Thermodynamics and Statistical Mechanics

(script accompanying the lecture held by Prof. Dr. Kree)

Prof. Dr. Reiner Kree
Christian Hettlage

Contents

Important Notice

This version of the script "Thermodynamics and Statistical Mechanics" is a preliminary one. Hence it should come as no surprise to the reader that, for example, in parts it might have been formatted far better. At least I hope all the serious and less serious errors have been found and removed.

Alas, all such hope must be in vain. Therefore you are kindly asked to report any errors, be they orthographic or related to physics, by sending an email to me (Christian Hettlage, that is). Your mail should be directed to hettlage@uni-sw.gwdg.de.

Of course, you may feel free to produce a hardcopy of this script. But please keep in mind that there is still a final version to come, which you then again might like to print. Hence, if you would like to save on paper and ink you might decide not to produce a printed version right now. On the other hand, reading a hardcopy probably enhances the chance to find some errors, and for that rather selfish reason, I suggest that using a printer right now might be a praiseworthy idea.

Finally, I hope you enjoy reading this version of the script and that you find it helpful in your quest of understanding statistical physics.

Chapter 1

Introduction

The concept of fundamental theories such as classical mechanics, electrodynamics, or quantum mechanics is to simplify physical systems as far as possible in order to obtain a set of fundamental rules containing all the physics, i.e. they attempt to understand the world (or, more humbly, the system under consideration) in terms of basic constituents and fundamental interactions. In what might perhaps be called its most glorious triumph, this reductionistic approach has led to the conclusion that the everyday world can be described by means of three particles (electron, up quark, and down quark) and four (strong, weak, electromagnetic, and gravitational) interactions only.

However, this approach runs into difficulty if we are to explain more complex systems: An understanding of a cow on the meadow (or the inner workings of the brain of the reader) in terms of quarks and electrons is just impossible, the reason being, of course, that there are just so many of these constituents (and thus possibilities for combining them).

This clearly shows the need for synthetic theories, which *statistical physics* is an example of. The idea here is *not* to try and know all there is to know about a given system (e.g. all the positions and velocities of all its atoms), but rather to concentrate on bulk properties, accepting the need to content oneself with some average description. While arguably such an approach may seem slightly unsatisfactory, it proves tremendously powerful. Indeed, statistical physics has been one of the most sucessful theories of the 20th century.

In order to give a flavor of what to expect from statistical physics., Fig. 1.1 shows its relation to various other fields of interest. From this figure, it can be gathered that the mechanism of statistical physics works for classical

Figure 1.1: Relation of statistical physics to other fields of interest.

and quantum mechanics alike. Furthermore, it can be used with some nonphysical input such as rules governing finance markets or social systems.

One of the major fields of application is given by condensed matter physics. In addition, statistical physics may be used to shed some light on the origin of the well-known laws of thermodynamics. Hence it should come as no surprise that initially thermodynamics constituted the main impetus for its establishment.

Finally, it should be noted that coincidentally statistical physics and quantum field theory are formally equivalent: Whereas the former is governed by the Boltzmann factor $e^{-H/k_{\mathrm{B}}T}$ (where H denotes the Hamiltonian, k_{B} the Boltzmann constant, and T the temperature), in the latter, time development is described by means of $e^{(i/\hbar)Ht}$ (where H now denotes the Hamiltonian operator). Accordingly, using the replacement

$$\frac{1}{k_{\mathrm{B}}T} \longleftrightarrow \frac{i}{\hbar}Ht$$

one can map statistical physics onto quantum field theory, and vice versa.

Chapter 2

Preliminaries

Our discussion of statistical physics will require classical and quantum mechanics, as well as some probability theory. We therefore revise each of these subjects in turn, stressing the concepts which will prove useful later on.

