and V. In other words the particles are modelled as spheres of radius r_0: for large distances the potential is attractive, while for short distances it is repulsive and prevents the atoms from overlapping. Note also that this is a very rough first approximation: the potential, when attractive, does not depend on the distance between the atoms, and φ is discontinuous. We will later see what happens when we make a more realistic approximation.

Therefore, we will have:

$$e^{-\beta\varphi(\vec{r})} = \begin{cases} 0 & |\vec{r}| < r_0 \\ e^{-\beta u} & |\vec{r}| \geq r_0 \end{cases}$$
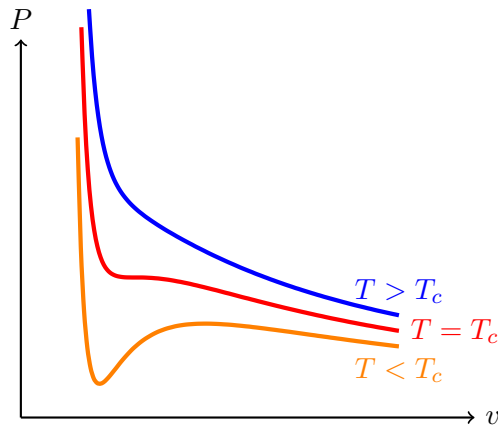
Figure 6.3: Van der Waals isotherms for $a = b = 1$

and calling V_{exc} the volume that must be excluded because the particles are hard spheres (namely the volume occupied by all the other atoms) the configurational partition function will be:

$$Q_N = \left[(V - V_{\text{exc}})e^{-\beta u} \right]^N$$

The configurational part of the free energy is thus:

$$F_N^{\text{conf.}} = -Nk_BT \left[\ln(V - V_{\text{exc}}) - \beta u \right]$$

We can determine the state equation of this system as usual (we derive only the configurational part of F_N because the kinetic one does not contain terms depending on V):

$$P = -\frac{\partial F_N^{\text{conf.}}}{\partial V} = \frac{Nk_BT}{V - V_{\text{exc}}} - N\frac{\partial u}{\partial V}$$

Now, since u is the attractive term between the particles we can suppose that it is proportional to the particle density (intuitively, if the system is dense the particles will be closer to each other and will interact strongly), and of course the excluded volume V_{exc} will be proportional to the number of particles; therefore, setting:

$$u = -a\frac{N}{V} \qquad\qquad V_{\text{exc}} = bN \qquad\qquad a, b > 0$$

we have:

$$P = \frac{Nk_BT}{V - Nb} - a\left(\frac{N}{V}\right)^2$$

which is exactly Van der Waals equation. Defining $v = V/N$ we can rewrite it as:

$$P = \frac{k_BT}{v - b} - \frac{a}{v^2} \tag{6.14}$$

Critical point of Van der Waals equation

The behaviour of the Van der Waals isotherms is shown in figure 6.3. As we can see this changes with the temperature and resembles that of real isotherms (see figure 1.2b); however, Van der Waals isotherms are always analytic and have a non physical behaviour in certain regions of (v, P) plane, called *spinodal curves*, if $T < T_c$: for some values of v we have $\partial P/\partial v > 0$, which is physically impossible. This is a consequence of the roughness of the approximation we have

made, since it can be shown that it doesn't ensure that the equilibrium state of the system glob-
ally minimizes the Gibbs free energy. As we will shortly see, however, this problem can be
solved "by hand" with *Maxwell's equal area rule*.

Let us now see how to determine the critical point of a system obeying Van der Waals equation.
First of all, from the isotherms represented in figure 6.3 we can see that the critical point is a
flex for the critical isotherm (i.e. the one with $T = T_c$); in other words, we can determine the
critical point from the equations:

$$\frac{\partial P}{\partial v} = 0 \qquad\qquad \frac{\partial^2 P}{\partial v^2} = 0$$

Equivalently, we can note that the equation $P(v) = P = \text{const.}$ is cubic in v. In fact, we can
rewrite the Van der Waals equation (6.14) as:

$$v^3 - \left(b + \frac{k_B T}{P}\right) v^2 + \frac{a}{P} v - \frac{ab}{P} = 0 \tag{6.15}$$

For $T > T_c$ equation (6.15) has one real solution and two imaginary ones, and for $T < T_c$ three
distinct real solutions; when $T = T_c$ the three solutions of the equation coincide. This means
that at the critical point equation (6.15) must be written in the form:

$$(v - v_c)^3 = v^3 - 3v_c v^2 + 3v_c^2 v - v_c^3 = 0$$

Equating the coefficients with the ones in equation (6.15) we get:

$$3v_c = b + \frac{k_B T_c}{P_c} \qquad\qquad 3v_c^2 = \frac{a}{P_c} \qquad\qquad v_c^3 = \frac{ab}{P_c}$$

from which we have:

$$v_c = 3b \qquad\qquad P_c = \frac{a}{27b^2} \qquad\qquad T_c = \frac{1}{k_B}\frac{8a}{27b}$$

We have found a very interesting result: in fact, if we can measure a and b at high temperatures
then we are able to determine the critical point of the system.

This model has also an interesting property, since it predicts that:

$$\frac{P_c v_c}{k_B T_c} = \frac{3}{8} = 0.375$$

which is a *universal number*, independent of a and b and so of the particular fluid considered.
Experimentally this ratio is approximately 0.29 for Argon, 0.23 for water and 0.31 for ^{4}He.
Therefore, even if it is very rough, this model leads to reasonable conclusions.

Law of corresponding states

We can also rewrite Van der Waals equation (6.14) in an adimensional form, rescaling the ther-
modynamic quantities of the system. In particular, defining:

$$\pi := \frac{P}{P_c} = P\frac{27b^2}{a} \qquad\qquad \nu := \frac{v}{v_c} = \frac{v}{3b} \qquad\qquad \tau := \frac{T}{T_c} = k_B T\frac{27b}{8a}$$

equation (6.14) becomes:

$$\left(\pi + \frac{3}{\nu^2}\right)(3\nu - 1) = 8\tau \tag{6.16}$$
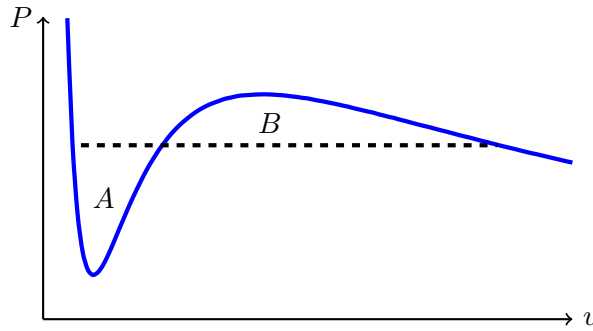
Figure 6.4: Maxwell's equal area rule

We have found another very interesting result: *when rescaled by their critical thermodynamic properties, all fluids obey the same state equation*. This is the *law of corresponding states* that we have already encountered in 1.5 (figure 1.6).

This is a form of universality, but substantially different from the one we have seen until now, which applies only in the neighbourhood of a critical point: in fact, the law of corresponding states applies *everywhere* on the phase diagram. It can even be shown that this law is a consequence of dimensional analysis, and is more general than what might seem: experimentally the law of corresponding states is well satisfied also by fluids which do not obey Van der Waals equation.

Maxwell's equal area rule

As we have previously anticipated, *Maxwell's equal area rule* is a method to "manually" remove the unphysical regions of Van der Waals isotherms.

From 1.2.2 we know that at the coexistence of two phases the chemical potentials and the pressures of the two phases must be equal; furthermore, from 1.1.3 we also know that the chemical potential is the Gibbs free energy per particle, namely $G = \mu N$, and in general we have also:

$$dG = -S dT + V dP + \mu dN$$

Now, differentiating $G = \mu N$ and subtracting this last equation we get:

$$d\mu = -\frac{S}{N} dT + \frac{V}{N} dP$$

Therefore, since along an isotherm $dT = 0$, we will have:

$$\mu_{\text{gas}} - \mu_{\text{liq.}} = \int_{\text{liq.}}^{\text{gas}} d\mu = \frac{1}{N} \int_{\text{liq.}}^{\text{gas}} V dP = 0$$

Looking also at figure 6.4, we see that this means that the horizontal segment of the isotherm must be drawn so that regions A and B have the same area (from which the name of the method).

Critical behaviour

Let us now study the behaviour of systems obeying Van der Waals equations near the critical point, computing one of the critical exponents.

Exponent β This exponent[13] can be computed from the shape of the coexistence curve for $T \to T_c^-$; this can be done using the law of corresponding states. In fact, defining:

$$t := \frac{T - T_c}{T_c} = \tau - 1 \qquad\qquad \varrho := \frac{V - V_c}{V_c} = \nu - 1$$

we can expand the law of corresponding states (6.16) near $\tau = \nu = 1$, or equivalently $t = \varrho = 0$ (namely in the neighbourhood of the critical point):

$$\pi = \frac{8\tau}{3\nu - 1} - \frac{3}{\nu^2} = \frac{8(1 + t)}{3(1 + \varrho) - 1} - \frac{3}{(1 + \varrho)^2} =$$

$$= (1 + t)\left(4 - 6\varrho + 9\varrho^2 - \frac{27}{2}\varrho^3 + \cdots\right) - (3 - 6\varrho + 9\varrho^2 - 12\varrho^3 + \cdots) \sim$$

$$\sim 1 + 4t - 6\varrho t + 9\varrho^2 t - \frac{3}{2}\varrho^3 - \frac{27}{2}\varrho^3 t + \cdots$$

Therefore:

$$\pi = 1 + 4t - 6\varrho t - \frac{3}{2}\varrho^3 + O(t\varrho^2, \varrho^4) \tag{6.17}$$

where the terms we have neglected are justified a posteriori (i.e. we will see that $\varrho \sim t^{1/2}$; we could have not neglected them, but the result of the computation doesn't change).
The strategy we want to apply is the following: since we want to determine how ϱ changes with t, we can determine the relation between the densities ϱ_g and ϱ_l in the gaseous and liquid phase from Maxwell's equal area rule. This way, from (6.17) we can determine the pressures in the two phases and express them in terms of ρ_g or ϱ_l, and since $\pi_l = \pi_g$ at the coexistence we can obtain from this equation the behaviour of ϱ in terms of t.

Therefore, from Maxwell's equal area rule we have:

$$\int_{\text{liq.}}^{\text{gas}} v dP = \int_{\text{liq.}}^{\text{gas}} (\varrho + 1)v_c P_c d\pi = 0$$

where we have used the definitions of π and ϱ.
If we now set our system on an isotherm with $t < 0$ and t small (so that the expansion in (6.17) makes sense), the variation of π in terms of ϱ is:

$$d\pi = -6td\varrho - \frac{9}{2}\varrho^2 d\varrho$$

We thus have[14]:

$$\int_{\varrho_l}^{\varrho_g} \varrho\left(-6t - \frac{9}{2}\varrho^2\right) d\varrho = 0 \qquad \Rightarrow \qquad 3\varrho_g^2\left(t + \frac{\varrho_g^2}{8}\right) = 3\varrho_l^2\left(t + \frac{\varrho_l^2}{8}\right)$$

Since t is small we can neglect it, and so:

$$\varrho_g^2 = \varrho_l^2 \qquad \Rightarrow \qquad \varrho_g = \pm\varrho_l$$

---

[13]Remember that by definition β describes the behaviour of the order parameter in the neighbourhood of the critical temperature, so we will have $\varrho \sim |t|^{\beta}$.

[14]If we didn't neglected the term of the expansion of π, we would have found:

$$-3\varrho_g^2\left(t - 2t\varrho_g + \frac{3}{8}\varrho_g^2\right) = -3\varrho_l^2\left(t - 2t\varrho_l + \frac{3}{8}\varrho_l^2\right)$$

Again, the terms linear in t can be neglected since t is small (and ϱ_g, ϱ_l are just numbers). (???)

Remembering that:

$$\varrho_l = \frac{v_l - v_c}{v_c} \qquad\qquad \varrho_g = \frac{v_g - v_c}{v_c}$$

we see that the only acceptable solution is $\varrho_g = -\varrho_l$ (since the volume of a gas is larger than that of a liquid).

Therefore, substituting ϱ_g and $\varrho_l = -\varrho_g$ into π we get:

$$\pi_g = 1 + 4t - 6t\varrho_g - \frac{3}{2}\varrho_g^3 \qquad\qquad \pi_l = 1 + 4t + 6t\varrho_g + \frac{3}{2}\varrho_g^3$$

Since $\pi_g = \pi_l$ at the phase coexistence, we have:

$$3\varrho_g \left(4t + \varrho_g^2\right) = 0$$

and excluding of course the case $\varrho_g = 0$, in the end:

$$\varrho_g \sim t^{1/2}$$

Therefore:

$$\beta = \frac{1}{2}$$

which is what we could have expected from a mean field theory.

In fact, if we compute all the other critical exponent, we get exactly:

$$\alpha = 0 \qquad\qquad \beta = \frac{1}{2} \qquad\qquad \gamma = 1 \qquad\qquad \delta = 3 \qquad\qquad (6.18)$$

A more precise approximation

We have seen that the problem of Van der Waals equation comes from the rough approximation that we have made in (6.13).

A better formulation of Van der Waals mean field theory can be done using the potential:

$$\varphi(\vec{r}) = \begin{cases} +\infty & |\vec{r}| < r_0 \\ -ke^{-k|\vec{r}|} & |\vec{r}| \geq r_0 \end{cases} \qquad\qquad (6.19)$$

where k is a parameter that determines the range of the potential.

In this case (we don't do the computations, we just state the result) after having computed the partition function, in the limit $k \to 0$ the potential becomes infinitely ranged and weak. The interesting fact is that in this limit the theory becomes essentially the same that we have seen before, but with the exception that there are no unphysical regions and Maxwell's rule is no longer necessary, since this model naturally predicts the existence of the "plateaux" relative to the liquid-gaseous phase transition in (v, P) plane.

6.4.2 Mean field theories for weakly interacting systems

If a system is composed of weakly interacting particles we can use perturbative methods to compute the partition function of such systems.

For example, consider a fluid composed of N particles in a region of volume V, interacting through a generic two-body potential that depends only on the relative distance between the particles:

$$U(\{\vec{r}\}) = \frac{1}{2} \sum_{i \neq j = 1}^{N} \varphi(|\vec{r}_i - \vec{r}_j|)$$

Its Hamiltonian will be:

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\vec{p_i}^2}{2m} + \frac{1}{2} \sum_{i \neq j=1}^{N} \varphi(|\vec{r}_i - \vec{r}_j|)$$

and its partition function:

$$Z_N = \frac{1}{h^{3N} N!} \int \prod_{i=1}^{N} d\vec{q_i} d\vec{p_i} e^{-\beta \mathcal{H}} = \frac{1}{N! \Lambda^{3N}} Q_N$$

where Λ is the thermal wavelength and Q_N is the configurational sum:

$$Q_N = \int \prod_i d\vec{r_i} e^{-\beta U(\{\vec{r_i}\})}$$

Of course for ideal gases $U = 0$ and so $Q_N = V^N$, and the dependence on the temperature is included only in Λ; if we consider also the interaction terms we must insert a correction χ in the configurational contribution to the partition function[15]:

$$Q_N = V^N \cdot \chi(N, V, T)$$

which (depending on the possible presence of attractive terms in the interaction potential φ) can in general be also a function of the temperature T; furthermore the correction depends strongly on the gas density: if it is low the particles will not "perceive" the presence of the other ones and the ideal gas approximation is a good one, while for high densities the particles will be closer to each other and corrections to Q_N are necessary.

Let us note that inserting the correction χ the free energy of the system will be:

$$F = F_{\text{ideal}} - k_B T \ln \chi$$

Virial and cluster expansion

The *virial expansion* is a systematic approach that can be used to incorporate the corrections due to the interactions between the particles, and as we will shortly see it can be obtained from a much general method, the *cluster expansion*.

The virial expansion consists in expanding the thermodynamic quantities of a system in powers of the density ρ; for example, the virial expansion for the pressure of a gas is:

$$\frac{P}{k_B T} = \rho + B_2 \rho^2 + B_3 \rho^3 + \cdots$$

where B_n are called *virial coefficients*, and in general they can depend on the temperature. The virial expansion is very useful because the coefficients B_n can be experimentally measured (for example, in the case of the pressure they can be determined by properly fitting the isotherms of a system), and as we will see they can be related to microscopic properties of the interparticle interaction.

Let us see for example the virial expansion of the Van der Waals equation. From (6.14) we have:

$$\frac{P}{k_B T} = \frac{N}{V - Nb} - \frac{aN^2}{k_B T V^2}$$

which can be rewritten as:

$$\frac{P}{k_B T} = \frac{N}{V} \left(1 - b\frac{N}{V} \right)^{-1} - \frac{a}{k_B T} \left(\frac{N}{V} \right)^2 = \rho(1 - b\rho)^{-1} - \frac{a}{k_B T} \rho^2$$

---

[15]We can always insert a multiplicative correction. We could have also written it as an additive correction, but the core of the subject doesn't change.

where of course $\rho = N/V$ is the density of the system. Therefore, expanding the first term in a Taylor series:

$$\frac{P}{k_B T} = \rho + \left(b - \frac{a}{k_B T} \right) \rho^2 + b^2 \rho^3 + b^3 \rho^4 + \cdots \tag{6.20}$$

We can thus immediately identify the first virial coefficient:

$$B_2(T) = b - \frac{a}{k_B T}$$

The temperature T_B at which B_2 vanishes is in general called *Boyle temperature*; in this case $T_B = \frac{a}{b k_B}$.

Now, let us see how the cluster expansion works and how we can obtain the virial expansion from it. Of course, we start from the general configurational partition function:

$$Q_N = \int \prod_i d\vec{r}_i e^{-\beta U(\{\vec{r}_i\})}$$

The idea is to find a "small quantity" in terms of which we can expand Q_N; this quantity is the so called *Mayer function*:

$$f(\vec{r}) = e^{-\beta \varphi(\vec{r})} - 1$$

In fact, when the gas is ideal $f(\vec{r}) = 0$, and if the particles interact weakly φ is small, and so is $f(\vec{r})$. In particular, this expansion will work well for low densities (namely $|\vec{r}_i - \vec{r}_j|$ is large and so $\varphi(|\vec{r}_i - \vec{r}_j|) \to 0$) or high temperatures (namely $\beta \to 0$): in both cases, in fact, $e^{-\beta \varphi(|\vec{r}_i - \vec{r}_j|)} \to 1$ and $f(|\vec{r}_i - \vec{r}_j|) \to 0$.
Using the short notations $\varphi_{ij} = \varphi(|\vec{r}_i - \vec{r}_j|)$ and $f_{ij} = f(|\vec{r}_i - \vec{r}_j|)$ we have:

$$e^{-\beta \frac{1}{2} \sum_{i \neq j=1}^N \varphi_{ij}} = \prod_{i \neq j} e^{-\frac{\beta}{2} \varphi_{ij}} = \prod_{i \neq j} \sqrt{1 + f_{ij}} =$$
$$= \sqrt{1 + f_{12}} \sqrt{1 + f_{13}} \cdots \sqrt{1 + f_{1N}} \sqrt{1 + f_{21}} \sqrt{1 + f_{23}} \cdots$$
$$\cdots \sqrt{1 + f_{N1}} \sqrt{1 + f_{N2}} \cdots \sqrt{1 + f_{N,N-1}}$$

Now, since $f_{ij} = f_{ji}$ by definition, in the end we will have:

$$e^{-\beta \frac{1}{2} \sum_{i \neq j=1}^N \varphi_{ij}} = (1 + f_{12})(1 + f_{13}) \cdots = 1 + \frac{1}{2} \sum_{i \neq j=1}^N f_{ij} + \cdots$$

where the missing terms contain the product of two or more Mayer functions. Therefore, the configurational contribution to the partition function will be:

$$Q_N = \int \prod_{k=1}^N d\vec{r}_k \left(1 + \frac{1}{2} \sum_{i \neq j=1}^N f_{ij} + \cdots \right) = V^N + \int \prod_{k=1}^N d\vec{r}_k \frac{1}{2} \sum_{i \neq j=1}^N f_{ij} + \cdots =$$
$$= V^N + V^{N-2} \frac{1}{2} \sum_{i \neq j=1}^N \int d\vec{r}_i d\vec{r}_j f_{ij} + \cdots$$

where in the last step we have extracted the contributions of the integrals with $k \neq i, j$. Now, the remaining integral can be easily computed with the definition of the new variable $\vec{r} = \vec{r}_i - \vec{r}_j$:

$$\int d\vec{r}_i d\vec{r}_j f_{ij} = \int d\vec{r}_i d\vec{r}_j f(|\vec{r}_i - \vec{r}_j|) = \int d\vec{r}_i \int d\vec{r} f(|\vec{r}|) = V \int d\vec{r} f(|\vec{r}|) := -2V B_2$$

where in the last step we have *defined* the first virial coefficient. Of course, we should have set $\int d\vec{r} f(|\vec{r}|)$ equal to a generic coefficient, but in the end we should have found exactly that this coefficient is $-2B_2$: we have done this for the sake of simplicity.
Therefore:

$$B_2 = -\frac{1}{2} \int f(|\vec{r}|) d\vec{r} \tag{6.21}$$

From this we see precisely how the first virial coefficient, which as we have already stated can be experimentally measured, is related to the microscopic properties of the interaction between the particles, represented by the Mayer function f.
It can also be shown that all the virial coefficients can be expressed in terms of integrals of products of Mayer functions; for example the second virial coefficient is:

$$B_3 = -\frac{1}{3} \int d\vec{r} d\vec{s} f(|\vec{r}|) f(|\vec{s}|) f(|\vec{r} - \vec{s}|)$$

Higher order coefficients involve the computation of increasingly difficult integrals, which can however be visualized in terms of graphs.

What we have seen now is how the cluster expansion works in general. Let us now apply it in order to find the virial expansion for real gases.
From what we have found, the configurational partition function of the system becomes:

$$Q_N = V^N - V^{N-1} B_2 \sum_{i \neq j = 1}^{N} 1 + \cdots$$

The remaining sum is equal to $N(N-1)$: in fact, for any of the N values that i can assume, j can have $N-1$ values. Therefore:

$$Q_N = V^N - V^{N-1} N(N-1) B_2 + \cdots$$

and, considering that $N - 1 \approx N$ for large N, the complete partition function of the system will be:

$$Z_N = \frac{V^N}{N! \Lambda^{3N}} \left(1 - \frac{N^2}{V} B_2 + \cdots \right)$$

We recognise in this expression that $(1 + B_2 N^2 / V + \cdots)$ is the correction χ to the ideal gas partition function that we have mentioned earlier; therefore, the free energy of the system will be:

$$F = F_{\text{ideal}} - k_B T \ln \left(1 - \frac{N^2}{V} B_2 + \cdots \right)$$

and its pressure:

$$P = -\frac{\partial F}{\partial V}\Big|_{N,T} = \frac{N k_B T}{V} + k_B T \frac{\frac{N^2}{V^2} B_2}{1 - \frac{N^2}{V} B_2} + \cdots = \frac{N k_B T}{V} \left(1 + \frac{\frac{N}{V} B_2}{1 - \frac{N^2}{V} B_2} + \cdots \right)$$

Now, neglecting the terms involving B_n with $n \geq 3$, and expanding in terms of B_2 (in fact $B_2 \sim f$, which is small)[16]:

$$P = \frac{N k_B T}{V} \left(1 + \frac{N}{V} B_2 + \cdots \right) \tag{6.22}$$

---

[16]Referring to what we have stated previously, if we set $\int d\vec{r} f(|\vec{r}|) = \mathfrak{F}$ with $\mathfrak{F}$ a generic constant, then we should have found:

$$Q_N = V^N \left(1 + \frac{N(N-1)}{V} \cdot \frac{\mathfrak{F}}{2} + \cdots \right)$$

and proceeding like we have done now, in the end:

$$P = \frac{N k_B T}{V} \left(1 - \frac{N}{V} \cdot \frac{\mathfrak{F}}{2} + \cdots \right)$$

and so from (6.20) we see that indeed $\mathfrak{F} = -2B_2$.

This expansion contains only low-order terms in the density N/V, so strictly speaking it is valid only for low densities. However, we can use a "trick" in order to extend its range; in fact, remembering that the McLaurin expansion of $(1-x)^{-1}$ is $1+x+\cdots$, from (6.22) we can write:

$$\frac{PV}{Nk_BT} = 1 + B_2\rho + \cdots \approx \frac{1}{1 - B_2\rho}$$

and now re-expand $(1 - B_2\rho)^{-1}$, so that we can express all the virial coefficients in terms of the first one:

$$\frac{1}{1 - B_2\rho} = 1 + B_2\rho + (B_2)^2\rho^2 + (B_2)^3\rho^3 + \cdots \qquad \Rightarrow$$

$$\Rightarrow \qquad \frac{PV}{Nk_BT} = \frac{P}{\rho k_BT} = 1 + B_2\rho + (B_2)^2\rho^2 + (B_2)^3\rho^3 + \cdots \qquad \Rightarrow$$

$$\Rightarrow \qquad \frac{P}{k_BT} = \rho + B_2\rho^2 + (B_2)^2\rho^3 + (B_2)^3\rho^4 + \cdots$$

So in the end:

$$B_3 \sim (B_2)^2 \qquad B_4 \sim (B_2)^3 \qquad \cdots \qquad B_n \sim (B_2)^{n-1}$$

Computation of virial coefficients for some interaction potentials

Let us now see this method in action by explicitly computing some coefficients B_2 for particular interaction potentials.

Hard sphere potential As a first trial, we use a hard sphere potential (6.13) similar to the one we have seen for the derivation of the Van der Waals equation:

$$\varphi(\vec{r}) = \begin{cases} +\infty & |\vec{r}| < r_0 \\ 0 & |\vec{r}| \geq r_0 \end{cases}$$

(the difference with (6.13) is that now the potential is purely repulsive, and has no attractive component).
In this case:

$$f(\vec{r}) = e^{-\beta\varphi(\vec{r})} - 1 = \begin{cases} -1 & |\vec{r}| < r_0 \\ 0 & |\vec{r}| \geq r_0 \end{cases}$$

Therefore, using (6.21) and shifting to spherical coordinates:

$$B_2 = -\frac{1}{2}\int f(|\vec{r}|)d\vec{r} = -\frac{1}{2}\int_0^{+\infty} 4\pi r^2\left(e^{-\beta\varphi(r)} - 1\right)dr = 2\pi\int_0^{r_0} r^2 dr = \frac{2}{3}\pi r_0^3$$

In this case B_2 does not depend on the temperature, and in the end we have:

$$PV = Nk_BT\left(1 + \frac{2}{3}\pi r_0^3 \frac{N}{V}\right)$$

Square well potential We now use a slight refinement of the previous potential:

$$\varphi(\vec{r}) = \begin{cases} +\infty & |\vec{r}| < r_0 \\ -\varepsilon & r_0 < |\vec{r}| < r_0 + \delta \\ 0 & |\vec{r}| > r_0 + \delta \end{cases}$$

This can be seen as a hard sphere potential where the spheres have an attractive shell of thickness δ.

We thus have:

$$f(\vec{r}) = \begin{cases} -1 & |\vec{r}| < r_0 \\ e^{\beta \varepsilon} - 1 & r_0 < |\vec{r}| < r_0 + \delta \\ 0 & |\vec{r}| > r_0 + \delta \end{cases}$$

so that:

$$B_2 = -\frac{1}{2} \int f(|\vec{r}|) d\vec{r} = -\frac{1}{2} \int 4\pi r^2 f(r) dr = -2\pi \left[\int_0^{r_0} (-r^2) dr + \int_{r_0}^{r_0+\delta} \left(e^{\beta \varepsilon} - 1 \right) r^2 dr \right] =$$

$$= -2\pi \left\{ -\frac{r_0^3}{3} + \frac{e^{\beta \varepsilon} - 1}{3} \left[(r_0 + \delta)^3 - r_0^3 \right] \right\} = B_2^{\text{h.s.}} - \frac{2}{3}\pi \left(e^{\beta \varepsilon} - 1 \right) \left[(r_0 + \delta)^3 - r_0^3 \right]$$

where $B_2^{\text{h.s.}}$ is the first virial coefficient of the hard sphere potential we have previously seen. Now, if the temperature is sufficiently high, namely $\beta \varepsilon \ll 1$, we can approximate $e^{\beta \varepsilon} - 1 \approx \beta \varepsilon$, so that:

$$B_2 = B_2^{\text{h.s.}} - \frac{2}{3}\pi \beta \varepsilon r_0^3 \left[\left(1 + \frac{\delta}{r_0} \right)^3 - 1 \right]$$

For the sake of simplicity, defining:

$$\lambda := \left(1 + \frac{\delta}{r_0} \right)^3 - 1$$

we will have, in the end:

$$\frac{PV}{Nk_BT} = 1 + B_2 \rho = 1 + \left(B_2^{\text{h.s.}} - \frac{2}{3}\frac{\pi \varepsilon}{k_BT} r_0^3 \lambda \right) \rho$$

so in this case B_2 actually depends on the temperature.

Lennard-Jones potential   This potential is a quite realistic representation of the interatomic interactions. It is defined as:

$$\varphi(\vec{r}) = 4\varepsilon \left[\left(\frac{r_0}{|\vec{r}|} \right)^{12} - \left(\frac{r_0}{|\vec{r}|} \right)^6 \right]$$

which contains a long-range attractive term (the one proportional to $1/|\vec{r}|^6$, which can be justified in terms of electric dipole fluctuations) and a short-range repulsive one (proportional to $1/|\vec{r}|^{12}$, which comes from the overlap of the electron orbitals).

With this interaction potential, the first virial coefficient is:

$$B_2 = -2\pi \int_0^{+\infty} r^2 \left\{ e^{-\frac{4\varepsilon}{k_BT} \left[\left(\frac{r_0}{|\vec{r}|} \right)^{12} - \left(\frac{r_0}{|\vec{r}|} \right)^6 \right]} - 1 \right\} dr$$

which is not analytically computable. However, it can be simplified defining the variables $x = r/r_0$ and $T^* = k_BT/\varepsilon$ so that:

$$B_2 = \frac{2}{3}\pi r_0^3 \frac{4}{T^*} \int_0^\infty x^2 \left(\frac{12}{x^{12}} - \frac{6}{x^6} \right) e^{-\frac{4}{T^*} \left(\frac{1}{x^{12}} - \frac{1}{x^6} \right)} dx$$

Now, we can expand the exponential and integrate term by term; this gives an expression of B_2 as a power series of $1/T^*$:

$$B_2 = -2A \sum_{n=0}^{\infty} \frac{1}{4n!} \Gamma \left(\frac{2n-1}{4} \right) \left(\frac{1}{T^*} \right)^{(2n+1)/4}$$

where Γ is the Euler gamma function and A is a constant.
Note that the attractive part of the Lennard-Jones potential has introduced in B_2 a dependence on the temperature.

6.5 Landau theory

6.5.1 Introduction

Landau theory is a *phenomenological* mean field theory that aims at describing the occurence of phase transitions in a unitary framework.

We have already stated that even very different many-particle systems exhibit *universality*: we have seen this explicitly in 1.5, but also in the study of mean field theories we have that two very different systems, namely an Ising magnet in the Weiss approximation and a fluid in the Van der Waals theory, have the very same critical exponents (see equations (6.7) and (6.18)).
The reason of this fact can be understood if we expand the state equations of these systems in terms of their order parameters. In fact, considering the Van der Waals theory for fluids we have seen in 6.4.1 that:
$$\pi = 1 + 4t - 6\eta t - \frac{3}{2}\eta^3 + O(\eta^4, \eta^2 t)$$

where we have called $\eta := \varrho$ the order parameter.
In a similar way we can expand the state equation of an Ising model in the Weiss approximation. In this case we get:
$$\frac{H}{k_B T} = \eta t + \eta^3 + O(t\eta^3)$$

where $\eta := m$ is again the order parameter.
Therefore we can see how the two equations of state, when expanded around a critical point, behave in the same way, i.e. they involve the same powers of the order parameter η: this is why these two very different systems exhibit the same critical exponents.