2.1 Classical mechanics

A classical system of N (point) particles can be described by a set of N position vectors $\boldsymbol{q}_i$ and conjugated momenta $\boldsymbol{p}_i$. Assuming that the particles live in d dimensions, such a system can hence be interpreted as a single point (p, q) in the $2 \cdot d \cdot N$ dimensional phase space Γ made of the position and momentum coordinates. Here, q and p denote the vectors $(\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N)$ and $(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N)$, respectively (cf. Fig. 2.1). As classical mechanics is governed by an ordinary differential equation of first order (i.e. Newton's second law), so that the whole dynamics of a given system can be described by stating the position and the velocity of its constituents at some time t_0, it is evident that through each point of Λ there runs exactly one trajectory describing a time evolution.

Figure 2.1: Phase space. Each system of classical mechanics can be viewed as a point (p, q) in phase space. The physically allowed momenta and positions may constitute a subset of the phase space, which we denote by Γ. For simplicity, only two dimensions are shown.

In any textbook on classical mechanics (e.g., cf. [4]) it is shown that this time evolution $(q(t), p(t))$ of a system is given by the *canonical equations*

$$\frac{\mathrm{d}q_r}{\mathrm{d}t} = \frac{\partial H}{\partial p_r} \qquad , \qquad \frac{\mathrm{d}p_r}{\mathrm{d}t} = -\frac{\partial H}{\partial q_r},$$

where the q_r and p_r denote the $d \cdot N$ elements of the vectors q and p. If one employs the *Poisson bracket*

$$\{u, v\} \equiv \sum_{r=1}^{d\cdot N} \left[\frac{\partial u}{\partial p_r} \frac{\partial v}{\partial q_r} - \frac{\partial u}{\partial q_r} \frac{\partial v}{\partial p_r} \right], \tag{2.1}$$

one may paraphrase these equations as

$$\dot{q}_r = \{H, q_r\} \qquad , \qquad \dot{p}_r = \{H, p_r\}.$$

Here, H denotes the Hamiltonian. A typical example would be the Hamiltonian

$$H(p, q, t) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + \sum_{i=1}^{N} V(\boldsymbol{q}_i) + \frac{1}{2} \sum_{\substack{i,j=1 \\ (i \neq j)}}^{N} U(\boldsymbol{q}_i - \boldsymbol{q}_j)$$

of N particles of the same mass m in an external field with a two-particle interaction. Knowing the dynamics of a system, one may compute the time evolution of some observable A. If the observable depends on the positions and momenta only, $A = A(p, q)$, one obtains

$$\begin{aligned} \frac{\mathrm{d}A(p,q)}{\mathrm{d}t} = \frac{\mathrm{d}A(p(t), q(t))}{\mathrm{d}t} &= \sum_{r=0}^{d\cdot N} \left[\frac{\partial A}{\partial p_r} \frac{\mathrm{d}p_r}{\mathrm{d}t} + \frac{\partial A}{\partial q_r} \frac{\mathrm{d}q_r}{\mathrm{d}t} \right] \\ &= \sum_{r=0}^{d\cdot N} \left[\frac{\partial A}{\partial p_r} \left(-\frac{\partial H}{\partial q_r} \right) + \frac{\partial A}{\partial q_r} \frac{\partial H}{\partial p_r} \right]. \end{aligned}$$

In the last step, the canonical equations have been used. From Eq. 2.1 then follows the neat formula

$$\dot{A}(p, q) = \{H, A\}.$$

From the aforementioned it should be clear that a point $(p, q) \in \Gamma$ may be regarded as a complete microscopic description. Therefore, it is referred to as a *microstate*. Now for any system encountered in everyday life, $N \gg 1$

Figure 2.2: Ensemble in phase space. While a single point in phase space corresponds to a microstate, a statistical ensemble can be regarded as a cloud of points.

holds valid. (Remember that for example in 1 gram of water there are about 3×10^{22} molecules!) This means, however, that any attempt to describe auch a system by means of its microstate is doomed to fail. Indeed, one can even say that, given any realistic device, the preparation of a specifically chosen microstate is impossible.

In order to get around this problem, we turn from the discussion of a single system to that of a set of similar systems. To this end, let us assume that initially we have $\mathcal{N}$ identical systems. If we subject them to some realistic preparation procedure (such as "add a heat energy of 1 kJ"), the outcome will be a set of $\mathcal{N}$ slightly different systems, which we call an ensemble. In terms of phase space diagrams, such an ensemble corresponds to a cloud of points, as illustrated in Fig. 2.2.

A moment's thought will show that if you increase the number $\mathcal{N}$ of systems in an ensemble, the average distance between neighbouring points in this cloud of points decreases. In the limit of an infinite number of systems, $\mathcal{N} \longrightarrow \infty$, the discrete distribution in phase space will be indistinguishable from a continuous one. Hence it makes sense to describe an ensemble of sufficiently many systems by a continuous function $\rho(p,q,t)$ defined by

$$\forall B \subset \Gamma : \int_B \rho(p,q,t) \mathrm{d}^{d \cdot N} p \, \mathrm{d}^{d \cdot N} q \propto \text{number of ensemble points in } B$$

and the normalization

$$\int_\Gamma \rho(p,q,t) \mathrm{d}^{d \cdot N} p \, \mathrm{d}^{d \cdot N} q = 1.$$

It is easy to see that the ρ thus defined is just a probability density in phase space. Evidently it can be used to describe the average behaviour of sufficiently many systems. However, as will be discussed more thoroughly and in more detail later on, a *single* large system may be considered as an ensemble of smaller subsystems, so that its macroscopic properties arise from averaging over an ensemble. $\rho(p,q,t)$ therefore is called a *macrostate* of classical mechanics.

Figure 2.3: Example of a distribution function $F(x)$.

Figure 2.4: Example of a histogram corresponding to the distribution function shown in Fig. 2.3.

We shall now turn to the statistical treatment of measured values in a given macrostate. To this end, let us assume that we have an ensemble ρ of $\mathcal{N}_\rho$ systems and an observable $A(p,q)$. If, for the sake of simplicity, we take the measuring device used to be ideal, the measurement of A performed on the ensemble yields $\mathcal{N}_\rho$ values $A'_\mu \equiv A(p_\mu, q_\mu)$. Then the *mean* (or *average*) $\langle A \rangle_\rho$ is given by

$$\langle A \rangle_\rho \equiv \frac{1}{\mathcal{N}_\rho} \sum_{\mu=1}^{\mathcal{N}_\rho} A'_\mu, \tag{2.2}$$

whereas the *variance* is $\sigma^2_{A,\rho}$ defined as

$$\sigma^2_{A,\rho} \equiv \left\langle (A - \langle A \rangle_\rho)^2 \right\rangle_\rho = \left\langle A^2 \right\rangle_\rho - \langle A \rangle^2_\rho . \tag{2.3}$$

It is left as an exercise for the reader to prove the right hand side of Eq. 2.3 by means of Eq. 2.2. In addition, we may introduce the *distribution* $F_{A,\rho}$:

$$F_{A,\rho}(A') \equiv \text{number of systems } \mu \text{ with } A_\mu \leqslant A' \tag{2.4}$$

See Fig. 2.3 for an example of a distribution. We note in passing that distributions are by definition monotonously increasing function. If we divide the range of possible values of A into a number of bins ΔA_i and plot the number of A_μ's in these bins, we obtain a *histogram.* Fig. 2.4 shows an example.