From these very simple examples we can prefigure that the properties and the possible occurrence of phase transitions of every universality class can be described once we have expanded the state equation of the system in terms of its order parameter (whatever it is, and independently of its nature).
These kinds of observations led Landau to suggest that such considerations can be done for *all* phase transitions, at least qualitatively. The resulting theory, which we now proceed to study, is the so called *Landau theory* for phase transitions.

6.5.2 Assumptions of Landau theory

Landau theory is based on some assumptions, which we now introduce:

– There exists an *order parameter η* for the system, such that:

$$\begin{cases} \eta = 0 & T > \overline{T} \\ \eta \neq 0 & T < \overline{T} \end{cases}$$

– There exists a function $\mathcal{L}$ called *Landau free energy*[17], which is an analytic function of the coupling constants $\{K_i\}$ of the system and of the order parameter η.
Therefore, near $\overline{T}$ we can expand $\mathcal{L}$ for small η:

$$\mathcal{L} = \sum_{n=0}^{\infty} a_n \eta^n = a_0 + a_1 \eta + a_2 \eta^2 + a_3 \eta^3 + a_4 \eta^4 + O(\eta^5)$$

– $\mathcal{L}$ has to be consistent with the symmetries of the system

– The equilibrium states of the system are the global minima of $\mathcal{L}$ with respect to η

We also assume that the thermodynamic properties of the system can be obtained by differentiating $\mathcal{L}$, just like we can do with thermodynamic potentials.[18]

From these assumption we can explicitly construct $\mathcal{L}$, depending of course on the system considered.

Note also that the general formulation of the Landau theory does not depend on the dimensionality of the system (although we will see that once a system has been chosen some details can depend on it).

6.5.3 Landau theory for the Ising model

To make things clearer, let us now consider the Ising model without any external field and see how we can determine the form of $\mathcal{L}$ from the general assumptions we have introduced.

Construction of $\mathcal{L}$

First of all, since the equilibrium configurations of the system must be minima of $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial \eta} = a_1 + 2a_2 \eta + 3a_3 \eta^2 + 4a_4 \eta^3 = 0$$

where we have chosen to stop the expansion at the fourth order[19]. Now, since this equation must hold for all T and for $T > \overline{T}$ we have $\eta = 0$, we see that $a_1 = 0$.

Considering now the constraint on the symmetries of the system, in 5.3.3 we have seen that the Ising model is invariant under parity, i.e. its Hamiltonian is simultaneously even in H and $\{S_i\}$: $\mathcal{H}(H, \{S_i\}) = \mathcal{H}(-H, \{-S_i\})$. Thus, in absence of external fields the Hamiltonian of the Ising model is even; this means that also $\mathcal{L}$ must be invariant under parity, namely an even function of η: $\mathcal{L}(-\eta) = \mathcal{L}(\eta)$. Therefore all the odd terms of the expansion are null:

$$a_{2k+1} = 0 \qquad \forall k \in \mathbb{N}$$

Finally, since we have assumed that $\mathcal{L}$ is an analytic function of η then its expansion cannot contain terms like $|\eta|$.

---

[17]To be more precise, $\mathcal{L}$ is the Landau free energy density; the "real" Landau free energy should be $L = V\mathcal{L}$.

[18]Strictly speaking, Landau free energy is *not* really a thermodynamic potential: the correct interpretation of $\mathcal{L}$ (see [3], 5.6.1) is that it is a *coarse grained* free energy (not the *exact* one); see 6.6.2 for coarse graining procedures.

[19]This choice has been made because we suppose η to be small and also because in both the Weiss and Van der Waals theories we have seen that the relevant physics involves third and fourth powers of the order parameter. Of course, the choice of where to stop the expansion depends on the system considered (in particular, it depends also on the dimensionality of the system).

In conclusion, the minimal form of the Landau free energy for the Ising model is:

$$\mathcal{L} = a_0 + a_2\eta^2 + a_4\eta^4 \tag{6.23}$$

where a_0, a_2 and a_4 can in general be functions of J_{ij} (the coupling constants) and T.
However, (6.23) can be further simplified and we can also explicitly show its dependence on the temperature.
In fact, first of all we can note that a_0 is the value of $\mathcal{L}$ in the paramagnetic state (when $T > \overline{T}$, $\eta = \overline{\eta} = 0$ and $\mathcal{L}(\overline{\eta}) = a_0$), and so for simplicity we can set $a_0 = 0$ (it's just a constant shift in the energy).
Then, we must also require $a_4 > 0$ because otherwise $\mathcal{L}$ can be minimized by $\eta \to \pm\infty$, which makes no sense physically.
Finally, expanding a_2 and a_4 in T near $\overline{T}$:

$$a_2 = a_2^0 + t\frac{a}{2} + \cdots \qquad\qquad a_4 = \frac{b}{4} + \cdots$$

where we have renamed the constants this way for later convenience, and $t = (T - \overline{T})/\overline{T}$ is the reduced temperature; in the expansion of a_4 we have neglected any explicit dependence on $T - \overline{T}$ because as we will see it will not dominate the behaviour of the thermodynamics near $\overline{T}$[20].
Taking $a_2^0 = 0$ we finally have that the form of the Landau free energy for the Ising model is:

$$\mathcal{L} = \frac{a}{2}t\eta^2 + \frac{b}{4}\eta^4 \tag{6.24}$$

If an external field H is present, then:

$$\mathcal{L} = \frac{a}{2}t\eta^2 + \frac{b}{4}\eta^4 - H\eta \tag{6.25}$$

Critical properties of the Landau theory for the Ising model

Let us now see what does the Landau theory for the Ising model predict.
First of all, in the absence of external fields from (6.24) we have that the equilibrium states are determined by:

$$\frac{\partial\mathcal{L}}{\partial\eta} = a\eta t + b\eta^3 = \eta(at - b\eta^2) = 0 \qquad \Rightarrow \qquad \eta = \overline{\eta} = \begin{cases} 0 & t > 0 \text{ (i.e. } T > \overline{T}) \\ \pm\sqrt{-at/b} & t < 0 \text{ (i.e. } T < \overline{T}) \end{cases}$$

We therefore see that a phase transition really occurs: when the temperature is greater than the critical one, the only global minimum of $\mathcal{L}$ is $\overline{\eta} = 0$, while for $T < \overline{T}$ two possible new equilibrium states appear (which of course are the two possible values that the spontaneous magnetization can assume in the magnetic phase).

Let us therefore see what critical exponents does the Landau theory for the Ising model predict.

Exponent β This is immediately determined from what we have just seen:

$$\eta = \pm\left(-\frac{a}{b}t\right)^{1/2} \sim (-t)^{1/2} \qquad \text{when} \quad t < 0 \qquad \Rightarrow \qquad \beta = \frac{1}{2}$$

---

[20]In other words, the occurrence of the phase transition at $T = \overline{T}$ depends only on how a_2 depends on t.

Exponent α The specific heat at constant external field of the system is $C_H = -T\partial^2\mathcal{L}/\partial T^2$; as we have seen, $\mathcal{L} = 0$ for $t > 0$ (since we have set $a_2^0 = 0$) while $\mathcal{L} = \mathcal{L}(\bar{\eta}) = -a^2 t^2/4b$ for $t < 0$. Therefore:

$$C_H = \begin{cases} 0 & t > 0 \\ \frac{a^2}{2b\overline{T}^2}T & t < 0 \end{cases}$$

We thus see that C_H has a jump discontinuity at $T = \overline{T}$, so:

$$\alpha = 0$$

Exponent δ Considering now also an external field, the state equation of the system will be given by the differentiation of (6.25):

$$\frac{\partial\mathcal{L}}{\partial\eta} = at\eta + b\eta^3 - H = 0 \qquad \Rightarrow \qquad at\eta + b\eta^3 = H$$

At the critical point $t = 0$ and so we have $H \propto \eta^3$. Therefore:

$$\delta = 3$$

Exponent γ If we now differentiate the state equation with respect to H we get:

$$at\frac{\partial\eta}{\partial H} + 3b\eta^2\frac{\partial\eta}{\partial H} = 1$$

Since $\partial\eta/\partial H = \chi_T$, then:

$$\chi_T = \frac{1}{at + 3b\eta^2}$$

If we now set $H = 0$ then for $t > 0$ we will have $\eta = \bar{\eta} = 0$ and thus $\chi_T = (at)^{-1}$; on the other hand, if $t < 0$ then $\eta = \bar{\eta} = \pm(-at/b)^{1/2}$ and therefore $\chi_T = (-2at)^{-1}$. In both cases $\chi_T \sim t^{-1}$, and thus:

$$\gamma = 1$$

Therefore, the Landau theory for the Ising model gives the critical exponents:

$$\alpha = 0 \qquad\qquad \beta = \frac{1}{2} \qquad\qquad \gamma = 1 \qquad\qquad \delta = 3$$

which, as we expected, are identical to those we have found within Weiss mean field theory (see equation (6.7)).

6.5.4   First-order phase transitions in Landau theory

As we have seen, Landau theory is based on the assumption that the order parameter is small near the critical point, and we have seen in the example of the Ising model how it can describe a continuous phase transition (in fact, for $t \to 0$ we have $\eta \to 0$). However, because of the symmetry properties of the Ising model we have excluded any possible cubic term; what we now want to do is to consider a more general form of $\mathcal{L}$ which includes also a cubic term in η, and see that this leads to the occurrence of a first-order phase transition.

We have seen that since the order parameter is null for $T > \overline{T}$ the Landau free energy can't contain any linear term in η. Let us therefore consider:

$$\mathcal{L} = \frac{a}{2}t\eta^2 - w\eta^3 + \frac{b}{4}\eta^4 - H\eta \qquad\qquad t = \frac{T - \overline{T}}{2}$$

(a) First-order transition

(b) Same transition for lower values of the temperature
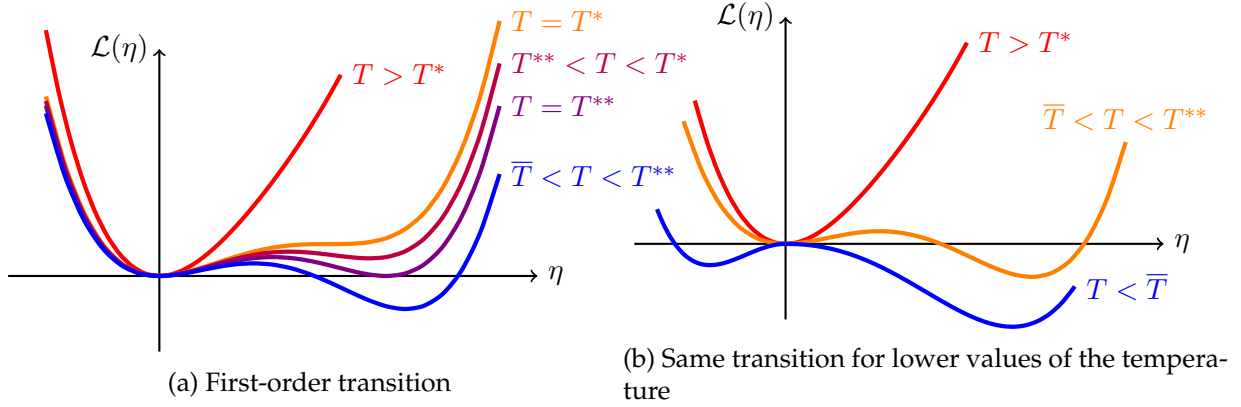
Figure 6.5: Behaviour of $\mathcal{L}$ for different values of T

where we choose $w > 0$ and we have redefined t for convenience; as in the previous case, we must have $b > 0$ so that η has finite values in the equilibrium configurations. The temperature $\overline{T}$ is the one at which we have the continuous transition if $w = 0$, but as we will see it doesn't have great significance now.

The equilibrium configurations of the system, in absence of external fields, will be given by:

$$\frac{\partial \mathcal{L}}{\partial \eta} = at\eta - 3w\eta^2 + b\eta^3 = 0 \qquad \Rightarrow \qquad \eta(at - 3w\eta + b\eta^2) = 0$$

The solutions of this equation are $\eta = \overline{\eta} = 0$ (disordered phase) and:

$$\overline{\eta}_\pm = \frac{3w \pm \sqrt{9w^2 - 4atb}}{2b} = c \pm \sqrt{c^2 - \frac{at}{b}}$$

which correspond to the ordered phase, and for the sake of simplicity in the last step we have defined:

$$c := \frac{3w}{2b}$$

However, these two last solutions are possible only if:

$$c^2 - \frac{at}{b} > 0 \qquad \Rightarrow \qquad t < t^* := \frac{bc^2}{a}$$

Let us note that since t^* is positive, this will occur at temperatures higher than $\overline{T}$ (because for $T = \overline{T}$ we have $t = 0$).

Intuitively, since $\mathcal{L} \to \infty$ for $\eta \to +\infty$ and $\mathcal{L} \to 0$ for $\eta \to 0$, we understand that $\overline{\eta}_-$ will be a maximum while $\overline{\eta}_+$ a minimum (see also figure 6.5).

We therefore have that the introduction of the cubic term brings to an asymmetry in $\mathcal{L}$ which leads to the formation of another minimum at $\eta = \overline{\eta}_+$ for $T < T^*$.

Let us try to understand how $\mathcal{L}$ behaves as a function of T, by also explicitly plotting $\mathcal{L}$ as shown in figure 6.5[21]. If we start from $T > T^*$, then the system will be in the disordered phase and $\mathcal{L}$ will have only one minimum at $\eta = \overline{\eta} = 0$. When T becomes smaller than T^* a new minimum appears at $\eta = \overline{\eta}_+$, but at the beginning we will have $\mathcal{L}(\overline{\eta}_+) > 0$ so this is only a local minimum (since $\mathcal{L}(0) = 0$): in this range of temperatures the ordered phase is *metastable*. If we further decrease the temperature, we will reach a temperature T^{**} for which $\mathcal{L}(\overline{\eta}_+) = 0 = \mathcal{L}(0)$:

---

[21]Of course all these considerations can be made more rigorous with a complete study of the function $\mathcal{L}$, which we don't do since it is rather straightforward and not illuminating.

at this point the ordered and disordered phase coexist, so this is the temperature of a new transition! If we now further decrease the temperature to values lower than T^{**}, $\mathcal{L}(\overline{\eta}_+)$ becomes negative and so now $\overline{\eta}_+$ is the global minimum of $\mathcal{L}$: the ordered phase becomes stable and the disordered phase metastable.

If now T becomes smaller than $\overline{T}$, $\mathcal{L}$ develops a new minimum for $\eta < 0$, but it is only a local minimum (the asymmetry introduced by $-w\eta^3$ ensures that $\overline{\eta}_+$ is always the global minimum). This means that also for $T < \overline{T}$ the disordered phase with $\eta = \overline{\eta}_+$ continues to be the stable one, and so no phase transition occurs at $\overline{T}$ any more; this is what we meant when we said that $\overline{T}$ is not a relevant temperature any more.

Therefore, we see that lowering the temperature of the system the value of η for which $\mathcal{L}$ has a global minimum changes *discontinuously* from $\eta = 0$ to $\eta = \overline{\eta}_+$: this is a *first-order transition*.

As we have seen, the temperature T^{**} at which this first-order transition occurs is defined by two conditions: it must be a minimum of $\mathcal{L}$ and such that the value of $\mathcal{L}$ in that minimum is zero. Thus we can determine T^{**} as follows:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \eta} = \eta(at - 3w\eta + b\eta^2) = 0 \\ \mathcal{L}(\eta) = \eta^2 \left(\frac{a}{2}t - w\eta + \frac{b}{4}\eta^2\right) = 0 \end{cases} \quad \overset{\eta \neq 0}{\Longrightarrow} \quad \begin{cases} at - 3w\eta + b\eta^2 = 0 \\ \frac{a}{2}t - w\eta + \frac{b}{4}\eta^2 = 0 \end{cases}$$

Solving this system for t and η, we get:

$$\eta = \overline{\eta}^{**} = 2\frac{w}{b} \qquad\qquad t = t^{**} = 2\frac{w^2}{ab}$$

Since by definition $t = (T - \overline{T})/2$, we have:

$$T^{**} = \overline{T} + 4\frac{w^2}{ab}$$

Now, since the one we have found is a first order transition, there must be an associated latent heat. Landau theory allows us to determine it; in fact, if we determine the entropy density at the transition:

$$s = -\frac{\partial \mathcal{L}}{\partial T}_{\big|\overline{\eta}^{**}} = -4a\frac{w^2}{b^2}$$

then the latent heat absorbed during the process will be:

$$Q = -T^{**}s = 4aT^{**}\frac{w^2}{b^2}$$

Finally, we can also determine the susceptibility of the system. In the presence of an external field, the equation of state of the system is:

$$\frac{\partial \mathcal{L}}{\partial \eta} = 0 \qquad \Rightarrow \qquad at\eta - 3w\eta^2 + b\eta^3 = H$$

If we now derive both sides with respect to H, since $\chi_T = \partial\eta/\partial H$ we have:

$$at\chi_T - 6w\eta\chi_T - 3b\eta^2\chi_T = 1 \qquad \Rightarrow \qquad \chi_T = \frac{1}{at - 6w\eta + 3b\eta^2}$$

Multicritical points in Landau theory

It is possible for a system to have more "disarranging parameters" than the sole temperature T; let us call one such field Δ. In this case the phase diagram of the system becomes richer, with coexistence and critical lines that intersect in points called *multicritical points*; one of the
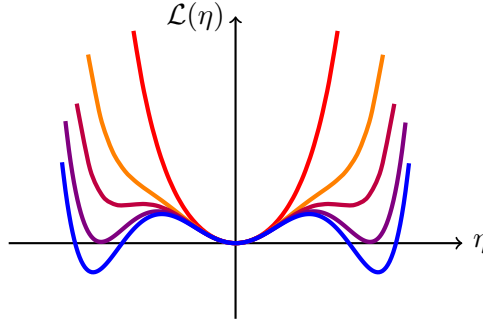
Figure 6.6: First-order transition with an even $\mathcal{L}$

most common examples of a multicritical point is the *tricritical point*, which divides a first-order transition line from a second-order one.

An example of a system of the type we are considering is the Blume-Emery-Griffiths model, which we have studied in 6.3.3. In that case the additional "disarranging field" was the concentration x of ^{3}He, and the tricritical point is the one we called (x_t, T_t).

Such a phenomenology can be obtained within Landau theory also with terms different from a simple cubic one; in particular, we can have first order phase transitions even when the system is invariant under parity, like in the case of the Ising model. In fact in that situation we required the coefficient of η^4 to be always positive, but if this is not true then $\mathcal{L}$ will be:

$$\mathcal{L} = \frac{a}{2}\eta^2 + \frac{b}{4}\eta^4 + \frac{c}{6}\eta^6 \tag{6.26}$$

where a, b and c are functions of T and Δ, and c must always be positive for the stability of the system (otherwise, like in the case previously considered, the minimization of $\mathcal{L}$ leads η to infinity).

Now, we know that if a changes sign and b is kept positive (which can be done varying the values of T and Δ in a way such that a goes to zero faster than b, depending of course on their explicit expressions) then a critical transition occurs since in this case $\eta = 0$ becomes a local maximum for $\mathcal{L}$, and it develops two new global minima. Therefore, the solution of the equation $a(T, \Delta) = 0$ will give a line of critical points in (T, Δ) plane.

However, if b becomes negative while a is still positive (which again can be done varying T and Δ so that b vanishes faster than a) then something rather different happens: as we can see from figure 6.6, in this case as b approaches zero $\mathcal{L}$ develops two new symmetric local minima at $\overline{\eta}_\pm$ (similarly to the case analysed before, with the difference that now the situation is perfectly symmetric since $\mathcal{L}$ is even) and they will become the new global minima as $\mathcal{L}(\overline{\eta}_\pm) = 0$, which happens when b changes sign: this way the equilibrium value of the order parameter change discontinuously from zero to a non-zero quantity so a first-order transition has indeed happened.

This means that when both a and b are null the system goes from exhibiting a continuous critical transition to a discontinuous first-order one; in other words, the *tricritical point* (T_c, Δ_c) can be determined from the solution of the equations $a(T, \Delta) = 0$ and $b(T, \Delta) = 0$.

To conclude let us consider again a system whose Landau free energy is given by (6.26) (to be concrete we still think about an Ising model), where $c > 0$ and a, b are in general functions of the reduced temperature t (and also of the other "disarranging" parameter, which we now neglect). We now want to show that we can understand how the phase diagram of the system is in (a, b) space, i.e. that we can draw where the phase transition lines are and so we are able to visually represents where the various phases of the system are in (a, b) plane.
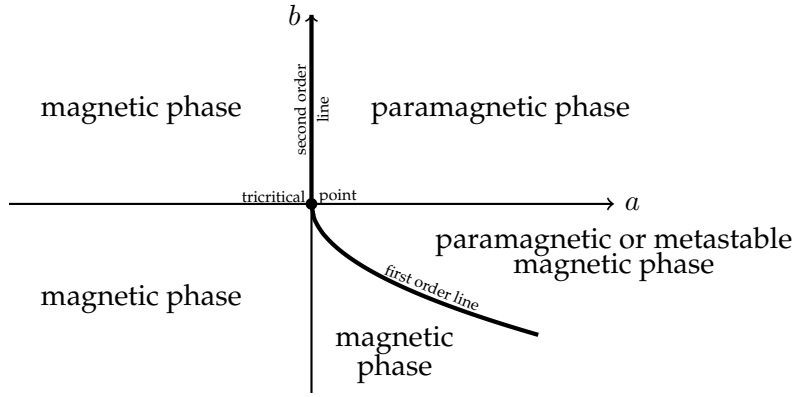
Figure 6.7: Phase diagram of the system in (a, b) space

First of all, we can note that when $a, b > 0$ the only minimum of $\mathcal{L}$ is $\overline{\eta} = 0$, so the system is in the paramagnetic phase. Furthermore if $a < 0$ and $b > 0$ the system is in the magnetic phase, and a second order transition has occurred; therefore we can surely say that the half-line $(a = 0, b > 0)$ is a second order transition line.

We must thus determine where the first order transition line lies in (a, b) space.

In order to do so, we first note that the extrema of $\mathcal{L}$ are given by:

$$0 = \frac{\partial \mathcal{L}}{\partial \eta} = \eta(a + b\eta^2 + c\eta^4) \qquad \Rightarrow \qquad \overline{\eta}_{\pm}^2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

(and of course they exist only when the temperature is such that $b^2 - 4ac > 0$) and since:

$$\frac{\partial^2 \mathcal{L}}{\partial \eta^2}\bigg|_{\overline{\eta}_{\pm}} = \pm \overline{\eta}_{\pm}^2 \cdot 2\sqrt{b^2 - 4ac}$$

we have that $\pm\overline{\eta}_+$ are maxima while $\pm\overline{\eta}_-$ are minima.

The first order transition happens when $\mathcal{L}(\pm\overline{\eta}_+) = \mathcal{L}(0) = 0$, so:

$$\frac{a}{2}\overline{\eta}_+^2 + \frac{b}{4}\overline{\eta}_+^4 + \frac{c}{6}\overline{\eta}_+^6 = 0 \qquad \Rightarrow \qquad \frac{a}{2} + \frac{b}{4}\overline{\eta}_+^2 + \frac{c}{6}\overline{\eta}_+^4 = 0$$

Now, from the condition $\partial \mathcal{L}/\partial \eta = 0$ we can express $\overline{\eta}_+^4$ as a function of $\overline{\eta}_+^2$, and we get $\overline{\eta}_+^4 = -(a + b\overline{\eta}_+^2)/c$. Substituting we get:

$$\overline{\eta}_+^2 = -4\frac{a}{b}$$

and substituting again in $\overline{\eta}_+^2 = (-b + \sqrt{b^2 - 4ac})/(2c)$ in the end we get:

$$b = -4\sqrt{\frac{ac}{3}}$$

so the first order transition line is a parabola in (a, b) plane (in particular it will lie in the fourth quadrant). In the end the situation is that shown in figure 6.7. As we can see the tricritical point of the system, being the point that divides the first-order from the second-order transition line, is the origin $(0, 0)$ of the parameter space.

We conclude by noting that in such situations strange things can happen; in particular we can show that if we move through the tricritical point along the a axis (thus keeping $b = 0$) the

critical exponents of the system change from the "trivial" ones predicted by mean field theories. In fact if $b = 0$ then the Landau free energy is $\mathcal{L} = a\eta^2/2 + c\eta^6/6$ and its minima are given by:

$$0 = \frac{\partial \mathcal{L}}{\partial \eta} = \eta(a + c\eta^4)$$

Excluding the case $\eta = 0$ and supposing that $a \sim t$ we get:

$$\eta \sim \left(-\frac{a}{c}\right)^{1/4} \sim (-t)^{1/4} \qquad \Rightarrow \qquad \beta = \frac{1}{4}$$

Furthermore, the state equation of the system in the presence of an external field h will be:

$$h = \frac{\partial \mathcal{L}}{\partial \eta} = \eta(a + c\eta^4)$$

and at the critical temperature $a = 0$, so that in the end:

$$h = cm^5 \qquad \Rightarrow \qquad \delta = 5$$

Analogous computations for the other critical exponents give $\gamma = 1$ and $\alpha = 1/2$.

We therefore see that the critical exponents do indeed change if the system passes through its tricritical point in the phase diagram.

6.5.5 Liquid crystals

We now proceed to study a particular physical system, *liquid crystals*, to which we will apply Landau theory of phase transitions. As we will see the symmetries of the system will allow the Landau free energy to include a cubic term in the order parameter (which we will properly define), and so we will be able to describe the first-order transition from an isotropic to a nematic phase (which we are now going to introduce).

What are liquid crystals?

Liquid crystals can be seen as an intermediate phase between a liquid and a solid: they are liquid like any other conventional fluid, but also have internal orientetional order like solid crystals. This orientational order provides them particular anisotropic properties from an optical, electric and magnetic point of view.
The most common structural characteristics of the molecules that constitute liquid crystals are the following:

- They have an elongated, anisotrpic shape

- Their axes can be considered rigid with good approximation

- They have strong electric dipoles or easily polarizable groups

Furthermore, it seems that the groups located at the extremities of a molecule are not relevant for the formation of phases.

The vast majority of the interesting phenomenology of liquid crystals concerns the geometry and dynamics of the preferred axis of orientation $\vec{n}(\vec{r})$, called *director* (and generally it is assumed to be unitary, i.e. $|\vec{n}(\vec{r})| = 1$). Since it just defines the direction of orientation, the direction of $\vec{n}$ is irrelevant, i.e. $\vec{n}$ and $-\vec{n}$ are physically equivalent.
There is a plethora of possible liquid crystal phases; the most common are:
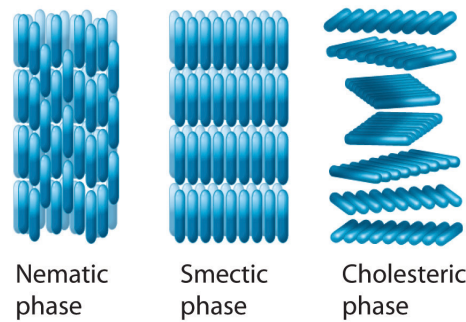
Figure 6.8: Graphical representation of the nematic, smectic and cholesteric phases of a liquid crystal (source)

nematic: This phase is characterized by a very strong long-ranged orientational order: the main axes of the molecules tend to orientate along a preferred direction, determined by the director. There is no long-ranged translational order of the molecular centers of mass, even if a short-ranged one can exist.

From an optical point of view, nematic liquid crystals are birefringent, i.e. they exhibit two different refractive indexes: one parallel to the director (called *ordinary refractive index*) and one orthogonal to it (*special refractive index*). These optical properties of the nematic phase are used to build devices like LCDs.

smectic: Also in this phase the molecules are aligned along a preferred direction, but contrarily to the nematic one this phase has also a spatial periodic order: the molecules are organised in layers. Furthermore, differently from nematic phases, smectic liquid crystals have non-uniform density and are generally more viscous.

cholesteric: It is similar to the nematic phase since it has a long-ranged orientational order, but the direction of $\vec{n}$ changes regularly in space; the typical configuration of a cholesteric liquid crystal has a director $\vec{n}(\vec{r})$ that rotates when $\vec{r}$ varies along a particular direction: for example, in a three-dimensional reference frame the molecules are orientated along the y direction in xy plane, but this direction roteates if z changes.

The structure of a cholesteric liquid crystal is characterised by the spatial distance along the torsion axis, called *pitch*, after which the director has rotated by an angle of 2π.[22] The pitch of the most common cholesteric liquid crystals is of the order of several hundred nanometers, so comparable with the wavelength of visible light; furthermore, it can also be very sensitive to changes in temperature, chemical composition, or external electromagnetic fields.

Note also that a nematic liquid crystal can be seen as a cholesteric one with infinite pitch; these two phases in fact are not independent from each other, and there is no real phase transition between them.

Definition of an order parameter for nematic liquid crystals

What we now want to do is to apply Landau theory to liquid crystals in order to study the transition from an isotropic to a nematic phase; therefore, we must define an order parameter for such a system. This is absolutely not trivial, and there are two ways to do it (even if as we will shortly see only one is really useful): a microscopic and a macroscopic one.

---

[22]Note that in reality the periodicity of the rotation is a half of the pitch, since the direction of $\vec{n}$ is irrelevant.

Microscopic approach Since in the nematic phase the important order is the orientational one, the director $\vec{n}(\vec{r})$ could be a good candidate as an order parameter; however, it only gives the mean orientation direction and no information about the *degree* of orientation, i.e. how much dispersion does the molecular orientational distribution have with respect to its mean value. Let us therefore look for something that also carries this information.

We can, for example, consider a small spherical portion of our system and try to "build" a probability distribution of the directions of orientation; this way we can compute its mean value (given by $\vec{n}$) but also higher moments like the standard deviation, which could be a good measure of the degree of orientation.

In order to estimate this distribution we can approximate every molecule with a rigid stick with cylindrical symmetry around a versor $\vec{a}$, which is therefore the orientation of a *single* molecule. If we suppose $\vec{n}$ to be parallel to the z axis, then we can identify $\vec{a}$ with the polar angles θ (colatitude) and ϕ (longitude). This way we can call $f(\theta, \phi)$ the orientational probability distribution, namely $f(\theta, \phi)d\Omega$ is the probability to find a molecule oriented along (θ, ϕ) in the solid angle $d\Omega = \sin\theta d\theta d\phi$. From the symmetries of the system, f has the following properties:

- Since the system has cylindrical symmetry f does not depend on φ, so it is only a function $f(\theta)$ of θ

- Since $\vec{n}$ and $-\vec{n}$ are physically equivalent, $f(\theta) = f(\pi - \theta)$

- $f(\theta)$ is even, again because the direction of $\vec{n}$ is irrelevant

Instead of a function, we could look for a *scalar* order parameter, since it is more easy to handle. The simplest possibility could be the mean value of the projection of the single molecule orientation $\vec{a}$ on the director $\vec{n}$:

$$\langle \vec{a} \cdot \vec{n} \rangle = \langle \cos\theta \rangle = \int f(\theta) \cos\theta d\Omega$$

However, this parameter is not really useful, since from the symmetry properties of f is null.

Now, in general a function g expressed in polar coordinates, i.e. which depends on θ and φ, can be expanded in multipoles, namely it can be written as a linear combination of spherical harmonics:

$$g(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} C_\ell^m Y_\ell^m(\theta, \phi)$$

where Y_ℓ^m are the spherical harmonics. The components with $\ell = 0$ are called *monopoles*, the ones with $\ell = 1$ *dipoles*, with $\ell = 2$ *quadrupoles*, $\ell = 3$ *octupoles* and so on. Remember that in general:

$$Y_\ell^m(\theta, \phi) = (-1)^m \sqrt{\frac{2\ell + 1}{4\pi} \frac{(\ell - m)!}{(\ell + m)!}} P_\ell^m(\cos\theta)e^{im\phi}$$

where $P_\ell^m(\cos\theta)$ are the Legendre polynomials.

Since our distribution $f(\theta)$ is even, we must discard any spherical harmonic involving $\sin\theta$, and among the monopoles also Y_1^0 must be discarded since it is proportional to $\cos\theta$ (and we have seen before that $\langle \cos\theta \rangle = 0$). Therefore, the first nontrivial multipole that we can use is the quadrupole Y_2^0:

$$Y_2^0(\theta, \phi) = \frac{1}{4}\sqrt{\frac{5}{\pi}} \left(3\cos^2\theta - 1\right)$$

with which we can define the parameter (neglecting the normalization constant of Y_2^0):

$$S := \frac{1}{2} \langle 3\cos^2\theta - 1 \rangle = \int f(\theta) \frac{3\cos^2\theta - 1}{2} d\Omega$$

This function is a good first approximation of the molecular degree of orientation, even if a more accurate description would require higher multipoles (i.e. higher even powers of $\cos\theta$). From its definition, we can deduce some properties of S:

– If $f(\theta)$ is sharply peaked around 0 and π, i.e. if nearly all the molecules are aligned, then $\cos\theta = \pm 1$ and $S = 1$

– If $f(\theta)$ is peaked around $\pi/2$, i.e. if the molecules lie on a plane orthogonal to the director and on this plane are randomnly oriented, $S = -1/2$

– If the molecular orientation si completely random, f is constantly equal to $1/2\pi$ and so:

$$S = \int \frac{1}{2\pi} \frac{3\cos^2\theta - 1}{2} d\Omega = \frac{3}{4\pi} \int \cos^2\theta d\Omega - 1$$

and since:

$$\int \cos^2\theta d\Omega = 2\pi \int_0^\pi \cos^2\theta \sin\theta d\theta = 2\pi \left(-\frac{1}{3}\cos^3\theta \right)_0^\pi = 2\pi \frac{2}{3}$$

then:

$$S = \frac{3}{4\pi} \frac{4\pi}{3} - 1 = 0$$

which makes sense since S is our candidate order parameter and in this case the system is in its disordered phase[23]

Therefore, once the director $\vec{n}$ and the parameter S are known, we can define the state of a nematic liquid crystal; for example, the director can be defined in terms of its polar angles θ and ϕ in a given reference frame:

$$\vec{n} = \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ \sin\theta \end{pmatrix}$$

and a theory for the nematic phase of the liquid crystal can be built using the variables:

$$\theta(\vec{r}, t) \qquad\qquad \phi(\vec{r}, t) \qquad\qquad S(\vec{r}, t)$$

(in fact, in general they depend on the point of the system considered, and also on time).

This approach, however, poses some problems: the fact that we are using polar coordinates leads to some mathematical problems, since when $\theta = \pi/2$ the angle ϕ is not well defined; furthermore, this approach is useful if we consider a uniaxial nematic liquid crystal but things start getting quite complicated if we try to describe biaxial nematic phases.

As we will now see a more useful approach consist in adopting a "macroscopic" perspective, which will however lead to the definition of a more complex (but also more powerful) order parameter.

---

[23]A remark: even if in principle is possible for the system to be in configurations where S is negative, it is more common that the equilibrium states of the liquid crystal have S positive, so we can think that as the temperature varies the order parameter goes from 0 (isotropic phase) to 1 (nematic phase).

Macroscopic approach From a macroscopic point of view we have already stated that an important difference between the disordered and nematic phases consists in the response functions when the liquid crystal is subjected to magnetic or electrical fields.

Supposing then that we have a liquid crystal subject to an external magnetic field $\vec{H}$, the magnetic response of the system will be measurable in terms of its magnetization $\vec{M}$, and in particular:

$$\vec{M} = \chi \vec{H}$$

where χ is the response function matrix, namely the magnetic susceptibility of the system. In componets we have:

$$M_\alpha = \chi_{\alpha\beta} H_\beta$$

where the indexes α, β stand for x, y or z.

If $\vec{H}$ is static then χ is symmetric, i.e. $\chi_{\alpha\beta} = \chi_{\beta\alpha}$; in the isotropic phase χ will also be diagonal, namely $\chi_{\alpha\beta} = \chi \delta_{\alpha\beta}$, while in the nematic phase:

$$\chi = \begin{pmatrix} \chi_\perp & 0 & 0 \\ 0 & \chi_\perp & 0 \\ 0 & 0 & \chi_\parallel \end{pmatrix}$$

where, as before, we have supposed that the director $\vec{n}$ is parallel to the z direction.

Therefore we could build an order parameter in terms of the susceptibility χ, and this parameter will necessarily have a tensorial nature[24] (since χ itself is in general a tensor), so it will not be a simple scalar like in the previous case.

Since we want our order parameter to vanish in the disordered phase, we can define it "removing" from χ its isotropic component. In other words, in components we can define:

$$Q_{\alpha\beta} := G \left(\chi_{\alpha\beta} - \frac{1}{3} \delta_{\alpha\beta} \operatorname{Tr} \chi \right)$$

where G is a constant. In this way Q is a good *tensorial order parameter*.

Let us note that its definition is completely general, and in fact it is useful also to describe other kinds of phases, not only the uniaxial nematic one.

Using what we have seen in the microscopic approach we can define a tensorial order parameter in terms of the director $\vec{n}$ and S:

$$Q_{\alpha\beta} := S \left(n_\alpha n_\beta - \frac{1}{3} \delta_{\alpha\beta} \right)$$

The advantage of this definition of the order parameter (which is the one we will use in the following) is that it also takes into account the degree of orientation and the mean direction.

We now just illustrate some properties of our newly defined order parameter.

By definition Q is symmetric and traceless, so in general way we can write it as:

$$Q = \begin{pmatrix} q_1 & q_2 & q_3 \\ q_2 & q_4 & q_5 \\ q_3 & q_5 & -q_1 - q_4 \end{pmatrix}$$

and from the expression of $\vec{n}$ in polar coordinates we readily have:

$$q_1 = S \left(\cos^2 \theta \cos^2 \phi - \frac{1}{3} \right) \qquad q_2 = S \cos^2 \theta \sin \phi \cos \phi \qquad q_3 = S \sin \theta \cos \theta \cos \phi$$

---

[24]We could have also foreseen this from the fact that for nematic phases the direction of $\vec{n}$ is irrelevant ($\vec{n}$ and $-\vec{n}$ ar equivalent), so the order parameter can't be a vector even if it has to indicate a preferred direction.

$$q_4 = S\left(\cos^2\theta\sin^2\phi - \frac{1}{3}\right) \qquad q_5 = S\cos\theta\sin\theta\sin\phi$$

and its diagonalization returns the eigenvalues $2S/3$, $-S/3$ and $-S/3$, so we can write:

$$Q_{\mathrm{D}} = \begin{pmatrix} \frac{2}{3}S & 0 & 0 \\ 0 & -\frac{1}{3}S & 0 \\ 0 & 0 & -\frac{1}{3}S \end{pmatrix} \tag{6.27}$$

where "D" stands for "diagnalized". In the more general case of a biaxial nematic phase, Q can again be diagonalized (it is still real and symmetric) in the form:

$$Q_{\mathrm{D}} = \begin{pmatrix} \frac{2}{3}S & 0 & 0 \\ 0 & -\frac{1}{3}(S+\eta) & 0 \\ 0 & 0 & -\frac{1}{3}(S-\eta) \end{pmatrix}$$

and we return to the uniaxial case when $\eta = 0$.

Landau-de Gennes theory for nematic liquid crystals

Since we now have a proper order parameter, we can formulate the Landau theory for the phase transitions of nematic liquid crystals (also called *Landau-de Gennes theory*). In particular we want to study the transition between the isotropic and nematic phase, and we call $T_{\mathrm{n.\text{-}i.}}$ the temperature at which it occurs.

As we have already stated, the Landau free energy $\mathcal{L}$ must be consistent with the symmetries of the system, so in this case it must be invariant under rotations.
Now, since Q transforms as a tensor under rotations and $\mathcal{L}$ must be a scalar, it will contain terms of the form $\mathrm{Tr}\,Q^p$; to the fourth order we will have (the linear term is absent because $\mathrm{Tr}\,Q = 0$ by definition):

$$\mathcal{L} = \frac{1}{2}A\,\mathrm{Tr}\,Q^2 + \frac{1}{3}B\,\mathrm{Tr}\,Q^3 + \frac{1}{4}C\left[\left(\mathrm{Tr}\,Q^2\right)^2 + \mathrm{Tr}\,Q^4\right]$$

In reality this expression, and in particular the quartic term, can be simplified: in fact it is a property (which we will not prove) of any $n \times n$ symmetric matrix that $\mathrm{Tr}\,Q^s$ with $s > n$ can be expressed as a polynomial of $\mathrm{Tr}\,Q^p$ with $p < n$, so in our case any $\mathrm{Tr}\,Q^s$ with $s \geq 4$ can be expressed in terms of $\mathrm{Tr}\,Q^2$ and $\mathrm{Tr}\,Q^3$ (we are automatically neglecting $\mathrm{Tr}\,Q$ since in our case it vanishes, but in general it must be considered).
Therefore, we can write the Landau free energy as:

$$\mathcal{L} = \frac{1}{2}A\,\mathrm{Tr}\,Q^2 + \frac{1}{3}B\,\mathrm{Tr}\,Q^3 + \frac{1}{4}C\left(\mathrm{Tr}\,Q^2\right)^2$$

or, in components:

$$\mathcal{L} = \frac{1}{2}AQ_{\alpha\beta}Q_{\beta\alpha} + \frac{1}{3}BQ_{\alpha\beta}Q_{\beta\gamma}Q_{\gamma\alpha} + \frac{1}{4}C(Q_{\alpha\beta}Q_{\beta\alpha})^2$$

Let us note that since our order parameter is a tensor its invariance under rotations does not exclude the possible existence of terms with odd powers of Q in $\mathcal{L}$, in particular the cubic one.
Now, from (6.27) in the case of a uniaxial nematic liquid crystal we have:

$$\mathrm{Tr}\,Q^2 = \frac{2}{3}S^2 \qquad\qquad \mathrm{Tr}\,Q^3 = \frac{2}{9}S^3 \qquad\qquad \left(\mathrm{Tr}\,Q^2\right)^2 = \frac{4}{9}S^4$$

so that, supposing that B and C do not depend on the temperature, while $A \propto T - \overline{T}$[25], we have:

$$\mathcal{L} = \frac{A}{3}(T - \overline{T})S^2 + \frac{2}{27}BS^3 + \frac{C}{9}S^4$$

This Landau free energy has exactly the same form of the one we studied in 6.5.4, with the substitutions:

$$a = \frac{2}{3}A \qquad\qquad -w = \frac{2}{27}B \qquad\qquad b = \frac{4}{9}C$$

Applying the results we have already found, we will have that the first-order transitions between the isotropic and nematic phases occurs at the temperature:

$$T_{\text{n.-i.}} = \overline{T} + 4\frac{w^2}{ab} = \overline{T} + \frac{2B^2}{27AC}$$

while the value of the scalar order parameter S at the transition is:

$$S_{\text{n.-i.}} = 2\frac{w}{b} = -\frac{B}{3C}$$

From what we have seen in the general case, there will also be a temperature $\tilde{T}$ such that for $T_{\text{n.-i.}} < T < \tilde{T}$ the Landau free energy $\mathcal{L}$ will have a local minimum different from $S = 0$: this corresponds to the nematic phase, which in this range of temperatures is metastable. This minimum occurs for the following value of the order parameter S:

$$S = \frac{3w}{2b} + \sqrt{\frac{9w^2}{4b^2} - \frac{at}{b}} = -\frac{B}{4C} + \sqrt{\frac{B^2}{16C^2} + \frac{T - \overline{T}}{9}\frac{A}{B}}$$

(note that for $T = T_{\text{n.-i.}}$ we have $S = S_{\text{n.-i.}}$).

6.6 Ginzburg-Landau theory

6.6.1 Introduction: Ginzburg criterion

As we have seen, the main assumption (and the most important problem) of mean field theories is that the fluctuations of the order parameter are completely neglected in the computation of the partition function; this approximation breaks down in the neighbourhoods of critical points, where as we have seen in 5.1.2 the correlation length becomes comparable with the size of the system.
What we would now like to do is to include these fluctuations in a mean field theoretical framework; this will lead to the so called *Ginzburg-Landau* theory.

As a first approach we can try to estimate how big is the error we make in mean field theories neglecting the fluctuations of the order parameter near a critical point, so that we can understand under which conditions mean field theories are actually good approximations.
To make things explicit, let us use the Ising model as a base for our considerations.
We have seen in 6.1 that the Weiss mean field theory for the Ising model is based on the assumption that $\langle S_i S_j \rangle = \langle S_i \rangle \langle S_j \rangle$, i.e. that the spins are statistically independent; therefore, a possible estimate of the error made with this assumption can be:

$$E_{ij} = \frac{|\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle|}{\langle S_i \rangle \langle S_j \rangle} \tag{6.28}$$

---

[25] As we have seen before in the general treatment of first-order phase transition within Landau theory, the temperature $\overline{T}$ has no particular physical relevance.

The numerator of E_{ij} is, by definition (see also 5.1.2), the two-point connected correlation function:

$$G_c(i,j) = \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle$$

If we neglect the fluctuations of the order parameter (which in our case is of course $\langle S_i \rangle$) we see that G_c is constantly zero: therefore, in order to have non-null correlation functions we need that the system exhibits some kind of inhomogeneity, not necessarily due to thermal fluctuations. In fact, the connected correlation function describes not only the spatial extension of the fluctuations of the order parameter, but also the way it varies in space in response to an external inhomogeneous field. Let us see this explicitly.

We know that from the partition function of the Ising model in an inhomogeneous external field $\vec{H}$, i.e.:

$$Z = \operatorname{Tr} e^{\beta \left(J \sum_{\langle ij \rangle} S_i S_j + \sum_i H_i S_i \right)}$$

we have:

$$\langle S_i \rangle = \frac{1}{\beta Z} \frac{\partial Z}{\partial H_i} = \frac{1}{\beta} \frac{\partial \ln Z}{\partial H_i}$$

Similarly:

$$\langle S_i S_j \rangle = \frac{1}{\beta^2 Z} \frac{\partial^2 Z}{\partial H_i \partial H_j}$$

and thus:

$$G_c(i,j) = \frac{1}{\beta^2 Z} \frac{\partial^2 Z}{\partial H_i \partial H_j} - \frac{1}{\beta^2 Z^2} \frac{\partial Z}{\partial H_i} \frac{\partial Z}{\partial H_j} = \frac{1}{\beta^2} \frac{\partial^2 \ln Z}{\partial H_i \partial H_j}$$

Therefore:

$$\frac{\partial \langle S_i \rangle}{\partial H_j} = \frac{\partial}{\partial H_j} \left(\frac{1}{\beta} \frac{\partial \ln Z}{\partial H_i} \right) = \frac{1}{\beta} \frac{\partial^2 \ln Z}{\partial H_i \partial H_j} = \beta G_c(i,j)$$

so $G_c(i,j)$ can indeed be seen as a response function.

If we now call:

$$M = \sum_i \langle S_i \rangle$$

we will have:

$$\frac{\partial M}{\partial H_j} = \sum_i \frac{\partial \langle S_i \rangle}{\partial H_j} = \beta \sum_i G_c(i,j)$$

If our system is invariant under translations and subject to a uniform field, then:

$$\frac{\partial M}{\partial H} = \sum_j \frac{\partial M}{\partial H_j} \frac{\partial H_j}{\partial H} = \beta \sum_{i,j} G_c(i,j)$$

and since $\chi_T = \partial M / \partial H$ we get:

$$\chi_T = \beta \sum_{i,j} G_c(i,j) = \beta \sum_{i,j} \left(\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle \right)$$

which is a version of the *fluctuation-dissipation theorem*.

Let us now try to understand when the error (6.28) done in mean field theories is negligible. Now, in general terms if we formulate a mean field theory for a system we will make the error (6.28) in the region where correlations are relevant, namely if $|\vec{r}|$ is the distance between two points of the system the error is made for $|\vec{r}| \leq \xi$, with ξ the correlation length. Supposing $T < T_c$, so that the order parameter η is non null, then:

$$E_{\text{TOT}} = \frac{\int_{|\vec{r}| \leq \xi} G_c(\vec{r}) d^d \vec{r}}{\int_{|\vec{r}| \leq \xi} \eta^2(\vec{r}) d^d \vec{r}}$$

where we have called d the dimensionality of our system.

Therefore, our mean field theory will be a good approximation if $E_{\text{TOT}} \ll 1$, i.e.:

$$\frac{\int_{|\vec{r}| \leq \xi} G_c(\vec{r}) d^d\vec{r}}{\int_{|\vec{r}| \leq \xi} \eta^2(\vec{r}) d^d\vec{r}} \ll 1$$

known as *Ginzburg criterion*.

In order to express it in a useful fashion, let us write it in terms of critical exponents; using also the version we have just found of the fluctuation-dissipation theorem we get (supposing our system is continuous)[26]:

$$\int_{|\vec{r}| \leq \xi} G_c(\vec{r}) d^d\vec{r} \sim k_B T_c \chi_T \sim t^{-\gamma} \qquad \int_{|\vec{r}| \leq \xi} \eta^2(\vec{r}) d^d\vec{r} \sim \xi^d |t|^{2\beta} \sim t^{2\beta - \nu d}$$

where we have used the definitions of critical exponents (see 1.5). Therefore, the Ginzburg criterion can be reformulated as:

$$E_{\text{TOT}} \sim t^{-\gamma + \nu d - 2\beta} \ll 1$$

and in the limit $t \to 0$ this is possible only if $-\gamma + \nu d - 2\beta > 0$, i.e.:

$$d > \frac{2\beta + \gamma}{\nu} := d_c$$

This means that Ginzburg criterion allows us to determine the *upper critical dimension d_c* of a system, namely the dimension above which mean field theories are good approximations[27]; if $d < d_c$ fluctuations become too relevant and mean field theories don't work.

Let us note that since it depends on the critical exponents, the upper critical dimension d_c ultimately depends on the universality class of the system considered; furthermore, in order to actually be able to compute d_c we must generalize Landau theory to systems with spatial inhomogeneities so that we are able to compute the critical exponent ν.

6.6.2 Functional partition function and coarse graining

A possible way to overcome the limitations of mean field theories (i.e. their inability to discuss the local properties of the order parameter of a given system, for example its fluctuations) can be the following: we could regard the profile of the order parameter $\eta(\vec{r})$ to be the "degree of freedom" of our system and compute the partition function as a *functional integral*; in other words from the microscopic configuration $\mathcal{C}$ of our system we can obtain $\eta(\vec{r})$ with a *coarse graining* procedure (we will immediately see what we mean by this) and then determine Z as a trace over all the possible configurations of our system, i.e. over all the possible forms of $\eta(\vec{r})$:

$$Z = \text{Tr}\, e^{-\beta \mathcal{H}(\mathcal{C})} = \int \mathcal{D}[\eta(\vec{r})] \sum_{\mathcal{C}}' e^{-\beta \mathcal{H}(\mathcal{C})} = \int \mathcal{D}[\eta(\vec{r})] e^{-\beta \mathcal{H}_{\text{eff.}}[\eta(\vec{r})]}$$

where with $\sum_{\mathcal{C}}'$ we mean a sum over all the possible microscopic configurations $\mathcal{C}$ compatible with the order parameter profile $\eta(\vec{r})$, and the last step implicitly defines the effective Hamiltonian $\mathcal{H}_{\text{eff.}}$:

$$e^{-\beta \mathcal{H}_{\text{eff.}}} := \sum_{\mathcal{C}}' e^{-\beta \mathcal{H}(\mathcal{C})}$$

---

[26] For the origin of the first equation, see what we have stated above and also equation (5.2).

[27] This also mean that above the upper critical dimension the critical exponents determined with mean field theories are exact, or at least in good agreement with experiments.

We therefore must understand how to determine $\eta(\vec{r})$; the idea of coarse graining procedures is the following: for a given microscopic configuration $\mathcal{C}$ we average the order parameter η over sufficiently wide "blocks", i.e. portions of the system with linear dimension ℓ much greater than its microscopic scale, which we call a (in the case of the Ising model, for example, a can be taken as the lattice constant), but still microscopic and in particular much smaller than the correlation length ξ, so that the order parameter is uniform in every block. In other words, coarse graining a system means dividing it into cells of linear dimension ℓ, with ℓ such that:

$$a \ll \ell \ll \xi(T) \leq L$$

(L being the linear dimension of our system) and averaging the order parameter η in every cell. This way we can obtain an expression for $\eta(\vec{r})$ (since ℓ is anyway microscopic with respect to the size of the system, so we can regard $\vec{r}$ as a continuous variable).

6.6.3   Coarse graining procedure for the Ising model

To make things clearer, let us see how the coarse graining procedure works for the Ising model. If we call $m_i = \langle S_i \rangle$ the local magnetization at the i-th site and d the dimensionality of the system, every "block" will have volume ℓ^d; we define for every block of the system centered in $\vec{r}$ the coarse grained magnetization as:

$$m_\ell(\vec{r}) = \frac{1}{N_\ell} \sum_{i \in \vec{r}} m_i$$

where $N_\ell = (\ell/a)^d$ is the number of spins (degrees of freedom in general) which belong to the block centered in $\vec{r}$; this definition is reasonable as long as N_ℓ is large. Since it has been built as an average, m_ℓ does not fluctuate much on microscopic scales but varies smoothly in space.
Of course, in general we need to specify ℓ in order to determine m_ℓ, but the coarse graining procedure we are applying will be useful only if the final results are independent of ℓ (at least at the spatial scales considered).

We now must express the partition function in terms of $m_\ell(\vec{r})$, and as we have stated before:

$$Z = \int \mathcal{D}[m_\ell(\vec{r})] e^{-\beta \mathcal{H}_{\text{eff.}}[m_\ell(\vec{r})]} \tag{6.29}$$

so we must now compute $\mathcal{H}_{\text{eff.}}$.
Since we now have a system made up of "blocks" this effective Hamiltonian will be composed of two parts: a bulk component relative to the single blocks and an interface component relative to the interaction between the blocks; let us consider them individually.

bulk component: Suppose that every block of volume ℓ^d is separate from the rest of the system; inside every one of them the magnetization is uniform (since the linear dimension of the blocks is much smaller than the correlation length), so we can use Landau theory for uniform systems. In the case of the Ising model, it led to (equation (6.24)):

$$\mathcal{L} = \frac{\overline{a}}{2} t m^2 + \frac{\overline{b}}{4} m^4$$

The total bulk energy is thus obtained summing over all the blocks:

$$\beta \mathcal{H}_{\text{eff.}}^{\text{bulk}} = \sum_{\vec{r}} \left(\frac{\overline{a}}{2} t m^2(\vec{r}) + \frac{\overline{b}}{4} m^4(\vec{r}) \right)$$

interaction component: We now must take into account the fact that adjacent blocks do interact. In particular since as we have stated m does not vary much on microscopic scales, the interaction between the blocks must be such that strong variations of magnetization between neighbouring blocks is energetically unfavourable. If we call $\vec{\delta}$ a vector of magnitude ℓ that points from one block to a neighbouring one, the most simple analytic expression that we can guess for such a term can be a harmonic one[28]:

$$\beta\mathcal{H}_{\text{eff.}}^{\text{int.}} = \sum_{\vec{r}}\sum_{\vec{\delta}} \frac{\overline{k}}{2}\left(m(\vec{r}) - m(\vec{r}+\vec{\delta})\right)^2$$

(the factor $1/2$ multiplying $\overline{k}$, just like the numeric factors multiplying $\overline{a}$ and $\overline{b}$, have been inserted for future convenience). We can also think of this as a first approximation of a general interaction between the blocks, namely as the first terms of a Taylor expansion of the *real* interaction energy.

Now, since the linear dimension of the blocks ℓ is much smaller than the characteristic length L of the system we can treat $\vec{r}$ as a continuous variable and thus substitute the sum over $\vec{r}$ with an integral:

$$\sum_{\vec{r}} \longrightarrow \frac{1}{\ell^d}\int d^d\vec{r}$$

(while the sum over $\vec{\delta}$ remains a sum, since for every $\vec{r}$ there is only a *finite* number of nearest neighbours). Therefore:

$$\beta\mathcal{H}_{\text{eff.}}^{\text{bulk}} = \frac{1}{\ell^d}\int\left(\frac{\overline{a}}{2}tm^2(\vec{r}) + \frac{\overline{b}}{4}m^4(\vec{r})\right)d^d\vec{r} \qquad \beta\mathcal{H}_{\text{eff.}}^{\text{int.}} = \frac{1}{\ell^d}\int\sum_{\vec{\delta}}\frac{\overline{k}}{2}\left(m(\vec{r}) - m(\vec{r}+\vec{\delta})\right)^2 d^d\vec{r}$$

Keeping in mind that $|\vec{\delta}| = \ell$, the interaction term can be rewritten in terms of $\vec{\nabla}m$:

$$\frac{1}{\ell^d}\int\sum_{\vec{\delta}}\frac{\overline{k}}{2}\left(m(\vec{r}) - m(\vec{r}+\vec{\delta})\right)^2 d^d\vec{r} = \frac{1}{\ell^{d-2}}\int\frac{\overline{k}}{2}\sum_{\vec{\delta}}\left(\frac{m(\vec{r}) - m(\vec{r}+\vec{\delta})}{\ell}\right)^2 d^d\vec{r} =$$

$$= \int\frac{\overline{k}}{2\ell^{2-d}}\sum_{\vec{\delta}}\left(\frac{\partial m}{\partial r_\delta}\right)^2 d^d\vec{r} = \int\frac{\overline{k}}{2\ell^{2-d}}\left(\vec{\nabla}m\right)^2 d^d\vec{r}$$

where we have called r_δ the components of $\vec{\delta}$.
Thus, if we now define for the sake of simplicity:

$$a := \frac{\overline{a}}{\ell^d} \qquad\qquad b := \frac{\overline{b}}{\ell^d} \qquad\qquad k := \frac{\overline{k}}{\ell^{2-d}}$$

we will have:

$$\beta\mathcal{H}_{\text{eff.}} = \int\left[\frac{a}{2}tm^2(\vec{r}) + \frac{b}{4}m^4(\vec{r}) + \frac{k}{2}\left(\vec{\nabla}m(\vec{r})\right)^2\right]d^d\vec{r}$$

Therefore, the (functional) partition function (6.29) of the system will be:

$$Z = \int e^{-\int\left(\frac{a}{2}tm^2 + \frac{b}{4}m^4 + \frac{k}{2}\left(\vec{\nabla}m\right)^2\right)d^d\vec{r}}\mathcal{D}[m_\ell(\vec{r})]$$

Let us now make a couple of considerations:

---

[28]Sometimes this approximation is called *elastic free energy*.

- If $m(\vec{r}) = m = $ const. the energy of the system has the same structure of the one used in Landau theory

- The term proportional to $(\vec{\nabla}m)^2$ is completely new but we could have introduced it intuitively to a Landau-like mean field functional, since the introduction of spatial variations in the order parameter has an energetic cost which must depend on how it varies in space, i.e. it depends on the gradient of m. In particular, it must involve $(\vec{\nabla}m)^2$ because of the symmetries of the model: since the system is isotropic and $\mathbb{Z}_2$-invariant, we must use combinations of derivatives that are invariant under rotations and parity, and $(\vec{\nabla}m)^2$ is the simplest of them[29].

If there is also an external magnetic field $\vec{h}(\vec{r}) = \beta\vec{H}(\vec{r})$, we must add to the Hamiltonian the term:

$$ - \int \vec{h}(\vec{r}) \cdot \vec{m}(\vec{r}) d^d\vec{r} $$

so that the partition function becomes:

$$ Z = \int e^{- \int \left(\frac{a}{2}tm^2 + \frac{b}{4}m^4 + \frac{k}{2}\left(\vec{\nabla}m\right)^2 - hm \right) d^d\vec{r}} \mathcal{D}[m_\ell(\vec{r})] \tag{6.30} $$

which is a functional of $m(\vec{r})$ and $h(\vec{r})$. As usual, all the thermodynamics of the system can be obtained from Z, provided that now we take *functional derivatives* instead of usual derivatives.

Saddle point approximation: Landau theory

We can now compute (6.30), as a first approach, using the saddle point approximation (see appendix B); as we will see this will reproduce a Landau-like mean field theory which will also take into account the presence of inhomogeneities. In particular thanks to the new term involving $\vec{\nabla}m$ we will be able to compute the fluctuation correlation function[30] and so also to determine the critical exponents η and ν.

Therefore we approximate (6.30) with the leading term of the integral, i.e. we must determine the function m_0 that maximizes the exponent, namely minimizes:

$$ \mathcal{L}[m(\vec{r})] := \beta\mathcal{H}_{\text{eff.}} - \int hm\, d^d\vec{r} = \int \left(\frac{a}{2}tm^2 + \frac{b}{4}m^4 + \frac{k}{2}\left(\vec{\nabla}m\right)^2 - hm \right) d^d\vec{r} \tag{6.31} $$

and then compute Z as:

$$ Z \approx e^{-\mathcal{L}[m_0(\vec{r})]} $$

where m_0 is determined imposing the stationarity of the functional $\mathcal{L}$ with respect to m:

$$ \frac{\delta\mathcal{L}}{\delta m}_{|m_0} = 0 $$

---

[29]At this point we could wonder why the interaction part of the Hamiltonian does not contain other terms, like $m\nabla^2 m$: in fact this is in principle perfectly acceptable since is of second order in m and is invariant under rotations and parity $m \to -m$. However, we have that:

$$ m\nabla^2 m = \vec{\nabla} \cdot (m\vec{\nabla}m) - (\vec{\nabla}m)^2 $$

and so when we integrate over $\vec{r}$, the first term vanishes in the thermodynamic limit supposing that the magnetization or its gradient goes to zero sufficiently rapidly as $|\vec{r}| \to \infty$. Therefore, we are left only with $(\vec{\nabla}m)^2$: the two terms are perfectly equivalent.

[30]We will do our computations on the Ising model, as usual.

This leads to the state equation of the system:

$$h(\vec{r}) = \frac{\delta\tilde{\mathcal{H}}}{\delta m}$$

where we have defined $\tilde{\mathcal{H}} := \beta\mathcal{H}_{\text{eff.}}$ for brevity.
If we now call $\mathfrak{h}$ the integrand of $\tilde{\mathcal{H}}$, since[31]:

$$\frac{\delta\tilde{\mathcal{H}}}{\delta m} = \frac{\partial\mathfrak{h}}{\partial m} - \vec{\nabla}\frac{\partial\mathfrak{h}}{\partial(\vec{\nabla}m)}$$

we have:

$$h(\vec{r}) = -k\nabla^2 m(\vec{r}) + atm(\vec{r}) + bm^3(\vec{r}) \tag{6.32}$$

Note that if $h = $ const. and $m_0 = $ const. (i.e. the system is uniform) we get the state equation that we have obtained with the Landau theory (see 6.5.3):

$$h = atm_0 + bm_0^3$$

Correlation function in the saddle point approximation We can now proceed to compute the correlation function within our approximations.
In order to do that, we take the (functional) derivative of (6.32) with respect to $h(\vec{r}\,')$, so that $\chi_T = \delta m/\delta h$ appears:

$$\delta(\vec{r} - \vec{r}\,') = \left[-k\nabla^2 + at + 3bm^2\right]\chi_T(\vec{r} - \vec{r}\,')$$

Now, from fluctuation-dissipation theorem we know that:

$$G(\vec{r} - \vec{r}\,') = k_B T \chi_T(\vec{r} - \vec{r}\,')$$

so that:

$$\beta\left[-k\nabla^2 + at + 3bm^2\right]G(\vec{r} - \vec{r}\,') = \delta(\vec{r} - \vec{r}\,') \tag{6.33}$$

Note that this means that the correlation function $G(\vec{r} - \vec{r}\,')$ can be interpreted as the Green's function of the operator written between the square brackets.