The generalization of Eqs. 2.2 to 2.4 to the case of infinitely many systems, i.e. to the case of a macrostate is straightforward, one basically just has to replace sums by integrals. But in order to avoid some cumbersome notation, we first introduce a short notation for the volume element in phase space:

Definition 1 *Let* Γ *be a* $2 \cdot d \cdot N$ *dimensional phase space. Then we denote its differential volume element by* $\mathrm{d}\Gamma_{\mathrm{ph}}$*, i.e.*

$$\mathrm{d}\Gamma_{\mathrm{ph}} \equiv \mathrm{d}^{d\cdot N} p \, \mathrm{d}^{d \cdot N} q.$$

In the following definitions, we assume that Γ is a phase space. We can now state the definition of the average:

Definition 2 *Let ρ be a macrostate and $A = A(p,q)$ an observable. Then the* average *(or* mean*) of A is given by*

$$\langle A \rangle_\rho \equiv \int_\Gamma A(p,q)\rho(p,q,t)\mathrm{d}\Gamma_{\mathrm{ph}}.$$

The definition of the variance remains completely unchanged:

Definition 3 *The* variance *$\sigma_\rho(A)$ of an observable $A = A(p,q)$ is defined for a macrostate ρ as*

$$\sigma^2_{A,\rho} \equiv \left\langle (A - \langle A \rangle_\rho)^2 \right\rangle_\rho = \left\langle A^2 \right\rangle_\rho - \langle A \rangle^2_\rho .$$

For the distribution we get

Definition 4 *The* distribution *$F_{A,\rho}$ of an observable $A = A(p,q)$ for a macrostate ρ is given by*

$$F_{A,\rho}(A') \equiv \int_\Gamma \Theta(A' - A(p,q))\rho(p,q,t)\mathrm{d}\Gamma_{\mathrm{ph}} = \langle \Theta(A' - A(p,q)) \rangle_\rho ,$$

where Θ is the Heaviside function,

$$\Theta(x) = \left\{ \begin{array}{ll} 0 & (x < 0) \\ 1 & (x \geq 1) \end{array} \right. .$$

So far, we have carefully distinguished between the observable A (which is a function) and its possible values (for which we used the symbol A'). In the following, we shall drop this distinction, so that A may stand for either the observable or its values. Furthermore, we omit the subscript $_A$. We then can proceed with introducing the probability density:

Definition 5 *The* probability density *p_ρ of an observable A for a macrostate ρ is the derivative of the distribution:*

$$p_\rho(A) \equiv \frac{\mathrm{d}F_\rho(A)}{\mathrm{d}A} = \int_\Gamma \delta(A - A(p,q))\rho(p,q,t)\mathrm{d}\Gamma_{\mathrm{ph}} = \langle \delta(A - A(p,q)) \rangle_\rho$$

While the concept of a distribution should be immediately clear, the reader may rightfully what is the meaning of the various statistical quantities introduced so far. Alas, concerning the mean and the variance, an answer has to be postponed until Sect. 2.3. But at least we can briefly comment on the probability density.

Consider a histogram bin from, say, A' to $A'+\Delta A$. Then the corresponding histogram value $h(A, A+\Delta A)$ can be computed by means of

$$\begin{aligned} h(A, A+\Delta A) &= \int_\Gamma \left[\Theta(A+\Delta A - A(p,q)) - \Theta(A - A(p,q))\right] \rho(p,q,t) \mathrm{d}\Gamma_{\mathrm{ph}} \\ &= F_\rho(A+\Delta A) - F_\rho(A). \end{aligned}$$

But then we can conclude that

$$\lim_{\Delta A \longrightarrow 0} \frac{h(A, A+\Delta)}{\Delta A} = \lim_{\Delta A \longrightarrow 0} \frac{F_\rho(A+\Delta A) - F_\rho(A)}{\Delta A} = \frac{\mathrm{d}F_\rho(A)}{\mathrm{d}A} = p_\rho(A).$$

In this sense we may say that the probability density is the limit of histograms for vanishing bin width. Finally, we conclude with two remarks:

(i) The probability density corresponding to the microstate (p_0, q_0) in a $2 \cdot d \cdot N$ dimensional phase space obviously is given by the delta function $\rho(p,q) = \delta^{d \cdot N}(p-p_0)\delta^{d \cdot N}(q-q_0)$. Putting this ρ into Definitions 1 to 4, we recover Eqs. 2.2 to 2.4, as should be expected.