In case of translationally invariant (i.e. uniform) systems, m is constant and equal to the equilibrium values given by Landau theory (see 6.5.3); in particular, depending on the sign of t there are two possible situations:

---

[31]This follows from the integral form of $\tilde{\mathcal{H}}$. In fact, if in general a functional is of the form:

$$F[\eta] = \int f(\eta, \vec{\nabla}\eta)d^d\vec{r}$$

then from the definition of functional derivative we have:

$$\int\frac{\delta F}{\delta\eta}\varphi(\vec{r})d^d\vec{r} = \frac{d}{d\varepsilon}\left[\int f\left(\eta + \varepsilon\varphi, \vec{\nabla}(\eta + \varepsilon\varphi)\right)d^d\vec{r}\right]_{|\varepsilon=0} = \int\left(\frac{\partial f}{\partial\eta}\varphi + \frac{\partial f}{\partial\vec{\nabla}\eta}\cdot\vec{\nabla}\varphi\right)d^d\vec{r}$$

where $\varphi(\vec{r})$ is an arbitrary function that vanishes on the boundary of integration. Integrating by parts we get:

$$\int\frac{\delta F}{\delta\eta}\varphi(\vec{r})d^d\vec{r} = \int\left(\frac{\partial f}{\partial\eta} - \vec{\nabla}\cdot\frac{\partial f}{\partial\vec{\nabla}\eta}\right)\varphi(\vec{r})d^d\vec{r}$$

and so finally:

$$\frac{\delta F}{\delta\eta} = \frac{\partial f}{\partial\eta} - \vec{\nabla}\cdot\frac{\partial f}{\partial\vec{\nabla}\eta}$$

$t > 0$: In this case $m(\vec{r}) = m_0 = 0$, so (6.33) becomes:

$$\left(-k\nabla^2 + at\right) G(\vec{r} - \vec{r}') = k_B T \delta(\vec{r} - \vec{r}')$$

Defining:

$$\xi_> := \sqrt{\frac{k}{at}}$$

this can be rewritten as:

$$\left(-\nabla^2 + \xi_>^{-2}\right) G(\vec{r} - \vec{r}') = \frac{k_B T}{k} \delta(\vec{r} - \vec{r}')$$

$t < 0$: In this case the magnetization is:

$$m_0 = \pm\sqrt{-\frac{at}{b}}$$

so (6.33) becomes:

$$\left(-k\nabla^2 - 2at\right) G(\vec{r} - \vec{r}') = k_B T \delta(\vec{r} - \vec{r}')$$

This can be rewritten in a form similar to the previous case; in fact, if we define:

$$\xi_< = \sqrt{-\frac{k}{2at}}$$

we get:

$$\left(-\nabla^2 + \xi_<^{-2}\right) G(\vec{r} - \vec{r}') = \frac{k_B T}{k} \delta(\vec{r} - \vec{r}')$$

We will shortly see that $\xi_>$ and $\xi_<$ are the expressions of the correlation length for $T > T_c$ and $T < T_c$, respectively. We can therefore see that in both cases we get:

$$\nu = \frac{1}{2}$$

Thus, for both the cases $t > 0$ and $t < 0$ the correlation function $G(\vec{r} - \vec{r}')$ can be obtained by solving the differential equation:

$$\left(-\nabla^2 + \xi^{-2}\right) G(\vec{r} - \vec{r}') = \frac{k_B T}{k} \delta(\vec{r} - \vec{r}') \tag{6.34}$$

which can be done with Fourier transforms.
If we use the following conventions for the Fourier transform:

$$\hat{f}(\vec{q}) = \int f(\vec{r} - \vec{r}')e^{-i\vec{q}\cdot(\vec{r}-\vec{r}')}d^d(\vec{r} - \vec{r}') \qquad\qquad f(\vec{r} - \vec{r}') = \frac{1}{(2\pi)^d}\int \hat{f}(\vec{q})e^{i\vec{q}\cdot(\vec{r}-\vec{r}')}d^d\vec{q}$$

then transforming both sides of (6.34) we get:

$$(q^2 + \xi^{-2})\hat{G}(\vec{q}) = \frac{k_B T}{k} \qquad \Rightarrow \qquad \hat{G}(\vec{q}) = \frac{k_B T}{k}\frac{1}{q^2 + \xi^{-2}}$$

where $q = |\vec{q}|$[32]. From this last equation we can also foresee that when $T = T_c$, since $\xi = \infty$ we have $\hat{G}(\vec{q}) \sim q^{-2}$ and so $G(\vec{r}) \sim 1/|\vec{r}|^{2-d}$, from which we have that the critical exponent η is

---

[32]In the following, for the sake of simplicity we will indicate the magnitude of a vector simply removing the arrow sign.

null (we will see that explicitly once we have computed G).

Therefore, renaming $\vec{x} = \vec{r} - \vec{r}\,'$ we can now determine $G(\vec{x})$ with the Fourier antitransform:

$$G(\vec{x}) = \int \frac{d^d\vec{q}}{(2\pi)^d} \frac{e^{i\vec{q}\cdot\vec{x}}}{q^2 + \xi^{-2}} \tag{6.35}$$

This integral is a bit tedious to compute, and in general its result depends strongly on the dimensionality d of the system; the general approach used to solve it is to shift to spherical coordinates in $\mathbb{R}^d$ and then complex integration for the remaining part, which involves $|\vec{q}|$.

In order to do some explicit computations, let us consider the case $d = 3$; we will then have:

$$G(\vec{x}) = \int \frac{d^3\vec{q}}{(2\pi)^3} \frac{e^{i\vec{q}\cdot\vec{x}}}{q^2 + \xi^{-2}} = \frac{1}{(2\pi)^3} \int_0^\infty \frac{q^2}{q^2 + \xi^{-2}} dq \int_{-1}^1 e^{iqx\cos\theta} d(\cos\theta) \int_0^{2\pi} d\varphi$$

Therefore:

$$G(\vec{x}) = \frac{1}{(2\pi)^2} \int_0^\infty \frac{q^2}{q^2 + \xi^{-2}} dq \left(\frac{e^{iqx}}{iqx}\right)_{-1}^1 = \frac{1}{(2\pi)^2 x} \int_0^\infty \frac{q\sin(qx)}{q^2 + \xi^{-2}} dq$$

This last integral can be computed, using the residue theorem, extending it to the complex plane:

$$\int_0^\infty \frac{q\sin(qx)}{q^2 + \xi^{-2}} dq = \frac{1}{2} \int_{-\infty}^{+\infty} \frac{q\sin(qx)}{q^2 + \xi^{-2}} dq = \frac{1}{2} \operatorname{Im} \int \frac{ze^{izx}}{z^2 + \xi^{-2}} dz = \frac{1}{2} \operatorname{Im} \int \frac{ze^{izx}}{(z + i\xi^{-1})(z - i\xi^{-1})} dz$$

Now, the integrand exhibits two poles at $\pm i\xi^{-1}$; we choose as the contour of integration Γ the upper semicircle in the complex plane, which contains only the pole at $i\xi^{-1}$ and so using the residue theorem we will have:

$$\frac{1}{2} \operatorname{Im} \int_\Gamma \frac{ze^{izx}}{(z + i\xi^{-1})(z - i\xi^{-1})} dz = \frac{1}{2} \operatorname{Im} \left[2\pi i \frac{i\xi^{-1} e^{-x/\xi}}{2i\xi^{-1}}\right] = \frac{1}{2} \operatorname{Im} \left[2\pi i \frac{e^{-x/\xi}}{2}\right] = \frac{\pi}{2} e^{-x/\xi}$$

Therefore, in the end we have:

$$G(\vec{x}) = \frac{1}{8\pi} \frac{e^{-|\vec{x}|/\xi}}{|\vec{x}|}$$

We see now clearly that the correlation function has indeed an exponential behaviour (as we have stated also in 5.1.2) and that ξ is really the correlation length; furthermore, $G(\vec{x}) \sim 1/|\vec{x}|$ and from the definition of the exponent η we have $G(\vec{x}) \sim 1/|\vec{x}|^{d-2+\eta}$, so since $d = 3$ we indeed have $\eta = 0$.

Therefore, we have seen that for the Ising model $\nu = 1/2$. If we also consider the values of the other critical exponents we see that the upper critical dimension for this model is $d_c = 4$. In other words, mean field theories are actually good approximations for the Ising model if $d \geq 4$. We will later see some other confirmations of this fact, particularly in 6.6.3.

Gaussian approximation

Until now even if we have introduced Ginzburg-Landau theory we are still neglecting the effects of the fluctuations since we are regarding the mean field theory approximation for non-homogeneous systems as a saddle point approximation of a more general theory; in other words, since we are approximating Z as $e^{-\mathcal{L}[m_0(\vec{r})]}$ we are still regarding the magnetization m as non fluctuating over the system. In order to include the fluctuations we must do more and go further the simple saddle point approximation.

The simplest way we can include fluctuations in our description is expanding (6.30) around

the stationary solution and keeping only quadratic terms; this means that we are considering fluctuations that follow a normal distribution around the stationary value. The important thing to note, however, is that in this approximation these fluctuations *are independent*, i.e. they do not interact with each other[33].

As we will see, with this assumption the values of some critical exponents will differ from the "usual" ones predicted by mean field theories.

Let us apply this approximation from easier cases to more complex ones (and finally to the one we are interested in).

Gaussian approximation for one degree of freedom   Let us consider a system with a single degree of freedom q, and call $\mathcal{H}(q)$ its Hamiltonian. Supposing that q_0 is a minimum for $\mathcal{H}$, i.e. $\partial \mathcal{H} / \partial q_{|q_0} = 0$, expanding $\mathcal{H}$ around q_0 we get:

$$\mathcal{H}(q) = \mathcal{H}(q_0) + \frac{1}{2} \frac{\partial^2 \mathcal{H}}{\partial q^2}\bigg|_{q_0} \delta q^2 + O(\delta q^3)$$

where $\delta q = q - q_0$ is the fluctuation of q around its stationary value q_0. The Boltzmann factor needed to compute the partition function will therefore be:

$$e^{-\beta \mathcal{H}(q)} \sim e^{-\beta \mathcal{H}(q_0) - \frac{1}{2} \frac{\delta q^2}{\lambda^2}}$$

where for the sake of simplicity we have defined:

$$\frac{1}{\lambda^2} := \beta \frac{\partial^2 \mathcal{H}}{\partial q^2}\bigg|_{q_0}$$

With this approximation, the partition function of the system results:

$$Z = \int_{-\infty}^{+\infty} e^{-\beta \mathcal{H}(q)} \sim e^{-\beta \mathcal{H}(q_0)} \int_{-\infty}^{+\infty} e^{-\frac{\delta q^2}{2\lambda^2}} dq$$

The last one is a gaussian integral, which readily gives:

$$\int_{-\infty}^{+\infty} e^{-\frac{(q-q_0)^2}{2\lambda^2}} dq = \sqrt{2\pi \lambda^2}$$

Therefore:

$$Z = e^{-\beta \mathcal{H}(q_0)} \sqrt{2\pi \lambda^2}$$

and since $Z = e^{-\beta F}$ the free energy of the system is:

$$F = \mathcal{H}(q_0) - \frac{k_B T}{2} \ln\left(2\pi \lambda^2\right)$$

We thus see that the introduction of the fluctuations in the degree of freedom q has led to the appearance of an entropic term in the free energy.

---

[33]In solid state physics this assumption is often called *random phase approximation*, while in field theory *free field approximation*.

Gaussian approximation for N degrees of freedom This is a simple generalization of the previous case; the Hamiltonian will now be a function of the N-component vector $\vec{q} = (q_1, \ldots, q_N)$, and calling $\vec{q}_0$ the minimum of $\mathcal{H}$, expanding around $\vec{q}_0$ we get:

$$\mathcal{H}(\vec{q}) = \mathcal{H}(\vec{q}_0) + \sum_{\alpha,\beta} \frac{1}{2} \left(q_\alpha - q_{0\alpha}\right) \frac{\partial^2 \mathcal{H}}{\partial q_\alpha \partial q_\beta}_{|\vec{q}_0} \left(q_\beta - q_{0\beta}\right) + \cdots$$

Now, the Hessian matrix:

$$\mathcal{H}_{\alpha\beta} = \frac{\partial^2 \mathcal{H}}{\partial q_\alpha \partial q_\beta}_{|\vec{q}_0}$$

is of course symmetric (thanks to Schwarz's theorem), so it can be diagonalized. Calling $\hat{\lambda}_i$ its eigenvalues and defining $\beta\hat{\lambda}_i := 1/\lambda_i^2$, we can write:

$$\beta\mathcal{H}(\vec{q}) = \beta\mathcal{H}(\vec{q}_0) + \frac{1}{2} \sum_{i=1}^{N} \frac{\tilde{q}_i^2}{\lambda_i^2}$$

where $\tilde{q}_i$ is the fluctuation from the eigenvector of the Hessian relative to the eigenvalue $\hat{\lambda}_i$. We therefore have:

$$Z = \int_{-\infty}^{+\infty} e^{-\beta\mathcal{H}(\vec{q})} d\vec{q} = e^{-\beta\mathcal{H}(\vec{q}_0)} \prod_{i=1}^{N} \int_{-\infty}^{+\infty} e^{-\frac{\tilde{q}_i^2}{2\lambda_i^2}} dq_i$$

The last integrals are all independent and of the same form as the previous case. Thus, in the end we have:

$$F = \mathcal{H}(\vec{q}_0) - \frac{k_B T}{2} \sum_{i=1}^{N} \ln\left(2\pi\lambda_i^2\right)$$

Gaussian approximation for infinite degrees of freedom Let us now move to the really interesting case, i.e. the case of infinite degrees of freedom.
In general terms (we shall shortly see this explicitly for the Ising model) we want to compute a partition function of the form:

$$Z = \int e^{-\beta\mathcal{H}(m(\vec{r}))} \mathcal{D}[m(\vec{r})]$$

and the Gaussian approximation will be obtained, in analogy with the previous cases, determining the extremal (uniform) solution $m(\vec{r}) = m_0$ and then expanding $\mathcal{H}$ around m_0 to the second order in the fluctuation $\delta m(\vec{r}) = m(\vec{r}) - m_0$.
Thus, we will in general obtain a Hamiltonian of the form:

$$\mathcal{H}(m(\vec{r})) = \mathcal{H}(m_0) + \frac{1}{2} \int \delta m(\vec{r}) G_0^{-1}(\vec{r}, \vec{r}\,') \delta m(\vec{r}\,') d^d\vec{r} d^d\vec{r}\,'$$

where:

$$G_0^{-1}(\vec{r}, \vec{r}\,') := \frac{\delta^2 \mathcal{H}}{\delta m(\vec{r}) \delta m(\vec{r}\,')} \Big|_{\substack{m(\vec{r})=m_0 \\ h=0}}$$

Therefore:

$$F = \mathcal{H}(m_0) + \frac{k_B T}{2} \int \ln\left(\frac{G_0^{-1}(\vec{r}, \vec{r}\,')}{2\pi}\right) d^d\vec{r}$$

Gaussian approximation for the Ising model in Ginzburg-Landau theory   Let us now apply
what we have just only stated to a concrete case, i.e. the Ginzburg-Landau theory for the Ising
model we were considering.
In this case:

$$\beta\mathcal{H}(m(\vec{r})) = \int \left[\frac{k}{2}\left(\vec{\nabla}m\right)^2 + \frac{a}{2}tm^2 + \frac{b}{4}m^4 - hm \right] d^d\vec{r} \tag{6.36}$$

and we have seen that the stationary solution m_0 is such that:

$$h = atm_0 + bm_0^3$$

Let us now include also the fluctuations of m, substituting $m = m_0 + \delta m$. To the second order
in δm we have:

$$\left(\vec{\nabla}m\right)^2 = \left[\vec{\nabla}(m_0 + \delta m)\right]^2 = \left(\vec{\nabla}\delta m\right)^2$$

$$m^2 = (m_0 + \delta m)^2 = m_0^2 + 2m_0\delta m + (\delta m)^2$$

$$m^4 = (m_0 + \delta m)^4 = m_0^4 + 4m_0^3\delta m + 6m_0(\delta m)^2 + 4m_0(\delta m)^3 + (\delta m)^4 =$$
$$= m_0^4 + 4m_0^3\delta m + 6m_0(\delta m)^2 + O(\delta m^3)$$

and so setting $h = 0$:

$$\beta\mathcal{H}(m(\vec{r})) = \int \left[\frac{a}{2}tm_0^2 + \frac{b}{4}m_0^4 + (atm_0 + bm_0^3)\delta m + \left(\frac{a}{2}t + \frac{3}{2}bm_0^3\right)\delta m^2 + \frac{k}{2}\left(\vec{\nabla}\delta m\right)^2 \right] d^d\vec{r}$$

Since $atm_0 + bm_0^3 = h = 0$, defining for simplicity $A_0 := atm_0^2/2 + bm_0^4/4$ and calling V the
volume of the system we get:

$$\beta\mathcal{H}(m(\vec{r})) = A_0V + \int \left[\left(\frac{a}{2}t + \frac{3}{2}bm_0^3\right)\delta m^2 + \frac{k}{2}\left(\vec{\nabla}\delta m\right)^2 \right] d^d\vec{r} \tag{6.37}$$

In order to compute this integral it is more convenient to shift to Fourier space.

Let us make some remarks on what happens when we apply Fourier transformations in this
case.
If our system is enclosed in a cubic box of volume $V = L^d$, we can define the Fourier compo-
nents of the magnetization as:

$$m_{\vec{k}} = \int_V e^{-i\vec{k}\cdot\vec{r}}m(\vec{r})d^d\vec{r} \tag{6.38}$$

where $\vec{k} = 2\pi\vec{n}/L$ and $\vec{n}$ is a vector whose components are integer numbers. We can therefore
expand the magnetization in a Fourier series:

$$m(\vec{r}) = \frac{1}{V}\sum_{\vec{k}} e^{i\vec{k}\cdot\vec{r}}m_{\vec{k}}$$

Substituting this expression of m in $m_{\vec{k}}$ we obtain an integral representation for the Kronecker
delta; in fact:

$$m_{\vec{k}} = \sum_{\vec{k}\,'} m_{\vec{k}\,'}\left(\frac{1}{V}\int_V e^{i(\vec{k}-\vec{k}\,')\cdot\vec{r}}d^d\vec{r}\right)$$

and this is true only if:

$$\frac{1}{V}\int_V e^{i(\vec{k}-\vec{k}\,')\cdot\vec{r}}d^d\vec{r} = \delta_{\vec{k},\vec{k}\,'} \qquad \Rightarrow \qquad \int_V e^{i(\vec{k}-\vec{k}\,')\cdot\vec{r}}d^d\vec{r} = V\delta_{\vec{k},\vec{k}\,'}$$

Let us now make two observations. First: since $m(\vec{r})$ is real we have that $m_{\vec{k}}^* = m_{-\vec{k}}$. Second: our coarse graining procedure is based on the construction of blocks which have a linear dimension that cannot be smaller than a, the characteristic microscopic length of the system; this means that not all the $\vec{k}$ are allowed, and in particular we must have $|\vec{k}| \leq \pi/a := \Lambda$.

Now, thinking about the functional integral form of the partition function, what does the trace $\int \mathcal{D}[m(\vec{r})]$ become in Fourier space?

Since $m(\vec{r})$ is expressed in terms of the Fourier modes $m_{\vec{k}}$, which are in general complex, the measure of the integral becomes:

$$\int_{-\infty}^{+\infty} \prod_{|\vec{k}|<\Lambda} d\left(\operatorname{Re} m_{\vec{k}}\right) d\left(\operatorname{Im} m_{\vec{k}}\right)$$

However, since $m(\vec{r})$ is real (i.e. $m_{\vec{k}}^* = m_{-\vec{k}}$) the real and imaginary parts of the Fourier modes are not independent, because we have:

$$\operatorname{Re} m_{\vec{k}} = \operatorname{Re} m_{-\vec{k}} \qquad\qquad \operatorname{Im} m_{\vec{k}} = -\operatorname{Im} m_{-\vec{k}}$$

This means that if we use the trace we have written above we would integrate twice on the complex plane; we must therefore change the measure so as to avoid this double counting. We can for example simply divide everything by 2, or restrict the integration on the region where for example the last coordinate of $\vec{k}$, let us call it k_z, is positive.

Therefore:

$$\operatorname{Tr} = \int \mathcal{D}[m(\vec{r})] = \frac{1}{2} \int_{-\infty}^{+\infty} \prod_{|\vec{k}|<\Lambda} d\left(\operatorname{Re} m_{\vec{k}}\right) d\left(\operatorname{Im} m_{\vec{k}}\right) =$$

$$= \int_{-\infty}^{+\infty} \prod_{|\vec{k}|<\Lambda,\, k_z>0} d\left(\operatorname{Re} m_{\vec{k}}\right) d\left(\operatorname{Im} m_{\vec{k}}\right) := \int_{-\infty}^{+\infty} \prod_{\vec{k}}{}' dm_{\vec{k}}$$

where the last step defines the symbolic notation $\prod_{\vec{k}}'$. In the end, we have:

$$Z = \int_{-\infty}^{+\infty} \prod_{\vec{k}}{}' e^{-\beta\mathcal{H}\left(m_{\vec{k}}\right)} dm_{\vec{k}}$$

Let us now compute the partition function of the system in the simpler case $T > T_c$[34], so that in the end we can determine the free energy of the system.

In this case $m_0 = 0$ so $\delta m = m$, and therefore substituting (6.38) in (6.37) we get (renaming $\vec{q}$ the coordinate in Fourier space, so as not to confuse it with the constant k):

$$\beta\mathcal{H} - A_0 V = \int d^d\vec{r} \left(\frac{k}{2V^2} \sum_{\vec{q},\vec{q}'} \vec{q}\cdot\vec{q}'\, e^{i(\vec{q}-\vec{q}')\cdot\vec{r}} m_{\vec{q}} m_{\vec{q}'}^* + \frac{at}{2V^2} \sum_{\vec{q},\vec{q}'} e^{i(\vec{q}-\vec{q}')\cdot\vec{r}} m_{\vec{q}} m_{\vec{q}'}^* \right) =$$

$$= \frac{1}{2V^2} \sum_{\vec{q},\vec{q}'} m_{\vec{q}} m_{\vec{q}'}^* \left(k\vec{q}\cdot\vec{q}' + at\right) \underbrace{\int e^{i(\vec{q}-\vec{q}')\cdot\vec{r}} d^d\vec{r}}_{V\delta_{\vec{q},\vec{q}'}} = \frac{1}{2V} \sum_{\vec{q}} m_{\vec{q}} m_{-\vec{q}} \left(k|\vec{q}|^2 + at\right) =$$

$$= \frac{1}{2V} \sum_{\vec{q}} \left(k|\vec{q}|^2 + at\right) |m_{\vec{q}}|^2$$

---

[34] We could have equivalently considered the case $T < T_c$, but it is a bit more complicated since $m_0 \neq 0$ and so there is another term that contributes to the free energy. In other words, if $T < T_c$ then in m (as defined in (6.38)) we are not considering the term with $\vec{k} = 0$ (which is exactly equal to m).

Therefore substituting in the expression of the partition function the exponentials factorize, and in the end:

$$Z = \prod_{\vec{q}}' \int_{-\infty}^{+\infty} e^{-\frac{\beta}{2V}(kq^2 + at)|m_{\vec{q}}|^2 + \beta A_0 V} dm_{\vec{q}}$$

Since $|m_{\vec{q}}|^2 = \mathrm{Re}^2\, m_{\vec{q}} + \mathrm{Im}^2\, m_{\vec{q}}$, changing variables to:

$$x = \mathrm{Re}\, m_{\vec{q}} \qquad\qquad y = \mathrm{Im}\, m_{\vec{q}}$$

the integration in $dm_{\vec{q}} = d\left(\mathrm{Re}\, m_{\vec{q}}\right) d\left(\mathrm{Im}\, m_{\vec{q}}\right)$ gives:

$$\int_{-\infty}^{+\infty} dx\, dy\, e^{-\frac{kq^2 + at}{2V}(x^2 + y^2)} = \pi \frac{2V}{kq^2 + at}$$

Thus:

$$Z = \prod_{\vec{q}}' \frac{2\pi V}{kq^2 + at} e^{-\beta A_0 V}$$

We therefore have that the free energy of the system is:

$$F = -k_B T \ln Z = A_0 V - \frac{k_B T}{2} \sum_{|\vec{q}| < \Lambda} \ln\left(\frac{2\pi V}{kq^2 + at}\right)$$

We can now compute the specific heat of the system, and so determine its critical exponent α. We therefore want to compute:

$$C_V = -T \frac{\partial^2}{\partial T^2} \frac{F}{V}$$

The derivatives are straightforward, and in the end we get:

$$\frac{C_V}{T} = \frac{k_B T a}{V T_c^2} \sum_{|\vec{q}| < \Lambda} \frac{1}{(kq^2 + at)^2} - \frac{k_B a^2}{2V T_c} \sum_{|\vec{q}| < \Lambda} \frac{1}{kq^2 + at} \tag{6.39}$$

Let us now consider the two terms separately and study their behaviour for $t \to 0^+$ (we are in fact considering $T > T_c$).
Neglecting the proportionality constants that we don't need, we can rewrite the first contribution as:

$$I_1 = \frac{1}{V} \sum_{|\vec{q}| < \Lambda} \frac{1}{(kq^2 + at)^2} = \int_{|\vec{q}| < \Lambda} \frac{d^d \vec{q}}{(2\pi)^d} \frac{1}{(kq^2 + at)^2}$$

where we have substituted the sum with an integral[35], since the density of states in Fourier space is high (it is proportional to V, see the footnote). Now, using the definition of $\xi := \xi_> =$

---

[35]The substitution:

$$\sum_{|\vec{q}| < \Lambda} \longrightarrow \frac{V}{(2\pi)^d} \int_0^{\Lambda} d^d \vec{q}$$

can be justified as follows:

$$\sum_{|\vec{q}| < \Lambda} = \sum_{|\vec{q}| < \Lambda} \frac{\Delta \vec{q}}{\Delta \vec{q}} = (\Delta \vec{q})^{-1} \sum_{|\vec{q}| < \Lambda} \Delta \vec{q}$$

Now, $\vec{q}$ is quantized since V is finite and we have periodic boundary conditions, and in particular $\Delta \vec{q} = \Delta q_1 \cdot \Delta q_2 \cdots \Delta q_d = (2\pi/L)^d = (2\pi)^d/V$. Therefore:

$$\sum_{|\vec{q}| < \Lambda} = \frac{V}{(2\pi)^d} \int_{|\vec{q}| < \Lambda} d\vec{q}$$

$\sqrt{k/at}$ that we have previously seen we have:

$$I_1 = \frac{1}{k^2} \int_{|\vec{q}|<\Lambda} \frac{d^d\vec{q}}{(2\pi)^d} \frac{1}{(q^2 + \xi^{-2})^2} \tag{6.40}$$

In order to understand the divergent behaviour of the integral for $t \to 0^+$, we use a "scaling trick"; we change variable defining:

$$\vec{x} = \xi\vec{q}$$

so that the integral becomes:

$$I_1 = \frac{1}{k^2} \int_{|\vec{x}|<\xi\Lambda} \frac{d^d\vec{x}}{(2\pi)^d} \xi^{-d} \left(\frac{x^2}{\xi^2} + \frac{1}{\xi^2}\right)^{-2} = \frac{\xi^{4-d}}{k^2} \int_{|\vec{x}|<\xi\Lambda} \frac{d^d\vec{x}}{(2\pi)^d} \frac{1}{(1+x^2)^2}$$

and for $t \to 0^+$ we know that $\xi \to \infty$, so the integral is computed for all values of $x = |\vec{x}|$. Now, this integral must be computed shifting to spherical coordinates, so we will have $d^d\vec{x} \propto x^{d-1}dx$ (a part from numerical factors that involve all the angles, which are integrated trivially since the integrand only depends on x). Therefore the integrand of I_1 is $x^{d-1}/(1+x^2)^2$, and its behaviour for large or small x is:

$$\frac{x^{d-1}}{(1+x^2)^2} \sim \begin{cases} 0 & x \to 0 \\ x^{d-5} & x \to \infty \end{cases}$$

where of course in the first case we have $d > 1$; for large x, since in general:

$$\int_{\overline{x}}^{\infty} \frac{dx}{x^a} < \infty \qquad \text{if } a > 1$$

then the integral will converge if $5 - d > 1$, i.e. $d < 4$.
To sum up, the integral that appears in I_1 converges for $1 < d < 4$ (since in this case the integrand does not diverge in the domain of integration); of course, when this integral converges its result is simply a number. We therefore have that:

$$I_1 \propto \xi^{4-d} \qquad d < 4$$

and since $\xi \to \infty$ for $t \to 0^+$, we see that I_1 brings to C_V a diverging contribution[36] for $T \to T_c^+$. We can also wonder what happens for $d > 4$. From what we have stated about the rescaled form of I_1 (namely equation (6.40)) we could think that the integral diverges, but we must also take into account the prefactor ξ^{4-d}, which tends to zero for $d > 4$ as the transition is approached (since $\xi \to \infty$). The net result is *finite*, as could also be argued from the original (unscaled) form of I_1; in fact if $\xi \to \infty$ in (6.40) then (in spherical coordinates) the integrand is proportional to $q^{d-1-4} = q^{d-5}$, and since:

$$\int_0^{\overline{x}} \frac{dx}{x^a} < \infty \qquad \text{for } a < 1$$

then I_1 converges if $5 - d < 1$, i.e. $d > 4$.
To sum up, we can say that the first contribution to the specific heat behaves as:

$$\frac{k_B T a}{V T_c^2} \sum_{|\vec{q}|<\Lambda} \frac{1}{(kq^2 + at)^2} \sim \begin{cases} t^{-\nu(4-d)} & d < 4 \\ \text{finite} & d > 4 \end{cases}$$

---

[36] A little remark: the origin of this divergence does *not* come from the behaviour of the integral for large wavelengths. In fact, in the original definition of I_1 the integral has an upper limit, Λ, so it cannot diverge because of the large q behaviour; if it diverges, it must be because of the behaviour for $q \to 0$, which corresponds to large wavelengths (this is also why this divergence is sometimes called *infrared divergence*).

where we have also used the definition of the exponent ν, i.e. $\xi \sim t^{-\nu}$.

Note that this also means that in the Gaussian approximation this term brings no corrections to the exponent α above four dimensions.

Let us now consider the second contribution to C_V/T in (6.39).

In particular, as we have done before we rewrite it substituting the sum with an integral and also using the definition of ξ, so that:

$$I_2 = \frac{1}{k} \int_{|\vec{q}|<\Lambda} \frac{d^d\vec{q}}{(2\pi)^d} \frac{1}{q^2 + \xi^{-2}}$$

Again, we are interested in its behaviour for $\xi \to \infty$, and as we have already noted it is only the behaviour of the integrand for $q \to 0$ that can cause any divergence.

As previously done, changing variable to $\vec{x} = \xi\vec{q}$:

$$I_2 = \frac{\xi^{2-d}}{k} \int_{|\vec{x}|<\xi\Lambda} \frac{d^d\vec{x}}{(2\pi)^d} \frac{1}{x^2 + 1}$$

and using spherical coordinates:

$$I_2 \propto \xi^{2-d} \int_0^{\xi\Lambda} \frac{x^{d-1}}{x^2 + 1} dx$$

The integrand behaves as:

$$\frac{x^{d-1}}{x^2 + 1} \sim \begin{cases} 0 & x \to 0 \\ x^{d-3} & x \to \infty \end{cases}$$

where again in the first case $d > 1$. Therefore, the integral in I_2 (*not* I_2 itself) in the limit $\xi \to \infty$ converges if $3 - d > 1$, i.e. $d < 2$; this means that for $d < 2$ and $t \to 0^+$ we have that I_2 behaves as:

$$I_2 \sim \xi^{2-d} \sim t^{-\nu(2-d)}$$

For the case $d > 2$ it is more convenient to consider the unscaled form of I_2. In this case, in the limit $\xi \to \infty$ we have (again in spherical coordinates):

$$I_2 \propto \int_0^\Lambda \frac{q^{d-1}}{q^2} dq$$

and the integrand converges for $q \to 0$ if $d > 2$.

Therefore for $d < 2$ the second contribution to (6.39) diverges, but in the same range of d the divergence of the first contribution is more relevant; on the other hand, for $2 \le d < 4$ only the first contribution diverges.

It is therefore the term containing I_1 that determines the divergence of the specific heat, and in particular for $d < 4$ we have $C_V \sim t^{-\nu(4-d)}$ and so we see that in the Gaussian approximation the inclusion of the fluctuations has changed the value of the critical exponent α to[37]:

$$\alpha = \nu(4 - d)$$

In order to compute it, however, we still must determine ν so we now proceed to compute the two-point correlation function in order to determine both η and ν.

---

[37]We stress again that the same calculations could have been done in the case $T < T_c$, but we have not done so only for the sake of simplicity.

Two-point correlation function in the Gaussian approximation We know that the (simple) correlation function is defined as:

$$G(\vec{r}, \vec{r}\,') = \left\langle m(\vec{r})m(\vec{r}\,') \right\rangle$$

so we first have to determine:

$$m(\vec{r})m(\vec{r}\,') = \frac{1}{V^2} \sum_{\vec{q}, \vec{q}\,'} e^{i\vec{q}\cdot(\vec{r}+\vec{r}\,')} m_{\vec{q}} m_{\vec{q}\,'}$$

Shifting to Fourier space, we have:

$$\left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle = \frac{\int dm_{\vec{q}_1} dm_{\vec{q}_2} \cdots dm_{\vec{q}} dm_{\vec{q}\,'} e^{-\beta\mathcal{H}} m_{\vec{q}} m_{\vec{q}\,'}}{\int dm_{\vec{q}_1} dm_{\vec{q}_2} \cdots dm_{\vec{q}} dm_{\vec{q}\,'} e^{-\beta\mathcal{H}}}$$

where:

$$\beta\mathcal{H} = A_0 V + \frac{1}{2V} \sum_{\vec{k}, \vec{k}\,'} \left(k|\vec{q}|^2 + at \right) |m_{\vec{q}}|^2$$

It is clear that in $\left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle$ all the integrals factorize since the Fourier modes are all independent (they are decoupled); therefore, all the integrals in the numerator that don't involve $\vec{q}$ or $\vec{q}\,'$ simplify with the same integrals in the denominator, so that in the end we are left with:

$$\left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle = \frac{\int e^{-\frac{k|\vec{q}|^2+at}{2V}|m_{\vec{q}}|^2} e^{-\frac{k|\vec{q}\,'|^2+at}{V}|m_{\vec{q}\,'}|^2} m_{\vec{q}} m_{\vec{q}\,'} dm_{\vec{q}} dm_{\vec{q}\,'}}{\int e^{-\frac{k|\vec{q}|^2+at}{V}|m_{\vec{q}}|^2} e^{-\frac{k|\vec{q}\,'|^2+at}{V}|m_{\vec{q}\,'}|^2} dm_{\vec{q}} dm_{\vec{q}\,'}}$$

(where the factor 2 in the denominator of the exponent has disappeared because we must remember that $m_{\vec{q}} = m_{-\vec{q}}^*$). There are now two possible cases:

$\vec{q} \neq \pm\vec{q}\,'$: in this case (which can be re-expressed as $|\vec{q}| \neq |\vec{q}\,'|$) the two coefficients $m_{\vec{q}}$ and $m_{\vec{q}\,'}$ are distinct, and in the numerator the double integral factorizes into two integrals of the form:

$$\int e^{-\frac{k|\vec{q}|^2+at}{V}|m_{\vec{q}}|^2} m_{\vec{q}} dm_{\vec{q}} = 0$$

since the integrand is odd. We therefore have:

$$G_{\vec{q}, \vec{q}\,'} = \left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle = 0 \qquad |\vec{q}| \neq |\vec{q}\,'|$$

$\vec{q} = \pm\vec{q}\,'$: in this case (equivalent to $|\vec{q}| = |\vec{q}\,'|$) we can either have $\vec{q} = \vec{q}\,'$ so that $\left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle = \left\langle m_{\vec{q}}^2 \right\rangle$, or $\vec{q} = -\vec{q}\,'$ so that $\left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle = \left\langle m_{\vec{q}} m_{\vec{q}}^* \right\rangle = \left\langle |m_{\vec{q}}^2| \right\rangle$.

Let us first consider the case $\vec{q} = \vec{q}\,'$. Using polar coordinates, we define $m_{\vec{q}} = |m_{\vec{q}}|e^{i\theta_{\vec{q}}}$ so that the measure $dm_{\vec{q}}$ in the complex plane becomes:

$$dm_{\vec{q}} = |m_{\vec{q}}| d|m_{\vec{q}}| d\theta_{\vec{q}}$$

Thus:

$$\left\langle m_{\vec{q}}^2 \right\rangle = \frac{\int e^{-\beta\mathcal{H}} m_{\vec{q}}^2 dm_{\vec{q}}}{\int e^{-\beta\mathcal{H}} dm_{\vec{q}}} \propto \int_0^{2\pi} e^{2i\theta_{\vec{q}}} d\theta_{\vec{q}} = 0$$

We are therefore left with the last case $\vec{q} = -\vec{q}\,'$: now, shifting to polar coordinates the integrals in both the numerator and denominator involving $\theta_{\vec{q}}$ factorize and simplify, so in the end renaming $x = |m_{\vec{q}}|$ in the integrals for simplicity:

$$\left\langle |m_{\vec{q}}|^2 \right\rangle = \frac{\int_0^\infty e^{-\frac{k|\vec{q}|^2+at}{V}x^2} x^2 \cdot x \, dx}{\int_0^\infty e^{-\frac{k|\vec{q}|^2+at}{V}x^2} x \, dx}$$

Noting that $xdx = dx^2/2$, if we change variable to $y = x^2$ and integrate we get[38]:

$$\left\langle |m_{\vec{q}}|^2 \right\rangle = \frac{\int_0^\infty e^{-\frac{k|\vec{q}|^2 + at}{V}y} y \, dy}{\int_0^\infty e^{-\frac{k|\vec{q}|^2 + at}{V}y} \, dy} = \frac{V}{k|\vec{q}|^2 + at}$$

Therefore since the correlation function in Fourier space is non null only when $\vec{q} = -\vec{q}\,'$, in general we can write:

$$G_{\vec{q},\vec{q}\,'} = \left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle = \frac{V}{k|\vec{q}|^2 + at} \delta_{\vec{q}, -\vec{q}\,'}$$

Going back to real space we have:

$$G(\vec{r}, \vec{r}\,') = \left\langle m_{\vec{r}} m_{\vec{r}\,'} \right\rangle = \frac{1}{V^2} \sum_{\vec{q}, \vec{q}\,'} e^{i(\vec{q}\cdot\vec{r} + \vec{q}\,'\cdot\vec{r}\,')} \left\langle m_{\vec{q}} m_{\vec{q}\,'} \right\rangle =$$

$$= \frac{1}{V^2} \sum_{\vec{q}, \vec{q}\,'} e^{i(\vec{q}\cdot\vec{r} + \vec{q}\,'\cdot\vec{r}\,')} \frac{V \delta_{\vec{q},\vec{q}\,'}}{k|\vec{q}|^2 + at} = \frac{1}{V} \sum_{\vec{q}} \frac{e^{i\vec{q}\cdot(\vec{r} - \vec{r}\,')}}{k|\vec{q}|^2 + at}$$

We see that appropriately substituting the sum with an integral (see the footnote on page 184) and defining:

$$\xi = \sqrt{\frac{k}{at}}$$

this correlation function acquires the same form of the one computed in mean field theory, i.e. equation (6.35). This means that the critical exponents ν and η now have the same values predicted by mean field theory, namely:

$$\nu = \frac{1}{2} \qquad\qquad \eta = 0$$

Interaction between fluctuations: expansion to the fourth order

We have therefore seen that mean field theories can be improved including the fluctuations of the order parameter around its extremal values; in particular with the Gaussian approximation we have stopped the expansion at the second order and this led to a change in the critical exponent of the specific heat, which now really diverges instead of exhibiting a jump discontinuity as simple mean field theories predict. However, the quartic term δm^4 that we ignore within the Gaussian approximation (and which basically represent the interactions between the fluctuations) becomes crucial when we approach a critical point.

We could thus wonder if the Gaussian approximation can be improved. In particular, reconsidering equation (6.36) with $h = 0$:

$$\beta \mathcal{H}(m(\vec{r})) = \int \left[\frac{k}{2} \left(\vec{\nabla} m \right)^2 + \frac{a}{2} t m^2 + \frac{b}{4} m^4 \right] d^d \vec{r}$$

then keeping all the terms in δm when expanding m around m_0, remembering that $atm_0 + bm_0^3 = 0$ and considering that the odd terms in δm give no contribution (since we integrate an odd function over an even domain), we have:

$$\beta \mathcal{H} = A_0 V + \int \left[\left(\frac{a}{2} t + \frac{3}{2} b m_0^3 \right) \delta m^2 + \frac{k}{2} \left(\vec{\nabla} \delta m \right)^2 \right] d^d \vec{r} + \frac{b}{4} \int \delta m^4 d^d \vec{r}$$

---

[38]In order to compute the integral in the numerator we have used a standard "trick":

$$\int_0^\infty e^{-kx} x \, dx = -\frac{\partial}{\partial k} \int_0^\infty e^{-kx} dx = -\frac{\partial}{\partial k} \left(\frac{e^{-kx}}{-k} \right)_0^{+\infty} = -\frac{\partial}{\partial k} \frac{1}{k} = \frac{1}{k^2}$$

A natural approach to compute the partition function Z would now consist in using a perturbative method in order to expand the term $\exp\left(-\frac{b}{4}\int \delta m^4 d^d\vec{r}\right)$ in powers of the parameter b. This is of course a reasonable approach if b is small; however, with a simple dimensional analysis we can show (and this is what we are going to do in the following) that for $t \to 0$ and $d < 4$ this parameter diverges. The approach of the Gaussian approximation is therefore inconsistent, at least from this point of view.

Dimensional analysis of Landau theory   We know that the partition function of the system is:

$$Z = \int e^{-\beta\mathcal{H}(m(\vec{r}))}\mathcal{D}[m(\vec{r})]$$

where, in the case $h = 0$:

$$\beta\mathcal{H}(m(\vec{r})) = \int \left[\frac{k}{2}\left(\vec{\nabla}m\right)^2 + \frac{a}{2}tm^2 + \frac{b}{4}m^4\right]d^d\vec{r}$$

It is now convenient to rescale the order parameter m so that the term proportional to $\vec{\nabla}m$ has only a numerical coefficient. This can be done defining:

$$\varphi = m\sqrt{k} \qquad\qquad r_0 = \frac{at}{k} \qquad\qquad u_0 = \frac{b}{4k^2} \qquad\qquad (6.41)$$

so that:

$$\mathcal{H}_{\text{eff.}} := \beta\mathcal{H} = \int \left[\frac{1}{2}\left(\vec{\nabla}\varphi\right)^2 + \frac{r_0}{2}\varphi^2 + u_0\varphi^4\right]d^d\vec{r}$$

where we have defined the *dimensionless*[39] effective Hamiltonian $\mathcal{H}_{\text{eff.}}$.
Now, since $\mathcal{H}_{\text{eff.}}$ is dimensionless all the three integrals that appear in $\mathcal{H}_{\text{eff.}}$ must be so; this means that φ, r_0 and u_0 must have precise dimensions.
In fact, from the first contribution we have that:

$$\left[\int \left(\vec{\nabla}\varphi\right)^2 d^d\vec{r}\right] = 1 \qquad \Rightarrow \qquad L^d\frac{[\varphi]^2}{L^2} = 1 \qquad \Rightarrow \qquad [\varphi] = L^{1-d/2}$$

and similarly:

$$\left[\int r_0\varphi^2 d^d\vec{r}\right] = \left[\int u_0\varphi^4 d^d\vec{r}\right] = 1$$

from which we have that:

$$[r_0] = L^{-2} \qquad\qquad [u_0] = L^{d-4}$$

We can therefore use r_0 to define a length scale independent of the dimension of the system, i.e. $r_0^{-1/2}$. Since $r_0 \propto t$ by definition and since $\xi \sim t^{-1/2}$ within mean field theories and Gaussian approximation, we see that this choice is equivalent to measuring lengths in units of the correlation length ξ (which we know is independent of the dimension of the system).
We now rescale the variables, including those defined in (6.41), by setting:

$$\phi = \frac{\varphi}{\ell^{1-d/2}} \qquad\qquad \vec{x} = \frac{\vec{r}}{\ell} \qquad\qquad u = \frac{u_0}{\ell^{d-4}}$$

where $\ell = r_0^{-1/2}$.
This way, the partition function can be written in the form:

$$Z = \int e^{-\mathcal{H}_0(\phi)-\mathcal{U}(\phi)}\mathcal{D}[\phi]$$

---

[39]Remember that β has the dimension of the inverse of an energy.

where:

$$\mathcal{H}_0 = \int \left[\frac{1}{2} \left(\vec{\nabla}\phi \right)^2 + \frac{1}{2}\phi^2 \right] d^d\vec{x} \qquad\qquad \mathcal{U} = \int u\phi^4 d^d\vec{x}$$

and $\mathcal{U}$, depending on the quartic term in ϕ, is the contribution due to the interaction between the fluctuations. It is the presence of this terms that prevents us from computing Z exactly and that forces us to resort to approximations.

The most standard procedure to apply in this case would be a perturbative method, namely to consider the dimensionless parameter u as small, i.e. $u \ll 1$, and expanding the exponential term containing $\mathcal{U}$:

$$Z = \int e^{-\mathcal{H}_0} \left(1 - \mathcal{U} + \frac{1}{2}\mathcal{U}^2 + \cdots \right) \mathcal{D}[\phi]$$

Written out explicitly, we have:

$$u = \frac{u_0}{r_0^{(d-4)/2}} = u_0 \left(\frac{at}{k} \right)^{(d-4)/2}$$

We thus immediately see that if $d < 4$ then u diverges when $t \to 0$, making the perturbative approach infeasible. On the other hand, if $d > 4$ we indeed have $u \to 0$ when $t \to 0$, so the Gaussian approximation is actually a good one.

Now, we can say that the perturbative expansion is reasonable if $u \leq 1$, which gives:

$$t^{(d-4)/2} \geq u_0 \left(\frac{a}{k} \right)^{(d-4)/2}$$

This, using also (6.41), can be rewritten as:

$$t^{(4-d)/2} \geq \frac{b}{4a^2\xi_1^d} \qquad\qquad \text{where} \quad \xi_1 := \xi(t=1) = \sqrt{\frac{a}{k}}$$

which is therefore a criterion that tells us if the perturbative approach is valid.
Note that from this analysis we have determined the upper critical dimension of the system ($d = 4$), so we can say that this last criterion is equivalent to the Ginzburg criterion.

A final remark.
The argument we have shown is not really convincing. In fact, we have only shown that every term of the perturbation theory diverges as $t \to 0$; however, this does not necessarily mean that the whole perturbation series is divergent. For example, consider the exponential series:

$$e^{-x} = \sum_{n=0}^{\infty} \frac{(-x)^n}{n!}$$

This is convergent for $x \to \infty$ ($\lim_{x\to\infty} e^{-x} = 0$), but each term of the sum diverges in the same limit.
We thus understand that we must be careful when handling perturbation expansions, since the appropriate resummation of divergent terms can bring to a convergent series.

Chapter 7

Scaling theory

We have seen in 5.1.2 that as a given system approaches a critical point, the distance ξ over which the fluctuations of the order parameter are correlated becomes comparable to the size of the whole system and the microscopic aspects of the system become irrelevant (which is the phenomenon of *universality* we have encountered in 1.5). This means that near a critical point the system has no longer characteristic lengths, besides ξ of course.
We can therefore expect that if we "move" a little bit from a critical point, for example changing the temperature by a small amount, the free energy of the system as a function will not change its *shape*, but will be *rescaled*.

This is the main idea of *scaling theory*; in order to understand it the concept of homogeneous function is essential, and in appendix E we recall the most important facts about them.

7.1 Widom's static scaling theory

We have seen in the first part of chapter 5 that when a phase transition occurs the free energy of the system is such that the response functions exhibit singularities, often in the form of divergences.
To make a concrete example (but of course all our statements are completely general) if we consider a magnetic system we can suppose to write its free energy density as:

$$f(T, H) = f_{\rm r}(T, H) + f_{\rm s}(t, h)$$

where $t = (T - T_c)/T_c$, $h = (H - H_c)/k_B T$, $f_{\rm r}$ is the "regular" part of the free energy (which does not significantly change near a critical point), while $f_{\rm s}$ is the "singular" one, which contains the non-analytic behaviour of the system near a critical point (i.e. $t \approx 0$ and $h \approx 0$).

Widom's *static scaling hypothesis* consists in assuming that the singular part $f_{\rm s}$ of the free energy is a *generalized homogeneous function*, i.e.:

$$f_{\rm s}(\lambda^{p_1} t, \lambda^{p_2} h) = \lambda f_{\rm s}(t, h) \qquad \forall \lambda \in \mathbb{R} \tag{7.1}$$

in appendix E we discuss the main properties of such functions.
Note that assuming that one thermodynamic potential is a generalized homogeneous function implies that all the other thermodynamic potentials are so.

The exponents p_1 and p_2 are not specified by the scaling hypothesis; however, we are shortly going to show that *all* the critical exponents of a system can be expressed in terms of p_1 and p_2; this also implies that if two critical exponents are known, we can write p_1 and p_2 in terms of them (since in general we will have a set of two independent equations in the two variables

p_1 and p_2) and therefore determine all the critical exponents of the system. In other words, we just need to know two critical exponents to obtain all the others.

As shown in appendix E, an important property of generalized homogeneous functions is that with a proper choice of λ we can remove the dependence on one of their arguments; for example, if in our case we choose $\lambda = h^{-1/p_2}$ then:

$$f_s(t, h) = h^{1/p_2} f_s\left(\frac{t}{h^{p_1/p_2}}, 1\right) \tag{7.2}$$

The ratio $\Delta := p_2/p_1$ is sometimes called *gap exponent*.

7.1.1   Relations between critical exponents

Let us now explore the consequences of (7.1) on the critical exponents of a system, again on a magnetic one for concreteness.

Exponent β   Since $M = \partial f/\partial H$, deriving both sides of (7.1) with respect to h^1 we get:

$$\lambda^{p_2} \frac{\partial}{\partial h} f_s(\lambda^{p_1} t, \lambda^{p_2} h) = \lambda \frac{\partial}{\partial h} f_s(t, h)$$

and thus:

$$\lambda^{p_2} M(\lambda^{p_1} t, \lambda^{p_2} h) = \lambda M(t, h) \tag{7.3}$$

In order to determine β, we set $h = 0$ so that (7.3) becomes:

$$\lambda^{p_2 - 1} M(\lambda^{p_1} t, 0) = M(t, 0)$$

and using property (7.2), we set $\lambda = (-t)^{-1/p_1}$ to get:

$$M(t, 0) = (-t)^{\frac{1 - p_2}{p_1}} M(-1, 0)$$

By definition of the β critical exponent, we have:

$$\beta = \frac{1 - p_2}{p_1} \tag{7.4}$$

Exponent δ   Using again (7.3), we can determine the exponent δ setting $t = 0$:

$$M(0, h) = \lambda^{p_2 - 1} M(0, \lambda^{p_2} h)$$

Now, using again (7.2) we set $\lambda = h^{-1/p_2}$ and get:

$$M(0, h) = h^{\frac{1 - p_2}{p_2}} M(0, 1)$$

so that:

$$\delta = \frac{p_2}{1 - p_2} \tag{7.5}$$

Now we can also solve (7.4) and (7.5):

$$p_1 = \frac{1}{\beta(\delta + 1)} \qquad\qquad p_2 = \frac{\delta}{\delta + 1} \tag{7.6}$$

from which we see that the gap exponent is:

$$\Delta = \frac{p_2}{p_1} = \beta\delta \tag{7.7}$$

---

[1]We should in principle derive with respect to H, but since $h \propto \beta H$, the β factors simplify on both sides.

Exponent γ In order to obtain the magnetic susceptibility, we derive twice (7.1) with respect to h, to get:

$$\lambda^{2p_2}\chi_T(\lambda^{p_1}t, \lambda^{p_2}h) = \lambda\chi_T(t, h) \tag{7.8}$$

The exponent γ describes the behaviour of χ_T for $t \to 0$ when no external field is present. What we can now see is that the scaling hypothesis leads to the equality of the exponents for $t \to 0^+$ and $t \to 0^-$ (which we just assumed for simplicity in 1.5).

Setting $h = 0$ and $\lambda = (-t)^{-1/p_1}$ we get:

$$\chi_T(t, 0) = (-t)^{-\frac{2p_2-1}{p_1}}\chi_T(-1, 0)$$

and if we call γ^- the critical exponent for $t \to 0^-$, we see that:

$$\gamma^- = \frac{2p_2 - 1}{p_1}$$

In order to compute the exponent γ^+ that describes the behaviour of χ_T for $t \to 0^+$, we set $\lambda = t^{-1/p_1}$, so that (7.8) becomes:

$$\chi_T(t, 0) = t^{-\frac{2p_2-1}{p_1}}\chi_T(1, 0)$$

so that indeed:

$$\gamma^+ = \frac{2p_2 - 1}{p_1}$$

We therefore see explicitly that:

$$\gamma^+ = \gamma^- = \gamma = \frac{2p_2 - 1}{p_1} \tag{7.9}$$

which, using (7.6), leads to:

$$\gamma = \beta(\delta - 1) \tag{7.10}$$

Exponent α In order to determine the behaviour of the specific heat (at constant external field) near the critical point, we derive (7.1) twice with respect to the temperature, so that:

$$\lambda^{2p_1}C_H(\lambda^{p_1}t, \lambda^{p_2}h) = \lambda C_H(t, h)$$

Setting $h = 0$ and $\lambda = (-t)^{-1/p_1}$:

$$C_H(t, 0) = (-t)^{-\left(2-\frac{1}{p_1}\right)}C_H(-1, 0)$$

so:

$$\alpha^- = 2 - \frac{1}{p_1}$$

Again, we can see that this exponent is equal to the one that we get for $t \to 0^+$; in fact, setting $\lambda = t^{-1/p_1}$:

$$C_H(t, 0) = t^{-\left(2-\frac{1}{p_1}\right)}C_H(1, 0)$$

so that indeed $\alpha^+ = \alpha^-$. Therefore:

$$\alpha = 2 - \frac{1}{p_1} \tag{7.11}$$

Griffiths and Rushbrooke's equalities   If we now substitute (7.6) into (7.11), we get:

$$\alpha + \beta(\delta + 1) = 2$$

This is *Griffiths equality*, which we have already encountered in 1.5.1 as an inequality.
On the other hand, Rushbrooke's equality is obtained using (7.4) and (7.10) to determine p_2 and p_1 and then substituting into (7.11):

$$\alpha + 2\beta + \gamma = 2$$

We therefore see, as anticipated in 1.5.1, that the static scaling hypothesis allows to show that they are indeed exact equalities.

An alternative expression for the scaling hypothesis   We can re-express (7.1) in another fashion often used in literature. If we set $\lambda = t^{-1/p_1}$, then:

$$f_s(1, t^{-p_2/p_1} h) = t^{-1/p_1} f_s(t, h)$$

From (7.7) and (7.11), we can rewrite this as:

$$f_s(t, h) = t^{2-\alpha} f_s\left(1, \frac{h}{t^\Delta}\right) \tag{7.12}$$

which is the most used form of the scaling hypothesis in statistical mechanics.

As we can notice, we have not considered the critical exponents η and ν; this will be done shortly in 7.2.

7.1.2   Rescaled state equation

Besides the relations between critical exponents, Widom's static scaling theory allows us to make predictions on the shape of the state equation of a given system. Let us now see how, again for a magnetic system.

We begin from (7.3):

$$M(t, h) = \lambda^{p_2 - 1} M(\lambda^{p_1} t, \lambda^{p_2} h)$$

If we set $\lambda = |t|^{-1/p_1}$:

$$M(t, h) = |t|^{\frac{1-p_2}{p_1}} M\left(\frac{t}{|t|}, \frac{h}{|t|^{p_2/p_1}}\right)$$

and using (7.6) and (7.7):

$$\frac{M(t, h)}{|t|^\beta} = M\left(\frac{t}{|t|}, \frac{h}{|t|^\Delta}\right) \tag{7.13}$$

We can therefore define the *rescaled magnetization* and the *rescaled magnetic field*:

$$\tilde{m} := |t|^{-\beta} M(t, h) \qquad\qquad \tilde{h} := |t|^{-\Delta} h \tag{7.14}$$

so that (7.13) becomes:

$$\tilde{m} = M(\pm 1, \tilde{h}) \tag{7.15}$$

where $+1$ corresponds to $t > 0$ (namely $T > T_c$) and -1 to $t < 0$ (i.e. $T < T_c$).

Equation (7.15) means that if we measure M and h and rescale them as in (7.14), all the experimental data should fall on the same curve *independently of the temperature T*; there are of course two possible curves (not necessarily equal), one for $T > T_c$ and one for $T < T_c$ (which correspond to $M(1, h)$ and $M(-1, h)$).
These predictions are in perfect agreement with experimental results[2], and are one of the greatest successes of Widom's static scaling theory.

---

[2]See for example [8], pg. 119

7.2 Kadanoff's scaling and correlation lengths

As we have seen, Widom's static scaling theory allows us to determine exact relations between critical exponents, and to interpret the scaling properties of systems near a critical point. However, this theory is based upon equation (7.1) but gives no physical interpretation of it; in other words, it does not tell anything about the physical origin of scaling laws. Furthermore, as we have noticed Widom's theory does not involve correlation lengths, so it tells nothing about the critical exponents ν and η.

We know (see 5.1.2) that one of the characteristic traits of critical phenomena is the divergence of the correlation length ξ, which becomes the only physically relevant length near a critical point. However, by now we are unable to tell if and how this is related to the scaling hypothesis (7.1); everything will become clearer within the framework of the Renormalization Group (see chapter 8), in which we will see that (7.1) is a consequence of the divergence of correlation length.

Nonetheless, before the introduction of the Renormalization Group Kadanoff proposed a plausibility argument for (7.1) applied to the Ising model, which we are now going to analyse.
We will see that Kadanoff's argument, which is based upon the intuition that the divergence of ξ implies a relation between the coupling constants of an effective Hamiltonian and the length on which the order parameter is defined, is correct in principle but not in detail because these relations are in reality more complex than what predicted by Kadanoff; furthermore, Kadanoff's argument does not allow an explicit computation of critical exponents.
We will have to wait for the Renormalization Group (chapter 8) in order to solve these problems.

7.2.1 Kadanoff's argument for the Ising model

Let us consider a d-dimensional Ising model with hypercubic lattice with lattice constant a; assuming nearest-neighbour interactions the Hamiltonian of the system will be:

$$-\beta\mathcal{H} = K \sum_{\langle ij \rangle} S_i S_j + h \sum_i S_i$$

where $K = \beta J$ and $h = \beta H$, as usual.

Since the values of the spin variables are correlated on lengths of the order of $\xi(T)$, the spins contained in regions of linear dimension ℓa, with ℓ such that:

$$a \ll \ell a \ll \xi(T)$$

will behave, statistically, as a single unit.
We can therefore imagine to carry out, similarly to what we have seen for the Ginzburg-Landau theory (see 6.6.2), a coarse graining procedure were we substitute the spin variables S_i inside a "block" of linear dimension ℓa (which will therefore contain ℓ^d spins) with a single *block spin*; the total number of blocks will of course be $N_{\mathrm{b}} = N/\ell^d$.
Considering the I-th block, we can define the block spin S_I as:

$$S_I = \frac{1}{|m_\ell|} \cdot \frac{1}{\ell^d} \sum_{i \in I} S_i$$

where the mean magnetization of the I-th block m_ℓ is:

$$m_\ell = \frac{1}{\ell^d} \sum_{i \in I} \langle S_i \rangle$$

(a) Original system

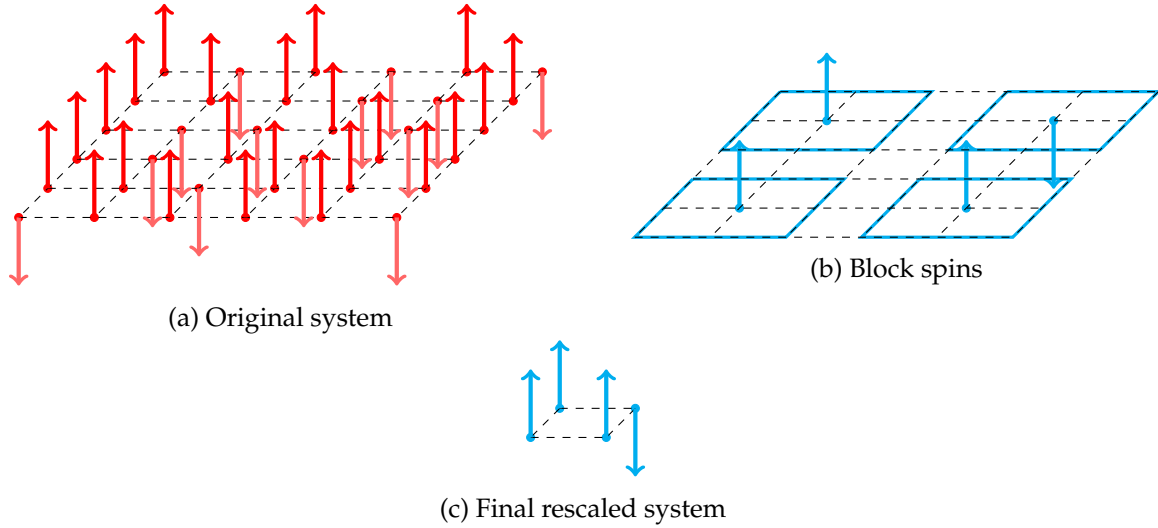(b) Block spins

(c) Final rescaled system

Figure 7.1: Kadanoff's coarse graining procedure in two dimensions

This way, the new block spin variables can assume only the values ± 1, just like the original ones.

In the end we are left with a system of block spins on a hypercubic lattice with lattice constant ℓa. We can therefore *rescale* the spatial distances between the degrees of freedom of our system:

$$\vec{r}_\ell = \frac{\vec{r}}{\ell} \tag{7.16}$$

In other words, since ℓa is now the characteristic length of the system we are measuring distances in units of ℓa (just like in the original one we measured distances in units of a).
The coarse graining procedure we have just seen is described in figure 7.1 for a two-dimensional Ising model.

Kadanoff's argument now proceeds with two assumptions.

Assumption 1. *In analogy to what happens in the original system, we assume that the block spins interact with the nearest neighbours and an external effective field (just like the original ones do).*

This means that the Hamiltonian of the coarse grained system of block spins has the same form of the original one, of course provided that the spins, coupling constants and external fields are redefined. If we call K_ℓ and h_ℓ these new constants (of course $K_1 = K$ and $h_1 = h$) the new effective Hamiltonian is:

$$-\beta \mathcal{H}_\ell = K_\ell \sum_{\langle IJ \rangle} S_I S_J + h_\ell \sum_{I=1}^{N/\ell^d} S_I$$

Since in the new system the lengths have been rescaled by a factor ℓ, as in (7.16), this means that also the correlation length of the system has to be measured in units of ℓa, and in particular we will have:

$$\xi_\ell = \frac{\xi}{\ell}$$

This means that the new system has a *lower* correlation length, and it will thus be *farther* from the critical point with respect to the original one, and so will have a new effective temperature t_ℓ.

Similarly, in the coarse grained system the magnetic field will be rescaled to an effective one:

$$h \sum_i S_i = h \sum_I \sum_{i \in I} S_i = h m_\ell \ell^d \sum_I S_I := h_\ell \sum_I S_I \tag{7.17}$$

which implies that there is a relation between the new magnetic field and the mean magnetization:

$$h_\ell = h m_\ell \ell^d \tag{7.18}$$

Since the Hamiltonian of the block spin system has the same form of the original one, the same will be true also for the free energy, provided that h, K and N are substituted with h_ℓ, K_ℓ and N/ℓ^d; in particular, considering the singular part f_s of the free energy density we will have:

$$\frac{N}{\ell^d} f_s(t_\ell, h_\ell) = N f_s(t, h)$$

and so:

$$f_s(t_\ell, h_\ell) = \ell^d f_s(t, h) \tag{7.19}$$

In order to proceed, we now need the second assumption.

Assumption 2. *We assume that:*

$$t_\ell = t\ell^{y_t} \qquad h_\ell = h\ell^{y_h} \qquad y_t, y_h > 0$$

The justification of this assumption lies in the fact that we are trying to understand the scaling properties of our system near a critical point, and these are the simplest possible relations between (t, h) and (t_ℓ, h_ℓ) that satisfy the following symmetry requirements:

- when $h \to -h$, then $h_\ell \to -h_\ell$

- when $h \to -h$, then $t_\ell \to t_\ell$

- when $t = h = 0$, then $t_\ell = h_\ell = 0$

The exponents y_t and y_h are for now unspecified, apart from the fact that they must be positive (so that the coarse grained system is indeed farther from the critical point with respect to the original one).

If we use this assumption in (7.19) we get:

$$f_s(t, h) = \frac{1}{\ell^d} f_s(t\ell^{y_t}, h\ell^{y_h})$$

This is very similar to Widom's scaling hypothesis (7.1), where the parameter λ is the inverse of the block volume ℓ^d.

Since ℓ has no specified value, we can choose the one we want and again we use the properties of generalized homogeneous functions to eliminate one of the arguments of f_s. In particular, setting $\ell = |t|^{-1/y_t}$ we get:

$$f_s(t, h) = |t|^{d/y_t} f_s\left(1, h|t|^{-y_h/y_t}\right)$$

where the gap exponent is now $\Delta = y_h/y_t$; comparing this equation with (7.12) we have:

$$2 - \alpha = \frac{d}{y_t} \tag{7.20}$$

This relation will be used in the following, after we have discussed the scaling of correlation functions.

Kadanoff's scaling for correlation functions   Let us now consider the correlation function of the block spin system:

$$G(\vec{r}_\ell, t_\ell) = \langle S_I S_J \rangle - \langle S_I \rangle \langle S_J \rangle$$

where $\vec{r}_\ell$ is the vector of the relative distance between the centers of the I-th and J-th block (measured in units of ℓa, as stated before).

We want now to see how this correlation length is related to the one of the original system $G(\vec{r}, t)$.

From (7.18) we have $m_\ell = h_\ell \ell^{-d}/h = \ell^{y_h - d}$, and using also (7.17) we get:

$$G(\vec{r}_\ell, t_\ell) = \langle S_I S_J \rangle - \langle S_I \rangle \langle S_J \rangle = \frac{1}{\ell^{2(y_h - d)} \cdot \ell^{2d}} \sum_{i \in I} \sum_{j \in J} (\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle) =$$

$$= \frac{\ell^{2d}}{\ell^{2(y_h - d)} \cdot \ell^{2d}} (\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle) = \ell^{2(d - y_h)} G(\vec{r}, t)$$

Introducing also the dependence on h, we have:

$$G\left(\frac{\vec{r}}{\ell}, t\ell^{y_t}, h\ell^{y_h}\right) = \ell^{2(d - y_h)} G(\vec{r}, t, h)$$

which is an equivalent of (7.1) for the correlation function.

Again, we can remove the dependence on t setting $\ell = t^{-1/y_t}$ so that:

$$G(\vec{r}, t, h) = t^{2\frac{d - y_h}{y_t}} G\left(\vec{r} t^{1/y_t}, 1, h t^{-y_h/y_t}\right)$$

Now, $\vec{r}$ scales with ℓ as all the lengths of our system, and since we have set $\ell = t^{-1/y_t}$ we have $rt^{1/y_t} = 1$. Therefore:

$$G\left(\vec{r} t^{1/y_t}, 1, h t^{-y_h/y_t}\right) = \left(rt^{1/y_t}\right)^{-2(d - y_h)} F_G\left(\vec{r} t^{1/y_t}, h t^{-y_h/y_t}\right)$$

where in the last step we have defined $F_G(a, b) = G(a, b, 1)$. Therefore:

$$G(\vec{r}, t, h) = \frac{1}{r^{2(d - y_t)}} F_G\left(\vec{r} t^{1/y_t}, h t^{-y_h/y_t}\right)$$

With the choice $\ell = t^{-1/y_t}$ we further have that the correlation length scales as:

$$\xi = \ell \xi_\ell = \xi_\ell t^{-1/y_t}$$

so we also have:

$$\nu = \frac{1}{y_t}$$

This equation together with (7.20) leads to the *hyperscaling relation*:

$$2 - \alpha = \nu d$$

Hyperscaling relations are known to be less robust than the normal scaling relations between critical exponents (for example, for Hamiltonians with long-ranged power law interactions hyperscaling relations don't hold).

Chapter 8

The Renormalization Group

Kadanoff's argument for the Ising model allows us to explain the scaling form of the free energy density and of the correlation length near the critical point. However, as we have seen, it is unable to predict the values of the scaling exponents y_t and y_h (and thus ultimately of the critical exponents), nor can it explain why universality occurs.

We will see that these problems are solved with the introduction of the *Renormalization Group* (done by K. G. Wilson at the beginning of the '70s), which we will call simply "RG" from now on for the sake of simplicity.
The RG is based upon the correct "intuition" of Kadanoff's argument that the coupling constants of a Hamiltonian change if we coarse-grain the system (or in other words we "look" at it on different spatial scales); however, this "intuition" strictly speaking is not correct since we have seen that in Kadanoff's procedure we assume that after the coarse-graining procedure the Hamiltonian of the system has *exactly* the same form: as we will see, this is not true in general because new terms can appear after we coarse-grain the system.

8.1 Basic ideas of the Renormalization Group

The application of the RG consists in the *recursive* enactment of a procedure made of two principal steps:

1. The first is an actual realization of a **coarse graining** procedure, also called *decimation*, like the one introduced by Kadanoff for the Ising model; in general this procedure must integrate the degrees of freedom of the system on scales of linear dimension ℓa which must be much larger than the characteristic microscopic scale a of the system but also much smaller than the correlation length ξ: $a \ll \ell a \ll \xi$.
   After the decimation, we are left with a new effective Hamiltonian.

2. The second consists in the **rescaling** (or *renormalization*) of the system, so that the "new" microscopic scale of the system is again a but in the "new units of measure". In other words we rescale the distances dividing them by ℓ:

$$\vec{r}_{\text{new}} = \vec{r}/\ell$$

   As we have seen, this means that the new correlation length $\xi_{\text{new}} = \xi/\ell$ is smaller than the original one, so our system is farther from criticality after the decimation. This way, the whole procedure can be seen (as realised by Kadanoff) as a transformation $K \to K'$ of the coupling constants of the Hamiltonian of the system.

To make an example, suppose we are given a Hamiltonian $\mathcal{H}[K]$ which depends on an arbitrary number of coupling constants $[K] = \vec{K} = (K_1, K_2, \dots)$ (in the case of an Ising model with

nearest-neighbour interaction and an external field there are only two coupling constants, $K = K_1$ and $h = K_2$). For what we have just stated the action of the RG can be expressed as a transformation of the form:

$$[K'] = R_\ell[K] \tag{8.1}$$

where R_ℓ is called *RG transformation*, also referred to as *recursion relation*. We suppose that the function R_ℓ is analytic (no matter how complicated it may be).

The set of transformations R_ℓ form a *semigroup*[1], because if we subsequently apply two transformations R_{ℓ_1} and R_{ℓ_2} on two different length scales ℓ_1 and ℓ_2 we have:

$$[K'] = R_{\ell_1}[K] \qquad\qquad [K''] = R_{\ell_2}[K'] = R_{\ell_2}R_{\ell_1}[K] \qquad \Rightarrow \qquad R_{\ell_2\ell_1}[K] = R_{\ell_2}R_{\ell_1}[K] \quad (8.2)$$

but in general the inverse of a given transformation R_ℓ does not exist.

There is no general way to construct the function R_ℓ: depending on the system and on the case considered we can choose different ways to carry out the decimation, and in general (as we will see) for a given system many different RG transformations can be built. In general such procedures can be done either in coordinate space (*real space Renormalization Group*) or in Fourier space (*momentum shell Renormalization Group*).

In terms of the coupling constants $[K]$ the partition function of the original system is:

$$Z_N[K] = \operatorname{Tr} e^{-\beta\mathcal{H}[K]}$$

while the free energy density (for the sill finite-sized system):

$$f_N[K] = -\frac{k_B T}{N}\ln Z_N[K]$$

Now, if the RG transformation integrates the degrees of freedom on the spatial scale ℓa then the number of degrees of freedom will decrease by a factor ℓ^d, if d is the dimensionality of the system; in other words, after the RG transformation R_ℓ we are left with $N' = N/\ell^d$ degrees of freedom.

Considering Kadanoff's block transformation (but the essence of our statements is valid in general, of course provided the trivial generalizations), the decimation is performed doing a "partial trace" of the degrees of freedom $\{S_i\}$ with the constraints that the block spins $\{S_I\}$ have fixed values (of course determined in the way we choose). Formally, we can write:

$$e^{-\beta\mathcal{H}'_{N'}([K'],S'_I)} = \operatorname{Tr}'_{\{S_i\}} e^{-\beta\mathcal{H}_N([K],S_i)} = \operatorname{Tr}_{\{S_i\}} P(S_i, S'_I) e^{-\beta\mathcal{H}_N([K],S_i)} \tag{8.3}$$

where Tr' is the constrained trace, while $P(S_i, S'_I)$ is the *projection operator*, which "incorporates" the constraints and allows us to write an unconstrained trace. In general this operator must be built "by hand".

For example, in the case of Kadanoff's block transformation we can assign the block spins S'_I their values with the "majority rule", i.e. we build (hyper)cubic blocks of side $(2\ell+1)a$ (so that each one contains an odd number of spins) and set:

$$S'_I = \operatorname{sgn}\left(\sum_{i\in I} S_i\right)$$

then $S'_I = \pm 1$ and the projection operator can be written as:

$$P(S_i, S'_I) = \prod_I \delta\left[S'_I - \operatorname{sgn}\left(\sum_{i\in I} S_i\right)\right]$$

---

[1]What we are now studying should be called *Renormalization Semigroup*, but it is simply known as *group*.

As we can see, doing an unconstrained trace with this operator is equivalent to performing the constrained trace.

The decimation procedure must in general satisfy three requirements:

- $e^{-\beta \mathcal{H}'_{N'}([K'], S'_I)} \geq 0$, so that $\mathcal{H}'$ can be indeed considered an effective Hamiltonian. From (8.3) we see that this requirement is satisfied if $P(S_i, S'_I) \geq 0$

- The effective Hamiltonian $\mathcal{H}'$ must have the same symmetry properties of the original one. This means (and this is the great improvement with respect to Kadanoff's argument) that the decimation can make some new terms appear in the coarse-grained Hamiltonian, as long as they respect the same symmetries of the original system. In more "formal" words, if $K_m = 0$ in $\mathcal{H}_N$ but its relative term is allowed by the symmetry group of $\mathcal{H}_N$ itself, then we can have $K'_m \neq 0$ in $\mathcal{H}'_{N'}$.
  For example, in 8.5.3 we will see that for a bidimensional Ising model with nearest-neighbour interactions and $H = 0$, after the decimation new four-spin interaction terms can appear, and they are still invariant under parity (which is the symmetry group of the initial Hamiltonian).
  In order to satisfy this requirement, also the projection operator $P(S_i, S'_I)$ must satisfy the symmetries of the original Hamiltonian

- The last requirement is that the decimation leaves invariant the partition function (*not* the Hamiltonian!):
$$Z_{N'}[K'] = Z_N[K]$$
  From (8.3) we see that this is true if:
$$\text{Tr}_{\{S'_I\}} P(S_i, S'_I) = \sum_{S'_I} P(S_i, S'_I) = 1$$

From the last requirement we can also see how the free energy density of the system changes under the action of the RG:

$$\frac{1}{N} \ln Z_N[K] = \frac{\ell^d}{N\ell^d} \ln Z_{N'}[K'] = \ell^{-d} \frac{1}{N'} \ln Z_{N'}[K'] \qquad \Rightarrow \qquad f_N[K] = \ell^{-d} f_{N'}[K']$$

which is the scaling form of the free energy density as obtained by Kadanoff (equation (7.19)).

8.2 Singular behaviour in the Renormalization Group

We have stated that in general the RG transformations R_ℓ are analytic, so we might ask: where does the singular behaviour of a system near a critical point come from?
Like what we have seen in 5.1.1 this occurs in the thermodynamic limit, which in this case is obtained when we apply the RG transformation an *infinite* number of times.

In general, after the n-th iteration of the RG the coarse-graining length of the system will be ℓ^n and the coupling constants $[K^{(n)}]$. As n increases the "vector" of coupling constants describes a "trajectory" in the space of all the possible coupling constants, often called *Hamiltonian space* or *theory space*; we call *RG flow* the set of all the trajectories that start from different initial conditions, i.e. different initial Hamiltonians.
In general, these trajectories can form strange attractors or complex limit cycles; however, it is almost always found that they are simply attracted towards or ejected from fixed points (cases where this doesn't occur are really exotic), so in the following we will assume that the RG flow only exhibits fixed points.

The study of the properties of the RG flow near these fixed points is crucial, since as we will see it is that that will allow us to actually explain universality and predict the values of the critical exponents.

We therefore proceed to study such points.

8.2.1 Fixed points of the Renormalization Group flow

Suppose we know R_ℓ (in 8.5 we will see explicitly how to construct them). If $[K^*]$ is a fixed point of the RG flow, by definition we have:

$$R_\ell[K^*] = [K^*]$$

Then, in general from the Hamiltonian of a system we can determine the correlation length ξ, and if $[K'] = R_\ell[K]$ we know that:

$$\xi[K'] = \xi[K]/\ell$$

Therefore for a fixed point we have:

$$\xi[K^*] = \xi[K^*]/\ell$$

which implies that $\xi[K^*]$ is either zero or infinity. A fixed point with $\xi = \infty$ is called *critical*, while if $\xi = 0$ *trivial*.

Clearly, every fixed point $[K^*]$ can have its own *basin of attraction*, i.e. a set of points that under the action of the RG flow tends to $[K^*]$.

An important result concerning the basin of attraction of critical fixed points is the following:

Theorem. *The correlation length is infinite for every point in the basin of attraction of a critical fixed point of the RG flow.*

Proof. Call $[K]$ the initial set of coupling constants, after n iterations of the RG the correlation length of the system will be such that:

$$\xi[K] = \ell\xi[K^{(1)}] = \cdots = \ell^n\xi[K^{(n)}] \qquad \Rightarrow \qquad \xi[K] = \ell^n\xi[K^{(n)}]$$

If we now take the limit $n \to \infty$ the right hand side diverges if $K^{(n)} \to K^*$, i.e. if $[K]$ belongs to the basin of attraction of $[K^*]$. Therefore, $\xi[K] = \infty$. $\qquad\square$

The basin of attraction of a critical fixed point is also called *critical manifold*.

We can argue that the fact that *all* the points of a critical manifold flow towards the *same* fixed point (i.e. the same Hamiltonian) is the basic mechanism on which universality is based upon, but this is by no means a complete explanation, since universality involves the behaviour of systems near a critical point and we still have said nothing about that.

We can however note the following fact: starting from any point in theory space, iterating the RG transformation and identifying the fixed point towards which the system flows, the phase of the original point in theory space (i.e. in the phase diagram) will be described by this fixed point. Therefore, every phase of the system is "represented" by a fixed point of the RG flow.

As we will later see (and will become clearer in 8.2.3, 8.3 and 8.5), critical fixed points describe the singular critical behaviour while trivial fixed points are related to the bulk phases of the system: therefore, the knowledge of the location and nature of the fixed points of the RG flow can give us hints on the structure of the phase diagram of the system, and the behaviour of the flow near critical fixed points allows us to calculate the values of the critical exponents.

8.2.2 Renormalization Group flow near a fixed point

In order to study the behaviour of the RG flow near a fixed point $\vec{K}^*$, let us take a slight perturbation from it, namely we set:

$$\vec{K} = \vec{K}^* + \delta\vec{K}$$

where $\delta\vec{K}$ is a small displacement. Applying the RG flow, in components we will have:

$$K'_j = R_\ell(\vec{K}^* + \delta\vec{K})_j = K_j^* + \sum_i \frac{\partial K'_j}{\partial K_i}\bigg|_{K^*} \delta K_i + O(\delta K^2) \tag{8.4}$$

Neglecting all the terms beyond the linear ones, we can write the action of the linearised RG transformation in terms of the displacements $\delta\vec{K}$ and $\delta\vec{K}'$ as:

$$\delta\vec{K}' = \boldsymbol{T}\delta\vec{K} \qquad \text{where} \quad \boldsymbol{T}_{ij} = \frac{\partial K'_j}{\partial K_i}\bigg|_{K^*}$$

Of course $\boldsymbol{T}$ is a square matrix but in general it is *not* symmetric, so it is not diagonalizable and its eigenvalues can be complex (and we also must distinguish between left and right eigenvectors).
However we suppose $\boldsymbol{T}$ to be symmetric (which, as before, is almost always the case) so that it can be diagonalized.
If we call $\lambda_\ell^{(\sigma)}$ and $\vec{e}^{(\sigma)}$ the σ-th eigenvalue and relative eigenvector of $\boldsymbol{T}^{(\ell)}$ (where we are explicitly writing the length scale of the decimation), in components the action $\boldsymbol{T}^{(\ell)}$ will be:

$$\boldsymbol{T}_{ij}^{(\ell)} e_j^{(\sigma)} = \lambda_\ell^{(\sigma)} e_i^{(\sigma)} \tag{8.5}$$

From the semigroup property (8.2) of the RG transformation we have:

$$\boldsymbol{T}^{(\ell)}\boldsymbol{T}^{(\ell')} = \boldsymbol{T}^{(\ell\ell')}$$

and so from (8.5):

$$\lambda_\ell^{(\sigma)}\lambda_{\ell'}^{(\sigma)} = \lambda_{\ell\ell'}^{(\sigma)}$$

This is a functional equation which can be solved in the following way: if we write the eigenvalues explicitly as functions of ℓ, namely $\lambda_\ell^{(\sigma)} = \lambda(\ell)^{(\sigma)}$, then differentiating with respect to ℓ' we will have

$$\lambda^{(\sigma)}(\ell)\lambda'^{(\sigma)}(\ell') = \ell\lambda'^{(\sigma)}(\ell\ell')$$

where with λ' we mean that λ has been differentiated with respect to its argument. Setting now $\ell' = 1$ and defining $\lambda'^{(\sigma)}(1) = y_\sigma^{-1}$ we get:

$$\frac{\lambda^{(\sigma)}(\ell)}{\lambda'^{(\sigma)}(\ell)} = \ell y_\sigma$$

which is easily solved to give:

$$\lambda_\ell^{(\sigma)} = \ell^{y_\sigma}$$

where, as we have defined it, y_σ is a number (to be determined) independent of ℓ.
To see how $\delta\vec{K}$ changes under the action of $\boldsymbol{T}$ let us find out how its components along the directions determined by the eigenvectors $\vec{e}^{(\sigma)}$ change[2]. In other words, we write:

$$\delta\vec{K} = \sum_\sigma a^{(\sigma)}\vec{e}^{(\sigma)} \qquad\qquad a^{(\sigma)} = \vec{e}^{(\sigma)} \cdot \delta\vec{K}$$

---

[2]Note: since $\boldsymbol{T}^{(\ell)}$ is diagonalizable its eigenvectors are orthonormal.

and applying $\boldsymbol{T}^{(\ell)}$:

$$\delta \vec{K}' = \boldsymbol{T}\delta\vec{K} = \boldsymbol{T}\sum_{\sigma} a^{(\sigma)}\vec{e}^{\,(\sigma)} = \sum_{\sigma} a^{(\sigma)}\lambda_\ell^{(\sigma)}\vec{e}^{\,(\sigma)} := \sum_{\sigma} a'^{(\sigma)}\vec{e}^{\,(\sigma)}$$

where in the last step we have defined the components $a'^{(\sigma)}$ of $\delta\vec{K}'$ along $\vec{e}^{\,(\sigma)}$. We therefore see that the behaviour of $\delta\vec{K}$ along the eigenvectors $\vec{e}^{\,(\sigma)}$ depends on the magnitudes of the eigenvalues $\lambda_\ell^{(\sigma)}$. In particular, we can distinguish three cases:

$|\lambda_\ell^{(\sigma)}| > 1$: this implies that $a'^{(\sigma)}$ grows

$|\lambda_\ell^{(\sigma)}| = 1$: this implies that $a'^{(\sigma)}$ doesn't change (its behaviour can depend on the higher orders in the expansion (8.4) that we have neglected)

$|\lambda_\ell^{(\sigma)}| < 1$: this implies that $a'^{(\sigma)}$ shrinks

These three cases are given, respectively, the following terminology:

- *relevant* eigenvalues/directions/eigenvectors

- *marginal* eigenvalues/directions/eigenvectors

- *irrelevant* eigenvalues/directions/eigenvectors

The number of irrelevant directions of a fixed point is equal to the dimension of its critical manifold, while the number of relevant directions is equal to its codimension.
Let us note that the eigenvalues and their possible relevance depend on the matrix $\boldsymbol{T}$, which in turn depends on the fixed point considered: this means that the terms "relevant", "irrelevant" or "marginal" must *always* be specified with respect to the particular fixed point considered.

8.2.3 Global properties of the Renormalization Group flow

We now want to show qualitatively that the global behaviour of the RG flow determines the phase diagram of the system.
Since as we have already stated previously every fixed point of the RG flow can correspond to a phase of the system, it is very important to classify the fixed points of a possible RG flow. This classification can be done using the codimension c and the correlation length ξ of the point. We shall now see some results.

- Fixed points with $c = 0$ and $\xi = 0$ are called *sinks* (since they have no relevant directions, so all the points in their neighbourhoods flow towards them), and correspond to stable bulk phases of the system; the nature of the coupling constants at the sink characterizes the relative phase.
  For example, a three-dimensional Ising model with nearest-neighbour interaction and in an external field turns out to have two sinks at $(H = +\infty, T = 0)$ and $(H = -\infty, T = 0)$ (note that considering K a coupling constant is equivalent to considering T as such), which correspond to the fact that, for all temperatures, in a positive (negative) external field the system has a positive (negative) magnetization

- There are two types of fixed points with $c = 1$, called *discontinuity* and *continuity fixed points*. In the case of the Ising model, both have $\xi = 0$ (but in other models we can have also fixed points with $c = 1$ and $\xi = \infty$).
  The former correspond to first order transitions where one of the order parameters changes abruptly. For example, in the same case as before the line of points $(H = 0, T < T_c)$ flows under the RG towards the discontinuity point $(H = 0, T = 0)$.

The latter represents a phase of the system, generally a disordered one like the paramagnetic phase of the Ising model; in this case the line $(H = 0, T > T_c)$ flows towards the continuity point $(H = 0, T = \infty)$. These points are generally not really interesting

– Fixed points with $c \geq 2$ can describe either points of multiple phase coexistence (if $\xi = 0$) or multicritical points (if $\xi = \infty$).
In the simplest case, i.e. $c = 2$, such fixed points correspond to triple points if $\xi = 0$ or critical points if $\xi = \infty$. In each case, a useful way to interpret the presence of two relevant directions is that these represent the two variables that must be tuned in order to place the system in the appropriate point (for example in the case of a magnet we must set $H = 0$ and $T = T_c$)

8.3 Universality in the Renormalization Group

Let us now see how the formalism of the RG can explain universality.

Suppose we start from a system with a Hamiltonian $\mathcal{H}$ which depends on some coupling constants $[K]$; suppose also that we can write a RG transformation which in general gives rise to no more than D couplings[3]. Under the action of the RG the initial physical Hamiltonian will move in the D-dimensional theory space and follow the RG flow.
Let us call $\mathcal{H}^*$ the fixed point towards which $\mathcal{H}$ tends, and assume it has only one relevant direction and $D-1$ irrelevant ones; linearising the flow near $\mathcal{H}^*$ we therefore identify $D-1$ linearly independent vectors which constitute the hyperplane tangent in $\mathcal{H}^*$ to its basin of attraction (i.e. the linearisation of the critical manifold near $\mathcal{H}^*$). In general, if we "zoom out" and "look" also in regions far from the fixed point $\mathcal{H}^*$, in theory space there will be a $D-1$-dimensional critical manifold $\mathcal{C}$.
Let us now consider a generic model in theory space; remembering that in general $\beta = 1/k_B T$ is included in the definition of the coupling constants K_i, if we change the temperature of the system all the K_i-s will change and thus the model will describe a trajectory $\mathcal{P}$ in the D-dimensional theory space, called *physical subspace* (because we can move along it by changing a physically accessible parameter like the temperature).
To make an explicit example, let us consider an Ising model with nearest- and next-nearest-neighbour interactions in the absence of any external field, so that:

$$-\beta\mathcal{H} = K_1 \sum_{\langle ij \rangle} S_i S_j + K_2 \sum_{i,j=\text{n.n.n.}} S_i S_j$$

In this case the physical subspace can be a straight line once the ratio K_1/K_2 has been fixed, and this line goes from $K_1 = K_2 = 0$ (when $T = \infty$) to $K_1 = K_2 = \infty$ (when $T = 0$).
The physical subspace will in general intersect the critical manifold $\mathcal{C}$ in some point $[K_c]$: *this is the critical point of our system*, and as it must be we have reached it by only varying its temperature.

Now, if we consider two different physical systems they will be characterized by two different physical subspaces $\mathcal{P}_1$ and $\mathcal{P}_2$ in theory space. In general, they will intersect the critical manifold $\mathcal{C}$ in different points, and so they will have different critical temperatures[4]; however, under the action of the RG they will flow towards the *same* critical fixed point, and since (as we are going to show explicitly in the following section) the critical behaviour of a system is determined by the properties of the RG flow near a fixed critical point, *these two systems will*

---

[3]For what we want to show it is not essential that D is finite, but we suppose it to be so for the sake of simplicity.
[4]This makes sense since we have already noticed that the exact value of the critical temperature depends strongly on the microscopic details of the system.
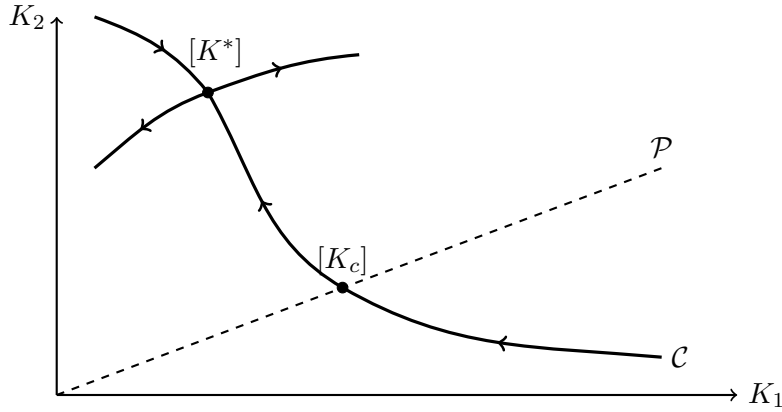
Figure 8.1: RG flow for the considered Ising model

behave identically near their critical points.
This is how universality is explained within the RG.

In the case of the Ising model we are considering, the situation can be represented in the two-dimensional theory space as shown in figure 8.1.

8.4   The origins of scaling and critical behaviour

Let us consider a fixed point of the RG flow of a generic system, and assume that it has two relevant directions corresponding to the coupling constants T[5], the temperature, and H, the external field. We suppose that T and H are transformed under the RG as:

$$T' = R_\ell^T(T, H) \qquad\qquad H' = R_\ell^H(T, H)$$

where R_ℓ^T and R_ℓ^H are analytic functions given by the coarse graining procedure.
The fixed points (T^*, H^*) of the flow will be given by the solutions of:

$$T^* = R_\ell^T(T^*, H^*) \qquad\qquad H^* = R_\ell^H(T^*, H^*)$$

Linearising the transformation around (T^*, H^*), in terms of the reduced variables $t = (T - T^*)/T^*$ and $h = (H - H^*)/H^*$ we have:

$$\begin{pmatrix} t' \\ h' \end{pmatrix} = \boldsymbol{T} \begin{pmatrix} t \\ h \end{pmatrix}$$

where:

$$\boldsymbol{T} = \begin{pmatrix} \partial R_\ell^T/\partial T & \partial R_\ell^T/\partial H \\ \partial R_\ell^H/\partial T & \partial R_\ell^H/\partial H \end{pmatrix}_{|T^*, H^*}$$

As previously stated we suppose $\boldsymbol{T}$ to be diagonalizable. We therefore write its eigenvalues as:

$$\lambda_\ell^t = \ell^{y_t} \qquad\qquad \lambda_\ell^h = \ell^{y_h}$$

Note that we can always do that, it is just a simple definition. In other words, we are defining y_t and y_h as:

$$y_t = \frac{\ln \lambda_\ell^t}{\ln \ell} \qquad\qquad y_h = \frac{\ln \lambda_\ell^h}{\ln \ell} \qquad\qquad (8.6)$$

---

[5]We have already stated that considering K as a coupling constant is equivalent to considering T as such.

This way we can write:

$$\begin{pmatrix} t' \\ h' \end{pmatrix} = \begin{pmatrix} \lambda_\ell^t & 0 \\ 0 & \lambda_\ell^h \end{pmatrix} \begin{pmatrix} t \\ h \end{pmatrix} \qquad \Rightarrow \qquad \begin{pmatrix} t' \\ h' \end{pmatrix} = \begin{pmatrix} \ell^{y_t} t \\ \ell^{y_h} h \end{pmatrix}$$

After n iterations we will have:

$$t^{(n)} = (\ell^{y_t})^n \, t \qquad\qquad h^{(n)} = (\ell^{y_h})^n \, h$$

and since ingeneral we know that $\xi(t', h') = \xi(t, h)/\ell$:

$$\xi(t, h) = \ell^n \xi(\ell^{n y_t} t, \ell^{n y_h} h) \tag{8.7}$$

This is the scaling law of the correlation length.
From this we can determine the critical exponent ν; in fact, setting $h = 0$ and choosing ℓ so that $t\ell^{n y_t} = b$ with b a positive real number[6], we have:

$$\ell^n = \left(\frac{b}{t}\right)^{1/y_t} \qquad \Rightarrow \qquad \xi(t) = \left(\frac{t}{b}\right)^{-1/y_t} \xi(b, 0)$$

Since in general $\xi \sim t^{-\nu}$, we get:

$$\nu = \frac{1}{y_t}$$

This is an extremely important result! In fact, from (8.6) we see that once the RG transformation R_ℓ is known, y_t is straightforward to compute and so we are actually able to calculate ν and predict its value!
We can do even something more (including giving y_h a meaning) from the scaling law of the free energy density. After n iterations of the RG we have:

$$f(t, h) = \ell^{-nd} f(t^{(n)}, h^{(n)}) = \ell^{-nd} f(\ell^{n y_t} t, \ell^{n y_h} h)$$

and choosing ℓ so that $\ell^{n y_t} t = b^{y_t}$, then:

$$f(t, h) = t^{d/y_t} b^{-d} f(b^{y_t}, b^{y_h} h/t^{y_h/y_t})$$

Comparing this to the form (7.12) of the scaling law we see that:

$$2 - \alpha = \frac{d}{y_t} \qquad\qquad \Delta = \frac{y_h}{y_t}$$

Irrelevant variables We may also include the irrelevant variables in the scaling law of f:

$$f(t, h, k_3, k_4, \dots) = \ell^{-d} f(t\lambda_\ell^t, h\lambda_\ell^h, k_3 \lambda_\ell^{(3)}, k_4 \lambda_\ell^{(4)} \dots)$$

where $\lambda_\ell^t, \lambda_\ell^h > 1$ and $\lambda_\ell^{(i \geq 3)}, \cdots < 1$. After n iterations we have:

$$f(t, h, k_3, k_4, \dots) = \ell^{-dn} f(t\ell^{n y_t}, h\ell^{n y_h}, k_3 \ell^{n y_3}, k_4 \ell^{n y_4}, \dots)$$

where $y_{i>3} < 0$ in accordance with the fact that $\lambda_\ell^{(i \geq 3)} < 1$.
Setting $\ell^{n y_t} t = b$ we have:

$$f(t, h, k_3, k_4 \dots) = b^{-d} t^{d/y_t} f(b^{y_t}, b^{y_h} h t^{-y_h/y_t}, b^{y_3} k_3 t^{-y_3/y_t}, k_4 t^{-y_4/y_t}, \dots)$$

---

[6]Remember that the value of ℓ is not fixed, so we can choose the one we prefer; in this case we are making this choice because ℓ does not necessarily have to be an integer.

$$S_1 \quad S_2 \quad S_3 \quad S_4 \quad S_5 \quad S_6$$

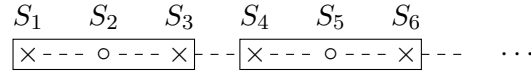Figure 8.2: Decimation to $N/3$ spins.
The spins over which we sum are indicated with a cross.

For $t \to 0$ the terms involving the irrelevant variables become vanishingly small, so we get:

$$f(t, h, k_3, k_4, \dots) = t^{d/y_t} b^{-d} f(b^{y_t}, b^{y_h} h t^{-y_h/y_t}, 0, 0, \dots)$$

Note that in the last step we have implicitly assumed that f is analytic in the limit $k_{i \geq 3} \to 0$. This assumption is however frequently false! When this happens, i.e. when the free energy density is singular in the limit $k_j \to 0$ for a particular irrelevant variable k_j, that variable is termed *dangerous irrelevant variable*.

For example, considering the Landau free energy of the Ising model obtained as a saddle-point approximation of the general functional partition function (equation (6.31)):

$$\mathcal{L} = \int \left(\frac{a}{2} t m^2 + \frac{b}{4} m^4 + \frac{k}{2} \left(\vec{\nabla} m \right)^2 - hm \right) d^d \vec{r}$$

the parameter b of the quartic term is a dangerous irrelevant variable (we have seen in 6.6.3 that problems arise when we try to treat it as a perturbative parameter).

8.5 Renormalization Group in coordinate space

We now want to see how we can build the RG transformation R_ℓ from the coarse graining procedure.

We will first use two different decimation techniques for a one-dimensional Ising model, and then see what happens for $d = 2$. As we have partially anticipated, we will see that as soon as we will no longer work in the "easy" case $d = 1$ things will get really complicated because at every iteration of the RG new interactions appear, and so the problem becomes rapidly impossible to handle; we will therefore be forced to resort to approximations.

8.5.1 Decimation to $N/3$ spins for a one-dimensional Ising model with $H = 0$

Let us consider a one-dimensional Ising model with nearest-neighbour interaction and periodic boundary conditions, without any external field ($H = 0$).

We choose to apply the coarse-graining procedure to our system by grouping spins in blocks of three; this way the $(i+1)$-th block (with $i = 0, 1, 2, \dots$) will be constituted by the spins S_{1+3i}, S_{2+3i} and S_{3+3i} (for example, the first block is $[S_1, S_2, S_3]$, the second one $[S_4, S_5, S_6]$ and so on). In order to define the new block spin we could use the majority rule, but we further simplify the problem requiring that the new block spin S'_I coincides with the central spin S_{2+3i} of the block. In other words, for every block we set:

$$P(S'_I; S_{1+3i}, S_{2+3i}, S_{3+3i}) = \delta_{S'_I, S_{2+3i}}$$

(for example for the first block we have $P(S'_1; S_1, S_2, S_3) = \delta_{S'_1, S_2}$).

Therefore, the coarse-graining procedure consists in summing over the spins at the boundaries of the blocks and leaving untouched the central ones (in figure 8.2 we represent the situation, where the spins over which we sum are indicated by a cross $\times$ and the ones leaved untouched by a circle $\circ$).

Now, using the notation introduced in 8.1 for the general theory, we have:

$$e^{-\beta\mathcal{H}'} = \text{Tr}_{\{S_i\}}\, P(S'_I, S_i)e^{-\beta\mathcal{H}} = \text{Tr}_{\{S_i\}} \prod_I \delta_{S'_I, S_{2+3i}} \cdot e^{K\sum_j S_j S_{j+1}} =$$

$$= \sum_{\{S_i=\pm 1\}} \delta_{S'_1, S_2} \delta_{S'_2, S_5} \cdots e^{KS_1 S_2} e^{KS_2 S_3} e^{KS_3 S_4} e^{KS_4 S_5} \cdots =$$

$$= \sum_{\{S_i=\pm 1\}} e^{KS_1 S'_1} e^{KS'_1 S_3} e^{KS_3 S_4} e^{KS_4 S'_2} \cdots$$

Let us therefore see how to perform the sum on the first two blocks, $[S_1, S_2, S_3] - [S_4, S_5, S_6]$:

$$\sum_{S_3, S_4=\pm 1} e^{KS'_1 S_3} e^{KS_3 S_4} e^{KS_4 S'_2} \tag{8.8}$$

From the definitions of $\cosh$ and $\sinh$ we can write:

$$e^{KS_a S_b} = \cosh K (1 + t S_a S_b) \qquad \text{where} \qquad t = \tanh K \tag{8.9}$$

so that (8.8) becomes:

$$\sum_{S_3, S_4=\pm 1} (\cosh K)^3 (1 + t S'_1 S_3)(1 + t S_3 S_4)(1 + t S_4 S'_2)$$

Expanding the product and keeping in mind that $S_i^2 = +1$, we get:

$$(1 + t S'_1 S_3)(1 + t S_3 S_4)(1 + t S_4 S'_2) =$$
$$= 1 + t S'_1 S_3 + t S_3 S_4 + t S_4 S'_2 + t^2 S'_1 S_4 + t^2 S'_1 S_3 S_4 S'_2 + t^2 S_3 S'_2 + t^3 S'_1 S'_2$$

and clearly all the terms containing S_3 or S_4 (or both) vanish when we perform the sum $\sum_{S_3, S_4=\pm 1}$. Therefore, the result of the partial sum (8.8) for the first two blocks is:

$$2^2 (\cosh K)^3 \prod_I (1 + t^3 S'_1 S'_2)$$

(where 2^2 comes from the fact that the constant terms 1 and $t^3 S'_1 S'_2$ must be summed 2^2 times, two for the possible values of S_3 and two for S_4).
Therefore, the partition function of the block spin system will be:

$$Z_{N'}[K'] = \text{Tr}_{\{S'_I\}} 2^{2N'} \cosh^{3N'} K (1 + t^3 S'_I S'_{I+1}) \tag{8.10}$$

where $N' = N/3$ is the new number of spin variables.
However, we know that in general $Z_{N'} = \text{Tr}_{\{S'_I\}} e^{-\beta\mathcal{H}'}$, so let us try to write (8.10) in this form.
We have:

$$2^2 (\cosh K)^3 (1 + t^3 S'_I S'_{I+1}) = 2^2 (\cosh K)^3 \frac{\cosh K'}{\cosh K'}(1 + t^3 S'_I S'_{I+1})$$

and renaming $t^3 := t'$, so that:

$$(\tanh K)^3 = \tanh K' \tag{8.11}$$

this term becomes:

$$2^2 \frac{(\cosh K)^3}{\cosh K'} \cosh K'(1 + t' S'_I S'_{I+1}) = 2^2 \frac{(\cosh 1 K)^3}{\cosh K'} e^{K' S'_I S'_{I+1}}$$

where we have used (8.9). Therefore:

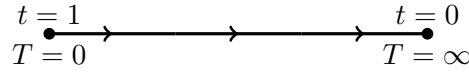$$2^2 \cosh^3 K (1 + t^3 S'_I S'_{I+1}) = e^{2\ln 2 + \ln \frac{(\cosh K)^3}{\cosh K'} + K' S'_I S'_{I+1}}$$

Figure 8.3: RG Flow of (8.11)

and we can write:

$$-\beta \mathcal{H}'(\{S_I'\}) = N' g(K, K') + K' \sum_I S_I' S_{I+1}'$$

where:

$$g(K, K') = 2 \ln 2 + \ln \frac{(\cosh K)^3}{\cosh K'}$$

The new effective Hamiltonian has therefore the same form of the original one with the redefined coupling constant K', and exhibits also a new term ($g(K, K')$) independent of the block spins.

Let us note that (8.11) is the recursion relation we are looking for:

$$K' = \tanh^{-1}(\tanh^3 K)$$

Rewritten in the form $t' = t^3$, its fixed points are given by:

$$t^* = t^{*3}$$

whose solutions are $t^* = 0$ and $t^* = 1$ (the case $t^* = -1$ is neglected because $K > 0$ and so $\tanh K > 0$). After all, however, $\tanh K \to 0^+$ if $K \to 0^+$ (i.e. $T \to \infty$) and $\tanh K \to 1^-$ if $K \to \infty$ (i.e. $T \to 0^+$): in other words, the fixed point $t^* = 0$ corresponds to $T = \infty$ while $t^* = 1$ to $T = 0$.

Since $\tanh K < 1 \quad \forall K \in \mathbb{R}$, starting from any initial point $t_0 < 1$ the recursion relation $t' = t^3$ makes t smaller every time, moving it towards the fixed point $t^* = 0$. We can thus conclude that $t^* = 1$ is an unstable fixed point while $t^* = 0$ is stable, as graphically represented in figure 8.3. Note that the fact that the flow converges towards $T = \infty$ means that on large spatial scales the system is well described by a Hamiltonian with a high effective temperature, and so the system will *always* be in the paramagnetic phase (except when $T = 0$).

Let us now see how the correlation length transforms.
We know that in general, if the decimation reduces the number of spins by a factor b (in the case we were considering above, $b = 3$) we have to rescale distances accordingly, and in particular:
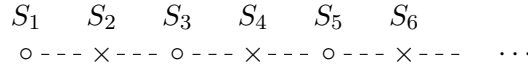
$$\xi(t') = \xi(t)/b$$

where in general $t' = t^b$. Since b is in general arbitrary, we can choose $b = \text{const.}/\ln t$ and thus:

$$\xi(t') = \xi(t^b) = \xi(e^{b \ln x}) = \xi(e^{\text{const.}}) = \left(\frac{\text{const.}}{\ln t}\right)^{-1} \xi(t)$$

Therefore:

$$\xi(t) = \frac{\text{const.}}{\ln t} \sim \frac{1}{\ln \tanh K}$$

which is the exact result we have found at the end of 5.4.4.

$$S_1 \quad S_2 \quad S_3 \quad S_4 \quad S_5 \quad S_6$$

$$\circ \text{ - - - } \times \text{ - - - } \circ \text{ - - - } \times \text{ - - - } \circ \text{ - - - } \times \text{ - - - } \quad \cdots$$

Figure 8.4: Decimation to $N/2$ spins, with the same notation as before

8.5.2 Decimation to $N/2$ spins for a one-dimensional Ising model with $H \neq 0$

Let us now see a different decimation procedure for the same one-dimensional Ising model, when $H \neq 0$.

This time, as shown in figure 8.4, the idea of the procedure is to sum over the spins that are on even sites and leaving unaltered those on odd sites.

We write the partition function as:

$$Z_N(K, h) = \text{Tr} \, e^{-\beta \mathcal{H}} = \sum_{\{S_i = \pm 1\}} e^{K \sum_i S_i S_{i+1} + h \sum_i S_i} =$$

$$= \sum_{S_1 = \pm 1} \cdots \sum_{S_N = \pm 1} e^{K(S_1 S_2 + S_2 S_3) + h S_2 + \frac{h}{2}(S_1 + S_3)} \cdot e^{K(S_3 S_4 + S_4 S_5) + h S_4 + \frac{h}{2}(S_3 + S_5)} \cdots$$

Indicating with S_i' the spins that are kept untouched and summing over the $N/2$ even spins:

$$Z_N(K, h) = \sum_{\{S_{2i+1}' = \pm 1\}} \sum_{\{S_{2i} = \pm 1\}} e^{S_2 \left[K(S_1' + S_3') + h \right] + \frac{h}{2}(S_1' + S_3')} e^{S_4 \left[K(S_3' + S_5') + h \right] + \frac{h}{2}(S_3' + S_5')} \cdots =$$

$$= \sum_{\{S_{2i+1}' = \pm 1\}} \left[e^{\left(K + \frac{h}{2}\right)(S_1' + S_3') + h} + e^{\left(-K + \frac{h}{2}\right)(S_1' + S_3') - h} \right] \cdot \left[e^{\left(K + \frac{h}{2}\right)(S_3' + S_5') + h} + e^{\left(-K + \frac{h}{2}\right)(S_3' + S_5') - h} \right] \cdots =$$

$$= \sum_{\{S_{2i+1}' = \pm 1\}} \prod_{i=0}^{N-1} \left[e^{\left(K + \frac{h}{2}\right)(S_{2i+1}' + S_{2i+3}') + h} + e^{\left(-K + \frac{h}{2}\right)(S_{2i+1}' + S_{2i+3}') - h} \right]$$

Now, since Z must not change after the RG transformation we can write:

$$Z_N(K, h) = e^{Ng(K,h)} Z_{N'}(K', h') = e^{Ng(K,h)} \text{Tr}_{\{S'\}} e^{-\beta \mathcal{H}'} \tag{8.12}$$

where:

$$-\beta \mathcal{H}'(K', h') = K' \sum_{i=0}^{N'-1} S_{2i+1}' S_{2i+3}' + h' \sum_{i=0}^{N'-1} S_{2i+1}'$$

Since $N' = N/2 \Rightarrow N = 2N'$, this means that:

$$Z_N(K, h) = e^{2N'g} \sum_{\{S_{2i+1}' = \pm 1\}} e^{K' \sum_i S_{2i+1}' S_{2i+3}' + h' \sum_i S_{2i+1}'} = \sum_{\{S_{2i+1}' = \pm 1\}} \prod_{i=0}^{N'-1} e^{K' S_{2i+1}' S_{2i+3}' + \frac{h'}{2}(S_{2i+1}' + S_{2i+3}') + 2g}$$

Therefore, we must have:

$$e^{\left(K + \frac{h}{2}\right)(S_{2i+1}' + S_{2i+3}') + h} + e^{\left(-K + \frac{h}{2}\right)(S_{2i+1}' + S_{2i+3}') - h} = e^{K' S_{2i+1}' S_{2i+3}' + \frac{h'}{2}(S_{2i+1}' + S_{2i+3}') + 2g}$$

and this equality must hold for all the possible values of S_{2i+1}' and S_{2i+3}'. In particular:

$$S_{2i+1}' = S_{2i+3}' = 1 \quad \Rightarrow \quad e^{2K + 2h} + e^{-2K} = e^{K' + h' + 2g}$$

$$S_{2i+1}' = S_{2i+3}' = -1 \quad \Rightarrow \quad e^h + e^{-h} = e^{K' - h' + 2g} \tag{8.13}$$

$$S_{2i+1}' S_{2i+3}' = -1 \quad \Rightarrow \quad e^h + e^{-h} = e^{-K' + 2g}$$
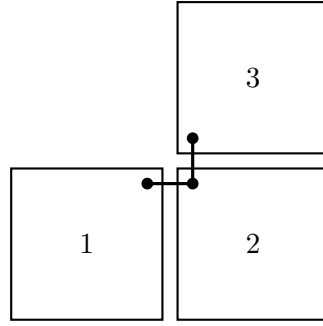
Figure 8.5: Introduction of a new effective interaction

The solutions of these equations are:

$$K' = \frac{1}{4} \ln \frac{\cosh(2K+h)\cosh(2K-h)}{\cosh^2 h} \qquad\qquad h' = h + \frac{1}{2}\ln\frac{\cosh(2K+h)}{\cosh(2K-h)}$$

$$g = \frac{1}{8}\ln\left[16\cosh(2K+h)\cosh(2K+h)\cosh^2 h\right]$$

which are the recursion relations for this decimation procedure.
Defining:

$$x = e^{-4K} \qquad\qquad y = e^{-2h} \qquad\qquad z = e^{-8g}$$

(where of course $0 \leq x, y, z \leq 1$) equations (8.5.2) can be more easily written as:

$$x' = x\frac{(1+y)^2}{(x+y)(1+xy)} \qquad\qquad y' = y\frac{x+y}{1+xy} \qquad\qquad z' = \frac{z^2 xy^2}{(x+y)(1+xy)(1+y)^2}$$

Note that x' and y' do not depend on z: this means that the constant g is not involved in the singular behaviour of the free energy density. In fact, from (8.12) we have:

$$f_N(K,h) = \frac{1}{2}f_{N'}(K',h') - k_B T g(K,h)$$

and since g does not influence the RG flow of the variables x and y (i.e. K and h), the critical properties of the system are not altered by g; since as we know these are determined by the behaviour of the singular part of f, g is part of the regular one.

8.5.3 Decimation in $d > 1$: proliferation of the interactions

As we have already stated, in $d = 1$ the recursion relations can be determined without great problems and they don't introduce new interactions. However this is not the case if $d > 1$, and the value of the new coupling constants can't be determined exactly, forcing us to use approximations.
Let us see with a generic example how the RG transformation can introduce new interactions in a two-dimensional Ising model with nearest-neighbour interactions.

Suppose we divide our system in blocks containing an odd number of spins and, similarly to what we have seen for $d = 1$ in 8.5.1, we sum over the spins on the boundary of the block and leave unchanged the one at the center. Looking at figure 8.5, we see that the spin on the corder of the block 2 is coupled to one spin in bock 1 and one in block 3. When we sum over the spin in block 2 an effective coupling between blocks 1 and 3 will be established: we therefore see that the coarse-graining procedure introduces next-nearest-neighbour interactions between the blocks, so new terms are appearing in the Hamiltonian (which of course, as already stated,

$$\begin{matrix} & & S'_{i+y} & & \\ \times & & \circ & & \times \\ S'_{i-x} & & \sigma_i & & S'_{i+x} \\ \circ & & \times & & \circ \\ & & S'_{i-y} & & \\ \times & & \circ & & \times \end{matrix}$$
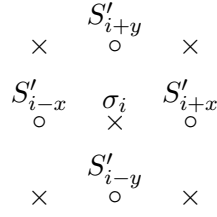
Figure 8.6: Decimation for the two-dimensional Ising model

respect the symmetries of the original one).
We therefore understand that the iteration of the RG will introduce increasingly complicated couplings: this is the so called *proliferation* of interactions.

Let us now see in detail how to face the problem of the proliferation for a two-dimensional Ising model with nearest-neighbour interaction and $H = 0$.
We choose to coarse-grain the system summing over a "chessboard" of spins, as shown in figure 8.6, which also defines the symbolic notation we are going to use.
We therefore have:

$$Z_N = \text{Tr}_{\{S_i\}} \, e^{K \sum_{\langle ij \rangle} S_i S_j} = \sum_{\{S'_i = \pm 1\}} \sum_{\{\sigma_i = \pm 1\}} e^{K \sum_i \left(S'_{i-x} + S'_{i+x} + S'_{i-y} + S'_{i+y} \right) \sigma_i}$$

and performing the sum over σ_i:

$$Z_N = \sum_{\{S'_i = \pm 1\}} \prod_i \left[e^{K(S'_{i-x} + S'_{i+x} + S'_{i-y} + S'_{i+y})} + e^{-K(S'_{i-x} + S'_{i+x} + S'_{i-y} + S'_{i+y})} \right] =$$

$$= \sum_{\{S'_i = \pm 1\}} \prod_i 2 \cosh \left[K(S'_{i-x} + S'_{i+x} + S'_{i-y} + S'_{i+y}) \right]$$

The argument of $\cosh$ can assume three possible values, which come from the fact that the $2^4 = 16$ possible spin configurations can be grouped into three categories: all spins aligned, three spins aligned and one not, and two spins aligned and two not. As before, we want to write $Z_N[K] = e^{N g(K,K')} Z_{N'}[K']$, and from what we have just seen we will have three conditions to satisfy so $\mathcal{H}'$ will contain at least three terms: in the coarse-grained Hamiltonian new terms are appearing.
We can therefore try for example with the following guess for rewriting Z_N:

$$2 \cosh[K(S'_{i-x} + S'_{i+x} + S'_{i-y} + S'_{i+y})] =$$

$$= z(K) e^{K'(S'_{i-x}S'_{i-y} + S'_{i-x}S'_{i+y} + S'_{i+x}S'_{i-y} + S'_{i+x}S'_{i+y}) + L'(S'_{i-x}S'_{i+x} + S'_{i-y}S'_{i+y}) + Q'S'_{i-x}S'_{i+x}S'_{i-y}S'_{i+y}} \quad (8.14)$$

This way, besides nearest-neighbour interactions (K) we are introducing also next-nearest-neighbour ones (K' and L') and also four-spin cluster interactions (Q'), as shown in figure 8.7. Note also that the final set of spins resides on a square whose side is $\sqrt{2}$ times the original one, so we have $\ell = \sqrt{2}$.
Inserting all the possible spin configurations in (8.14) we get the following equations:

$$2\cosh(4K) = z(K)e^{2K'+2L'+Q'} \qquad\qquad 2\cosh(4K) = z(K)e^{-Q'}$$
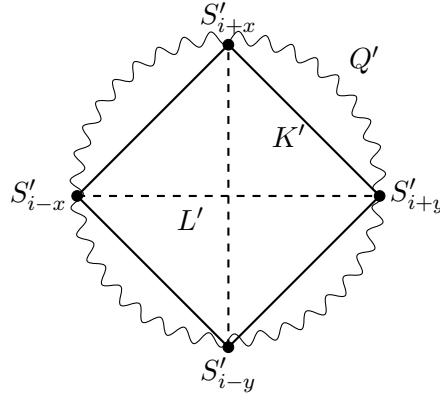
$$2 = z(K)e^{-2L'+Q'} \qquad\qquad 2 = z(K)e^{-2K'-2L'+Q'}$$

Figure 8.7: New interactions introduced
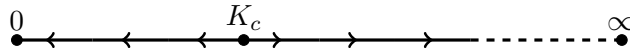

Figure 8.8: RG flow of (8.15)

Their solutions are given by:

$$z(K) = 2\cosh^{1/8}(4K)\cosh^{1/2}(2K) \qquad\qquad K' = \frac{1}{4}\ln\cosh(4K)$$

$$L' = \frac{1}{8}\ln\cosh(4K) \qquad\qquad Q' = \frac{1}{8}\ln\cosh(4K) - \frac{1}{2}\ln\cosh(2K)$$

If we now reiterate the procedure, more complicated interactions will appear and the problem becomes rapidly intractable. We therefore must do some approximations.

We choose to neglect Q' (also because it is the only coupling that can become negative and thus prevent the spins from aligning) and to omit the explicit dependence on L' defining a new constant K':

$$K'_{\text{new}} = K' + L'$$

This way the recursion relation involves only K:

$$K' = \frac{3}{8}\ln\cosh(4K)$$

Let us therefore see to which conclusions does this lead.

The fixed points are given by:

$$K^* = \frac{3}{8}\ln\cosh(4K^*) \tag{8.15}$$

and the non-trivial ($K^* \neq 0$) numerical solution of this equation is $K_c = 0.50698\ldots$; the exact value found by Onsager is $K_{\text{exact}} = 0.44069\ldots$, so our approximation is good enough.

If now the initial value K_0 of K is greater than K_c, then the sequence $K^{(n)} = \frac{3}{8}\ln\cosh(4K^{(n)})$ grows indefinitely, while if $K_0 < K_c$ it tends to zero. Thus, the fixed points $K^* = 0$ and $K^* = \infty$ are stable, while $K^* = K_c$ is unstable, as shown in figure 8.8.

Let us now linearise the recursion relation near K_c and compute a couple of critical exponents. On the base of what we have seen in 8.4, if we call $\delta K = K - K_c$ and $\delta K' = K' - K_c$ we have:

$$\delta K' = \lambda_t \delta K \qquad\qquad \text{where} \qquad \lambda_t = \frac{dK'}{dK}_{\big|K_c}$$

Therefore, since also $\ell = \sqrt{2}$, we get:

$$y_t = \frac{\ln\lambda_t}{\ln\ell} = \frac{1}{\ln\sqrt{2}}\ln\left(\frac{dK'}{dK}_{\big|K_c}\right) = \frac{\ln\left[\frac{3}{2}\tanh(4K_c)\right]}{\ln 2/2} = 1.070\ldots$$

and so:

$$\nu = \frac{1}{y_t} = 0.9345\dots$$

and from the hyperscaling relation:

$$\alpha = 2 - \frac{d}{y_t} = 0.131\dots$$

Onsager's exact result, as we know, gives $\alpha = 0$ (since the specific heat diverges logarithmically) and thus $\nu = 1$. We therefore see that our approximation is sufficiently accurate (even if improvable).

Appendix A

Volume of a hypersphere

In this appendix we will compute the volume of an ℓ-dimensional hypersphere of radius R, which we call $V_\ell(R)$. This is defined as:

$$V_\ell(R) = \int \Theta(R - |\vec{p}|)d\vec{p}$$

In order to calculate it we will use a "trick".
First, we change variable defining $\vec{z}$ so that $p_i = Rz_i$ and thus[1]:

$$V_\ell(R) = R^\ell \int \Theta(1 - |\vec{z}|)d\vec{z} = R^\ell V_\ell(1) \tag{A.1}$$

We then have:

$$\frac{d}{dR}V_\ell(R) = \ell R^{\ell-1}V_\ell(1) = \int \delta(|\vec{p}| - R)d\vec{p}$$

Multiplying both sides by e^{-R^2} and integrating over R from 0 to infinity:

$$V_\ell(1)\int_0^\infty e^{-R^2}\ell R^{\ell-1}dR = \int_0^\infty dR \int e^{-R^2}\delta(|\vec{p}| - R)d\vec{p}$$

The right hand side is equal to:

$$\int_0^\infty dR \int e^{-R^2}\delta(|\vec{p}| - R)d\vec{p} = \int d\vec{p} \int_0^\infty e^{-R^2}\delta(|\vec{p}| - R) = \int d\vec{p}\, e^{-|\vec{p}|^2}$$

Therefore:

$$\ell V_\ell(1)\int_0^\infty R^{\ell-1}e^{-R^2}dR = \int d\vec{p}\prod_{i=1}^\ell e^{-p_i^2} = \prod_{i=1}^\ell \underbrace{\int dp_i e^{-p_i^2}}_{\sqrt{\pi}} = \pi^{\ell/2}$$

If we now define $t = R^2$, in the integral in the left hand side of the equation we recognise the definition of Euler's Γ function[2]:

$$\ell\int_0^\infty R^{\ell-1}e^{-R^2}dR = \frac{\ell}{2}\int_0^\infty e^{-t}t^{\frac{\ell}{2}-1}dt = \frac{\ell}{2}\Gamma\left(\frac{\ell}{2}\right) = \Gamma\left(\frac{\ell}{2}+1\right)$$

---

[1] We also use the fact that $\Theta(R(1-z)) = \Theta(1-z)$, since $R > 0$.

[2] As a remainder, the Γ function is defined as:

$$\Gamma(n) = \int_0^\infty e^{-t}t^{n-1}dt$$

If $n \in \mathbb{N}$, then $\Gamma(n) = (n-1)!$.

This way the volume of an ℓ-dimensional sphere of unit radius is $V_\ell(1) = \pi^{\ell/2}/\Gamma\left(\frac{\ell}{2} + 1\right)$, and thanks to equation (A.1), in the end we have:

$$V_\ell(R) = \frac{R^\ell \pi^{\ell/2}}{\Gamma\left(\frac{\ell}{2} + 1\right)} \tag{A.2}$$

Let us verify that this formula returns the values we expect for $\ell = 2$ and $\ell = 3$. We first note that:

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-t} t^{-1/2} dt = \sqrt{\pi}$$

$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2} \cdot \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{3}{4}\sqrt{\pi}$$

This way:

$$V_2(R) = \frac{R^2 \pi^{2/2}}{\Gamma(2)} = \frac{\pi R^2}{1!} = \pi R^2 \qquad V_3(R) = \frac{R^3 \pi^{3/2}}{\frac{3}{4}\sqrt{\pi}} = \frac{4}{3}\pi R^3$$

which is exactly what we expected.

Appendix B

The saddle point approximation

In this appendix we are going to see how the saddle point approximation works in general. Let us define the class of integrals:

$$I_N = \int e^{Nf(x)} dx \qquad N \gg 1$$

Suppose that $f(x)$ has a unique maximum at $x = x^*$ and that $\lim_{|x| \to \infty} f(x) = -\infty$. Then, expanding f around x^* we have:

$$f(x) = f(x^*) + \frac{(x - x^*)^2}{2} f''(x^*) + \frac{(x - x^*)^3}{6} f'''(x^*) + \cdots$$

(and obviously $f'(x^*) = 0$ because x^* is a maximum). Setting $z = x - x^*$ we can write, stopping the expansion at the third order:

$$I_N \approx e^{Nf(x^*)} \int e^{N\frac{z^2}{2} f''(x^*) + N\frac{z^3}{6} f'''(x^*)} dz$$

Calling $y = z\sqrt{N}$ and remembering that $f''(x^*) < 0$ since x^* is a maximum, we have:

$$I_N = e^{Nf(x^*)} \frac{1}{\sqrt{N}} \int e^{-\frac{y^2}{2}|f''(x^*)| + \frac{y^2}{6\sqrt{N}} f'''(x^*)} dy$$

Therefore for very large N the term proportional to $f'''(x^*)$ in the exponential (like all the following terms of the complete expansion) is negligible, so:

$$I_N \approx e^{Nf(x^*)} \frac{1}{\sqrt{N}} \int e^{-\frac{y^2}{2}|f''(x^*)|} dy$$

and computing the Gaussian integral:

$$I_N \approx \sqrt{\frac{2\pi}{N|f''(x^*)|}} \cdot e^{Nf(x^*)}$$

Therefore we see that the saddle point approximation essentially states that an integral of the form I_N can be approximated, provided that N is large, with the value of the integrand calculated at its maximum (up to a multiplicative factor).

Appendix C

Stirling's approximation

We want to derive here Stirling's approximation, i.e. we want to show that for large n we have:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

We start by writing $n!$ with Euler's Γ function:

$$n! = \int_0^\infty t^n e^{-t} dt$$

Changing variable to $s = t/n$ we get:

$$n! = \int_0^\infty s^n n^n e^{-sn} n \, ds = \int_0^\infty s^n e^{-sn} n^{n+1} ds$$

which we rewrite as:

$$n! = \int_0^\infty e^{-sn + n\ln s + (n+1)\ln n} ds = e^{(n+1)\ln n} \int_0^\infty e^{n(\ln s - s)} ds$$

The last one is an integral in the form of those studied in appendix B; using the notation shown there we have $f(s) = \ln s - s$ and its maximum is at $s = 1$. Since $f''(1) = -1$, we get:

$$n! \sim e^{(n+1)\ln n} \sqrt{\frac{2\pi}{n \cdot |-1|}} e^{-n} = n^n n \sqrt{\frac{2\pi}{n}} e^{-n} = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Therefore:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Taking the logarithm we find another famous expression for Stirling's approximation:

$$\ln n! \sim n \ln n - n + \frac{1}{2}\ln(2\pi n)$$

Appendix D

A more convincing foundation of statistical mechanics

D.1 Introduction

As we have stated in 3.2.5, we would like to understand if a more rigorous foundation of statistical mechanics can be established. In particular we would like to see if the equal a priori probability in the phase space of a system is a general property of *any* physical system, and if so *why* does it occur.

To make things more tangible let us consider a system with Hamiltonian:

$$\mathcal{H}(\mathbb{Q}, \mathbb{P}) = \sum_i \frac{\vec{p_i}^2}{2m_i} + V(\mathbb{Q})$$

We know that from a microscopic point of view the system can be completely described if we solve Hamilton's equations:

$$\dot{\vec{q_i}} = \frac{\partial \mathcal{H}}{\partial \vec{p_i}} = \frac{\vec{p_i}}{m_i} \qquad \dot{\vec{p_i}} = -\frac{\partial \mathcal{H}}{\partial \vec{q_i}} = -\frac{\partial V}{\partial \vec{q_i}}$$

with given initial conditions that we call $(\mathbb{Q}_0, \mathbb{P}_0)$, so that in the end we know $(\mathbb{Q}(t), \mathbb{P}(t))$ for any t.

If we now call $O(\mathbb{Q}, \mathbb{P})$ a generic observable of the system we can define its *time average* as:

$$\overline{O}(\mathbb{Q}_0, \mathbb{P}_0) = \lim_{T \to \infty} \frac{1}{T} \int_0^T O(\mathbb{Q}(t), \mathbb{P}(t)) dt$$

We can therefore ask ourselves: is there a relation between the time average $\overline{O}$ and the ensemble average $\langle O \rangle$ of the observable O? And in particular are there any hypotheses under which these two averages are equal?

It is very important to find an answer to this question because $\overline{O}$, as can be seen from its definition, derives solely from the microscopic properties of the system while $\langle O \rangle$ comes from its statistical description given within the microcanonical ensemble. Therefore if we can find a relation between these two quantities we can establish a microscopic foundation of the microcanonical ensemble and thus of the whole statistical mechanics.

As we will see, we will be able to determine under which hypotheses $\overline{O} = \langle O \rangle$ but unfortunately to date it is still not known *when* these hypotheses do apply.

D.2 Liouville's theorem

The first important result that we want to prove is *Liouville's theorem*; this is a very important theorem which will be useful in the following, and ultimately justifies why we microscopically describe systems in phase space and not in other possible spaces[1]. It essentially states that the phase space volume is *locally conserved*, or in other words time evolution doesn't change it. Let's show it explicitly.

Theorem (Liouville). *Let $\mathcal{H}$ be the Hamiltonian of a physical system of N particles, and $\rho(\mathbb{Q}, \mathbb{P}, t)$ a generic probability density in phase space which can in general also depend explicitly on time. Let $(\mathbb{Q}_0, \mathbb{P}_0)$ be the initial conditions of the system and call $(\mathbb{Q}(t), \mathbb{P}(t))$ their time evolutions, namely the solutions of Hamilton's equations:*

$$\dot{\mathbb{Q}}(t) = \frac{\partial}{\partial \mathbb{P}} \mathcal{H}(\mathbb{Q}(t), \mathbb{P}(t)) \qquad \dot{\mathbb{P}}(t) = -\frac{\partial}{\partial \mathbb{Q}} \mathcal{H}(\mathbb{Q}(t), \mathbb{P}(t))$$

Then ρ is constant along $(\mathbb{Q}(t), \mathbb{P}(t))$, namely:

$$\frac{d}{dt} \rho(\mathbb{Q}(t), \mathbb{P}(t), t) = 0$$

Proof. Since ρ is a probability density it must satisfy the continuity equation[2]:

$$\frac{\partial}{\partial t} \rho(\mathbb{Q}(t), \mathbb{P}(t), t) = -\vec{\nabla} \cdot \vec{J}(\mathbb{Q}(t), \mathbb{P}(t), t)$$

where the *probability flow* is given by[3]:

$$\vec{J}(\mathbb{Q}(t), \mathbb{P}(t), t) = \rho(\mathbb{Q}(t), \mathbb{P}(t), t) \begin{pmatrix} \dot{\mathbb{Q}}(t) \\ \dot{\mathbb{P}}(t) \end{pmatrix}$$

The continuity equation can be rewritten as:

$$0 = \frac{\partial}{\partial t} \rho(\mathbb{Q}(t), \mathbb{P}(t), t) + \vec{\nabla} \cdot \vec{J}(\mathbb{Q}(t), \mathbb{P}(t), t) = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i} (\dot{q}_i \rho) + \frac{\partial}{\partial p_i} (\dot{p}_i \rho) \right] =$$

$$= \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i} \left(\frac{\partial \mathcal{H}}{\partial p_i} \rho \right) + \frac{\partial}{\partial p_i} \left(-\frac{\partial \mathcal{H}}{\partial q_i} \rho \right) \right]$$

If we now compute all the derivatives, since $\mathcal{H}$ is a sufficiently regular function we can use Schwartz's theorem to cancel all the terms with the second derivatives in $\mathcal{H}$. We therefore get:

$$0 = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left(\dot{q}_i \frac{\partial \rho}{\partial q_i} + \dot{p}_i \frac{\partial \rho}{\partial p_i} \right)$$

---

[1]For example, if we consider a system of particles of the same mass we could have equally well described it in the space of positions and velocities rather than with positions and momenta. In this case, in fact, a collision between particles can be seen as an exchange of velocity between the two and since all the particles have the same mass this is also equivalent to an exchange in momentum. However, if we consider particles of different mass this is not true any more and the only quantity exchanged between colliding particles is momentum. Therefore, position-velocity space does not have the same properties of phase space (in particular, Liouville's theorem does not apply).

[2]This is a general property of probability distributions: in fact probability can't just "disappear" and "reappear" in different parts of the phase space.

[3]This can be better understood through an analogy with fluids: if $\rho(\vec{r}, t)$ is the fluid density, then it must satisfy the continuity equation $\dot{\rho} = -\vec{\nabla} \cdot \vec{J}$ where $\vec{J} = \rho \dot{\vec{r}}$.

However, by definition of total derivative we have precisely:

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left(\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right)$$

So in the end:

$$\frac{d}{dt}\rho(\mathbb{Q}(t), \mathbb{P}(t), t) = 0$$

$\square$

As a consequence of Liouville's theorem we have that the probability density in phase space of an isolated system in equilibrium is such that:

$$\frac{d}{dt}\rho_{\text{eq}}(\mathbb{Q}(t), \mathbb{P}(t)) = 0$$

(in fact since the system is in equilibrium its probability density doesn't depend explicitly on time). Therefore ρ_{eq} is a conserved quantity; we have however seen that the only conserved quantity for an isolated system at equilibrium is its energy[4] so ρ_{eq} must be some function F of the Hamiltonian $\mathcal{H}$:

$$\rho_{\text{eq}}(\mathbb{Q}, \mathbb{P}) = F(\mathcal{H}(\mathbb{Q}, \mathbb{P}))$$

In the microcanonical ensemble we have postulated that $F(\mathcal{H}) = \delta(\mathcal{H} - E)/\Omega(E, V, N)$, but this is just a particular case (there is nothing that urges us to suppose so).
Therefore, the a priori equal probability postulate is *compatible* with Liouville's theorem but does not necessarily derive from it.

D.3 Ergodicity

Let us note that Liouville's theorem does not prevent the microstate of a system to be confined in a particular region of the constant-energy phase space hypersurface, or to move "mostly" in that region[5]. In this case we surely have:

$$\overline{O}(\mathbb{Q}, \mathbb{P}) = \lim_{T \to \infty} \frac{1}{T} \int_0^T O(\mathbb{Q}(t), \mathbb{P}(t)) dt \neq \langle O(\mathbb{Q}, \mathbb{P}) \rangle$$

Such phenomena *do* occur in physical systems, in particular those with few degrees of freedom. Planetary systems are an example since the orbits of the planets are approximately stable and thus the representative point of the system always remains in the same region of the phase space; if it "spanned" all the accessible phase space hypersurface of constant energy then the orbits would mess up completely and the system wouldn't be stable.
There are also however many-particle systems whose representative point remains almost always in the same region of phase space. An example is the *Fermi-Pasta-Ulam* system, a chain of N anharmonic oscillators (namely every particle of the system is subject to the potential[6] $U(x) = kx^2 + ux^4$). In this case it turns out that if we give some energy to a single particle, this energy is *not* distributed along all the system as one would expect if the representative point of the system spanned the entire accessible region of phase space. However this kind of system is a bit "pathological" since it has been shown that in the continuum limit the Fermi-Pasta-Ulam

---

[4]We are considering the system fixed and still, so that its momentum and angular momentum are zero.

[5]This means that the time evolution of the representative point is such that it is much more probable to find it in determinate regions of phase space than others.

[6]A simple harmonic potential would be too simple: with a proper change of coordinates, in fact, the system can be described as a set of N independent particles.

system has an *infinite* number of conserved quantities, not just one (as we require).

We are interested, however, in systems where such phenomena do not occur. In particular in order to establish a link with the microcanonical ensemble we would like to discuss systems where the representative point "spans" the whole accessible phase space, namely it spends on average the same time covering different regions of the constant energy hypersurface.

This property of dynamical systems is known with the name of *ergodicity*. Its most intuitive definition (although not the most useful, as we will shortly see) is the following:

Definition (Ergodicity, I). *A dynamical system is said to be* ergodic *if on the hypersurface of constant energy the time evolution of almost every point eventually passes arbitrarily close to any other point.*

The expression "almost every point" of the hypersurface means that we are considering it up to a set of null measure. This is needed to avoid problems with strange or unusual configurations: in this way for example we are excluding the possibility that all the particles move precisely at the same velocity in neat rows.

As we have said this definition is not very useful, nor it is really clear since it does not follow explicitly from it that the representative point covers different regions of phase space on average in the same time.

In order to give a much more useful definition of ergodicity we first must introduce the concept of *ergodic component* of a set (which we think of as the hypersurface of constant energy):

Definition D.1 (Ergodic component). *Let S be a subset of the phase space. Then a set $\Sigma \subset S$ is called* ergodic component *of S if it is invariant under time evolution, namely:*

$$(\mathbb{Q}(0), \mathbb{P}(0)) \in \Sigma \qquad \Longrightarrow \qquad (\mathbb{Q}(t), \mathbb{P}(t)) \in \Sigma \quad \forall t$$

Intuitively, different ergodic components of a same set are subsets that are not "mixed" together by time evolution.

Now, we can give another definition of ergodicity:

Definition D.2 (Ergodicity, II). *A dynamical system is said to be* ergodic *if the measure of every ergodic component Σ of the hypersurface of constant energy S is either zero or equal to the measure of S. In other words, if μ is the measure defined on phase space then the system is* ergodic *if for any ergodic component Σ of S we have $\mu(\Sigma) = 0$ or $\mu(\Sigma) = \mu(S)$.*

This means that a dynamical system is ergodic if the hypersurface of constant energy S is invariant under time evolution up to a set of null measure.

Even if it is not immediately clear these two definitions are equivalent, but we won't show that. What we want to do now is to show that if a system is ergodic according to this last definition then the time average and the microcanonical ensemble average of an observable coincide.

Theorem D.1. *If a system is ergodic according to definition D.2 and $O(\mathbb{Q}, \mathbb{P})$ is an observable, then time average and ensemble average coincide:*

$$\overline{O}(\mathbb{Q}, \mathbb{P}) = \langle O(\mathbb{Q}, \mathbb{P}) \rangle$$

Proof. First of all, let us note that:

$$\overline{O}(\mathbb{Q}(0), \mathbb{P}(0)) = \overline{O}(\mathbb{Q}(t_0), \mathbb{P}(t_0))$$

namely the time average does not depend on the initial instant we choose. In fact we have:

$$\overline{O}(\mathbb{Q}(0), \mathbb{P}(0)) = \lim_{T \to \infty} \frac{1}{T} \int_0^T O(\mathbb{Q}(t), \mathbb{P}(t)) dt =$$

$$= \lim_{T \to \infty} \left[\frac{1}{T} \int_0^{t_0} O(\mathbb{Q}(t), \mathbb{P}(t)) dt + \frac{1}{T} \int_{t_0}^T O(\mathbb{Q}(t), \mathbb{P}(t)) dt \right]$$

The first integral is a constant (it does not depend on T), so in the limit $T \to \infty$ the first term vanishes. Therefore, multiplying and dividing by $T - t_0$:

$$\overline{O}(\mathbb{Q}(0), \mathbb{P}(0)) = \lim_{T \to \infty} \frac{T - t_0}{T} \frac{1}{T - t_0} \int_{t_0}^{T} O(\mathbb{Q}(t), \mathbb{P}(t)) dt = \overline{O}(\mathbb{Q}(t_0), \mathbb{P}(t_0))$$

(since $\lim_{T \to \infty} (T - t_0)/T = 1$).

We now define:

$$R_a = \left\{ (\mathbb{Q}, \mathbb{P}) : \quad \overline{O}(\mathbb{Q}, \mathbb{P}) \leq a \right\} \qquad a \in \mathbb{R}$$

For what we have noted above this is an ergodic component of the hypersurface S since the time evolution maps R_a into itself.
Therefore, either $\mu(R_a) = 0$ or $\mu(R_a) = \mu(S)$; note also that if $a < a'$ then $R_a \subset R_{a'}$ and $\mu(R_a) = \mu(S) \Rightarrow \mu(R_{a'}) = \mu(S)$.
We now call a^* the smallest value of a such that $\mu(R_a) = \mu(S)$:

$$a^* = \inf \left\{ a \in \mathbb{R} : \quad \mu(R_a) = \mu(S) \right\}$$

and we want to show that $\mu(R_{a^*}) = \mu(S)$ and that $a^* = \overline{O}(\mathbb{Q}, \mathbb{P})$ for almost all $(\mathbb{Q}, \mathbb{P})$.

Let us therefore consider a sequence a_n (with of course $n \in \mathbb{N}$) monotonically increasing and such that $\lim_{n \to \infty} a_n = \infty$, and call $R_{a_n} = R_n$. Then:

$$R_n \subset R_{n+1} \qquad\qquad \bigcup_n R_n = S$$

From theorems of measure theory this means that $\mu(R_n) \to \mu(S)^-$. Since R_n are ergodic components of S, this means that:

$$\exists n_0 \in \mathbb{N} : \quad \mu(R_n) = \mu(S) \quad \forall n \geq n_0$$

and so that there are some finite $a \in \mathbb{R}$ such that $\mu(R_a) = \mu(S)$; we thus deduce that $\exists a^* < \infty$.
Let us now consider another sequence of ergodic components:

$$R'_n = R_{a^* + 1/n}$$

Then:

$$R'_n \supset R'_{n+1} \qquad\qquad \mu(R'_n) = \mu(S)$$

and the last equality is true because for every n there exists $a^* < a'_n < a^* + 1/n$ such that $\mu(R_{a'_n}) = \mu(S)$. Again, from theorems of measure theory this means that:

$$\mu\left(\bigcap_n R'_n \right) = \lim_{n \to \infty} \mu(R'_n)$$

but since $\mu(R'_n) = \mu(S)$:

$$\mu\left(\bigcap_n R'_n \right) = \lim_{n \to \infty} \mu(S) = \mu(S)$$

Therefore, in order to show that $\mu(R_{a^*}) = \mu(S)$ we must show that $\bigcap_n R'_n = \bigcap_{a > a^*} R_a = R_{a^*}$.
We surely have $R_{a^*} \subset R_a$ for $a > a^*$, therefore:

$$R_{a^*} \subset \bigcap_{a > a^*} R_a$$

If we now consider $(\mathbb{Q}, \mathbb{P}) \in \bigcap_{a > a^*} R_a$, then:

$$\overline{O}(\mathbb{Q}, \mathbb{P}) \leq a \quad \forall a > a^* \qquad \Rightarrow \qquad \overline{O}(\mathbb{Q}, \mathbb{P}) \leq a^* \qquad \Rightarrow \qquad (\mathbb{Q}, \mathbb{P}) \in R_{a^*}$$

This means that:

$$R_{a^*} \supset \bigcap_{a > a^*} R_a$$

and so:

$$R_{a^*} = \bigcap_{a > a^*} R_a$$

We thus have found that:

$$\mu(R_a) = \begin{cases} 0 & a < a^* \\ \mu(S) & a \geq a^* \end{cases}$$

Therefore, if $a > a^*$ then $R_{a^*} \subset R_a$ and $\mu(R_a \setminus R_{a^*}) = \mu(R_a) - \mu(R_{a^*}) = 0$.
This means that $\overline{O}(\mathbb{Q}, \mathbb{P}) = a^*$ everywhere but on the points $(\mathbb{Q}, \mathbb{P})$ of the sets:

$$\left\{ (\mathbb{Q}, \mathbb{P}) : \quad \overline{O}(\mathbb{Q}, \mathbb{P}) < a^* \right\} := R' \qquad\qquad \left\{ (\mathbb{Q}, \mathbb{P}) : \quad \overline{O}(\mathbb{Q}, \mathbb{P}) > a^* \right\} = (S \setminus R_{a^*}) \setminus R'$$

However, $\mu(R') = 0$ (this can be shown similarly as what we have done, using the sequence of ergodic components $R_n'' = R_{a^* - 1/n}$), and also the second set has null measure since $\mu(S \setminus R_{a^*}) = 0$.

Therefore, we now have to show that from the fact that $\overline{O}(\mathbb{Q}, \mathbb{P}) = a^*$ almost everywhere it follows that $\langle O \rangle = a^*$.
We have[7]:

$$\langle O \rangle = \frac{1}{\Delta \cdot \Omega(E)} \int_{E \leq \mathcal{H} \leq E + \Delta} O(\mathbb{Q}, \mathbb{P}) d\Gamma$$

If we call $(\mathbb{Q}(t), \mathbb{P}(t))$ the time evolution of $(\mathbb{Q}, \mathbb{P})$ so that $(\mathbb{Q}, \mathbb{P}) = (\mathbb{Q}(t = 0), \mathbb{P}(t = 0))$ and $\Gamma_0 = \{(\mathbb{Q}, \mathbb{P}) : \mathcal{H}(\mathbb{Q}, \mathbb{P}) \in [E, E + \Delta]\}$, then by definition we have:

$$\langle O(t) \rangle = \frac{1}{\Delta \cdot \Omega(E)} \int_{\Gamma_0} O(\mathbb{Q}(t), \mathbb{P}(t)) d\Gamma$$

Note that the integral is made over Γ_0 and so we are integrating over the initial conditions (in fact, in general $\mathbb{Q}(t)$ and $\mathbb{P}(t)$ will be functions of time and of the initial conditions, namely $\mathbb{Q}(t) = \hat{\mathbb{Q}}(t; \mathbb{Q}, \mathbb{P})$ and $\mathbb{P}(t) = \hat{\mathbb{P}}(t; \mathbb{Q}, \mathbb{P})$). Therefore, changing variables to $(\mathbb{Q}(t), \mathbb{P}(t))$:

$$\langle O(t) \rangle = \frac{1}{\Delta \cdot \Omega(E)} \int_{\Gamma_t} (\text{Jac}(t))^{-1} O(\mathbb{Q}(t), \mathbb{P}(t)) d\Gamma_t$$

where we have written the Jacobian of the change of coordinates in this way for convenience. However, since $\mathcal{H}(\mathbb{Q}(t), \mathbb{P}(t)) = \mathcal{H}(\mathbb{Q}, \mathbb{P})$ we have that $\Gamma_t = \Gamma_0$ and so:

$$\langle O(t) \rangle = \frac{1}{\Delta \cdot \Omega(E)} \int_{\Gamma_0} (\text{Jac}(t))^{-1} O(\mathbb{Q}, \mathbb{P}) d\Gamma$$

where we have renamed the integration variables to $(\mathbb{Q}, \mathbb{P})$.
Now:

$$(\text{Jac}(t))^{-1} = \det \frac{\partial(\mathbb{Q}, \mathbb{P})}{\partial(\mathbb{Q}(t), \mathbb{P}(t))} \qquad \Rightarrow \qquad \text{Jac}(t) = \det \frac{\partial(\mathbb{Q}(t), \mathbb{P}(t))}{\partial(\mathbb{Q}, \mathbb{P})}$$

---

[7] We consider a thin energy shell instead of an hypersurface because it makes things simpler.

namely:

$$\text{Jac}(t) = \begin{vmatrix} \frac{\partial q_1(t)}{\partial q_1} & \cdots & \frac{\partial p_{3N}(t)}{\partial q_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial q_1(t)}{\partial p_{3N}} & \cdots & \frac{\partial p_{3N}(t)}{\partial p_{3N}} \end{vmatrix}$$

For dynamical systems, $\text{Jac}(t) = 1$ for all t as a consequence of Hamilton's equations.
In fact calling for brevity $(\mathbb{Q}(t), \mathbb{P}(t)) := x$ and $(\mathbb{Q}, \mathbb{P}) := y$, then:

$$\frac{d}{dt}\text{Jac}(t) = \sum_{i=1}^{6N} J_i \qquad\qquad J_i = \det \frac{\partial(x_1, \ldots, \dot{x}_i, \ldots, x_{6N})}{\partial(y_1, \ldots, y_{6N})}$$

and we have:

$$\dot{x}_i = \begin{cases} \frac{\partial \mathcal{H}}{\partial p_i} & \text{if } x_i = q_i \\ -\frac{\partial \mathcal{H}}{\partial q_i} & \text{if } x_i = p_i \end{cases}$$

Furthermore, in general:

$$\frac{\partial \dot{x}_i}{\partial y_k} = \sum_{j=1}^{6N} \frac{\partial \dot{x}_i}{\partial x_j} \frac{\partial x_j}{\partial y_k}$$

and thus:

$$\frac{d}{dt}\text{Jac}(t) = \sum_{i=1}^{6N} \sum_{j=1}^{6N} \frac{\partial \dot{x}_i}{\partial x_j} \det \frac{\partial(x_1, \ldots, x_{i-1}, x_j, x_{i+1}, \ldots, x_{6N})}{\partial(y_1, \ldots, y_{6N})}$$

This determinant is null if $i \neq j$ (because the determinant of a matrix with two equal columns vanishes), otherwise it will be equal to a constant J. Therefore:

$$\det \frac{\partial(x_1, \ldots, x_{i-1}, x_j, x_{i+1}, \ldots, x_{6N})}{\partial(y_1, \ldots, y_{6N})} = J\delta_{ij}$$

and so:

$$\frac{d}{dt}\text{Jac}(t) = J\sum_{i=1}^{6N} \frac{\partial \dot{x}_i}{\partial x_i} = J\sum_{i=1}^{3N} \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) J\sum_{i=1}^{3N} \left(\frac{\partial^2 \mathcal{H}}{\partial q_i \partial p_i} - \frac{\partial^2 \mathcal{H}}{\partial p_i \partial q_i} \right) = 0$$

Therefore $\text{Jac}(t) = \text{const.} = \text{Jac}(0) = 1$, since for $t = 0$ the change of coordinates is the identity.

This means that:

$$\langle O(t) \rangle = \frac{1}{\Delta \cdot \Omega(E)} \int_{\Gamma_0} O(\mathbb{Q}, \mathbb{P}) d\Gamma$$

namely $\langle O(t) \rangle$ doesn't depend on time. Therefore:

$$\langle O \rangle = \langle O(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T \langle O(t) \rangle \, dt = \left\langle \lim_{T \to \infty} \frac{1}{T} \int_0^T O(\mathbb{Q}(t), \mathbb{P}(t)) dt \right\rangle = \langle \overline{O}(\mathbb{Q}, \mathbb{P}) \rangle = \langle a^* \rangle = a^*$$

This ultimately implies that $\langle O \rangle = \overline{O}$, which is what we wanted to show. $\square$

So, we now know that if a system is ergodic then the microcanonical ensemble is well defined. But how can we know if a system is ergodic or not? Unfortunately we still don't know: this is to date an open problem.
We can however cite two other important systems which can be not ergodic: magnets and glasses. A magnet (as also shown throughout chapter 5) can be considered as composed of small orientable magnetic dipoles (the spins of the atoms); at high temperatures the system is "disordered" and the dipoles are not aligned, but when the temperature becomes smaller

than the so called "critical" one T_c these dipoles align along any of the possible directions in space. The system thus spontaneously breaks its internal symmetry; such phenomena lead to ergodicity breaking: in fact when $T < T_c$ it can be shown that the time it takes the system to spontaneously rearrange its magnetization along another direction grows with the dimension of the system. This means that in the thermodynamic limit the system will *always* remain in the same configurations, and so its representative point will not visit all the available regions of phase space (note that the configuration of the system is now given by the spin configuration, not the positions of the particles).

The same argument applies to other kinds of phase transitions that break a symmetry of a given system, for example the solidification of a fluid.

Glasses are much more complicated systems, and many of their properties are still unknown. Their main characteristic is that they are nor crystalline solids nor fluids, so strictly speaking they are not at equilibrium: they tend to approach a crystalline configuration, but the process takes insanely huge amounts of time (glass dynamics is often referred to as "sluggish dynamics").

Appendix E

Homogeneous functions

E.1 Homogeneous functions of one or more variables

Let us begin with the definition of *homogeneous function*.

Definition E.1 (Homogeneous function). *A function $f(x)$ is said to be* homogeneous *if:*

$$f(\lambda x) = g(\lambda)f(r) \qquad \forall \lambda \in \mathbb{R} \tag{E.1}$$

where g is, for now, an unspecified function (we will shortly see that it has a precise form).

An example of homogeneous function is $f(x) = Ax^2$; in fact:

$$f(\lambda x) = A\lambda^2 x^2 = \lambda^2 f(x)$$

and so in this case $g(\lambda) = \lambda^2$.

A very interesting property of homogeneous functions is that once its value in a point x_0 and the function $g(\lambda)$ are known, the entire $f(x)$ can be reconstructed; in fact, any x can be written in the form λx_0 (of course with $\lambda = x/x_0$), so that:

$$f(x) = f(\lambda x_0) = g(\lambda)f(x_0)$$

We now want to show that $g(\lambda)$ has a precise form.

Theorem. *The function $g(\lambda)$ as in definition E.1 is:*

$$g(\lambda) = \lambda^p$$

Proof. From (E.1), for $\lambda, \mu \in \mathbb{R}$ we have on one hand that:

$$f(\lambda \mu x) = f(\lambda(\mu x)) = g(\lambda)f(\mu x) = g(\lambda)g(\mu)f(x)$$

but also:

$$f((\lambda \mu)x) = g(\lambda \mu)f(x)$$

and so:

$$g(\lambda \mu) = g(\lambda)g(\mu)$$

If we now suppose g to be differentiable[1], then differentiating with respect to μ this last equation we get:

$$\lambda g'(\lambda \mu) = g(\lambda)g'(\mu)$$

---

[1] In reality it would be sufficient the sole continuity of g, but in this case the proof becomes longer.

where by g' we mean the derivative of g with respect to its argument. Setting $\mu = 1$ and defining $p := g'(1)$ we have:

$$\lambda g'(\lambda) = g(\lambda)p \qquad \Rightarrow \qquad \frac{g'(\lambda)}{g(\lambda)} = \frac{p}{\lambda}$$

which yields:

$$\frac{d}{d\lambda} \ln g(\lambda) = \frac{p}{\lambda} \qquad \Rightarrow \qquad \ln g(\lambda) = p \ln \lambda + c \qquad \Rightarrow \qquad g(\lambda) = e^c \lambda^p$$

Now, $g'(\lambda) = p e^c \lambda^{p-1}$, so since $g'(1) = p$ by definition we have $p = p e^c$ and thus $c = 0$. Therefore:

$$g(\lambda) = \lambda^p$$

$\square$

A homogeneous function such that $g(\lambda) = \lambda^p$ is said to be *homogeneous of degree p*.

In case f is a function of more than one variable $x_1, \ldots, x_n$, the definition of homogeneous function changes to:

$$f(\lambda x_1, \ldots, \lambda x_n) = g(\lambda) f(x_1, \ldots, x_n)$$

E.2   Generalized homogeneous functions

In case, for example, of a function of two variables x and y, the definition of homogeneous function can be extended to:

$$f(\lambda^a x, \lambda^b y) = \lambda f(x, y) \qquad \forall \lambda, a, b \in \mathbb{R}$$

In this case f is said to be a *generalized homogeneous function* (of two variables).

Let us note that this is indeed the most general form for a generalized homogeneous function; in fact if $f(\lambda^a x, \lambda^b y) = \lambda^p f(x, y)$ then it is sufficient to call $\lambda^p = \sigma$ to get $f(\sigma^{a/p} x, \sigma^{b/p} y) = \sigma f(x, y)$, which is in the form of the definition we have given.

For the considerations that we make in chapter 7 it is important to note that from the definition of homogeneous function, since λ is arbitrary we can set $\lambda = y^{-1/b}$ and get:

$$f(x, y) = y^{1/b} f\left(\frac{x}{y^{a/b}}, 1\right)$$

which transforms the dependence of f from the two variables x and y to the sole ratio $x/y^{a/b}$.

Bibliography

[1] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Dover Publications, 2008

[2] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, 1985

[3] N. Goldenfeld, *Lectures on phase transitions and critical phenomena*, Westview Press, 1992

[4] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd edition, Butterworth-Heinemann, 1980

[5] L. Peliti, *Statistical mechanics in a nutshell*, Princeton University Press, 2011

[6] M. Plischke and B. Bergersen, *Equilibrium statistical physics*, 3rd edition, World Scientific Publishing, 2006

[7] J. P. Sethna, *Statistical mechanics: Entropy, Order Parameters, and Complexity*, Oxford University Press, 2006

[8] J. M. Yeomans, *Statistical Mechanics of Phase Transitions*, Clarendon Press, 1992