(ii) The probability density changes non-trivially under coordinate transformations. To see this, we consider the coordinate transformation $(p,q) \longrightarrow (p',q')$ and compute the mean of some observable $A(p,q)$

(which must be independent of the specific coordinates chosen) in both old and new coordinates:

$$\begin{aligned}
\langle A\rangle_\rho &= \int_\Gamma A(p,q)\rho(p,q)\mathrm{d}^{d\cdot N}p\,\mathrm{d}^{d\cdot N}q \\
&= \int_\Gamma A(p(p',q'),q(p',q'))\cdot\rho(p(p',q'),q(p',q'),t)\cdot \\
&\qquad\qquad \cdot\underbrace{\left|\det\left(\frac{\partial(p,q)}{\partial(p',q')}\right)\right|}_{\text{Jacobian}}\cdot\mathrm{d}^{d\cdot N}p'\mathrm{d}^{d\cdot N}q' \\
&\overset{(1)}{=} \int_\Gamma A'(p',q')\cdot\rho(p(p',q'),q(p',q'),t)\cdot \\
&\qquad\qquad \cdot\left|\det\left(\frac{\partial(p,q)}{\partial(p',q')}\right)\right|\cdot\mathrm{d}^{d\cdot N}p'\mathrm{d}^{d\cdot N}q' \\
&\overset{(2)}{=} \int_\Gamma A'(p',q')\mathrm{d}^{d\cdot N}p'\mathrm{d}^{d\cdot N}q'.
\end{aligned} \tag{2.5}$$

Here, step (1) follows from the fact that the observable transforms as

$$A'(p',q') = A(p(p',q'),q(p',q')),$$

whereas in step (2) we make use of the fact that the mean of the observable must be an invariant. As Eq. 2.5 must hold valid for all observables, for the probability density we obtain the transformation law

$$\rho'(p',q') = \rho(p,q)\left|\det\left(\frac{\partial(p,q)}{\partial(p',q')}\right)\right|.$$

2.2 Quantum mechanics

We saw in Sect. 2.1 that in classical mechanics, a microstate (i.e., as the reader will undoubtedly remember, “all there is to know about a given system”) can be regarded as a point in phase space. Things are somewhat more complicated in quantum mechanics. Here, a microstate corresponds to a normalized vector $|\psi\rangle$ in a (appropriately chosen) Hilbert space $\mathcal{H}$. To keep things simple, we shall assume throughout the following that the Hilbert space under consideration has a countable basis.

In order to paraphrase the basic postulate of the preceding paragraph in a slightly different form, we start with reminding the reader of the definition of a projecton operator.

Definition 6 *An operator $\hat{P}$ on a Hilbert space $\mathcal{H}$ is a* projection operator *(or* projector*) if it fulfils the following conditions:*

(i) $\hat{P} = \hat{P}^\dagger$

(ii) $\hat{P}^2 = \hat{P}$

In addition, the trace of a projector onto a direction *has the value 1:*

(iii) $\operatorname{tr} \hat{P} = 1$

The reader might like to check that this definition implies that in $\mathbb{R}^n$ a projector projects onto some fixed direction, which is after all what you would expect a projector to do.

Now consider a projector $\hat{P}$ onto some arbitrary direction. Then from condition (ii) of Def. 6 it follows that for an eigenvector $|p\rangle$ of $\hat{P}$

$$(p^2 - p)\,|p\rangle = \hat{P}^2\,|p\rangle - \hat{P}\,|p\rangle = \hat{P}\,|p\rangle = \hat{P}\,|p\rangle = 0,$$

so that $p^2 - p = 0$. Hence, $\hat{P}$ can only have 0 and 1 as its eigenvalues. But then condition (iii) of Def. 6 implies that there is exactly one eigenvector $|\psi\rangle$ with eigenvalue 1, as otherwise the trace would be either zero or larger than 1. Accordingly, $\hat{P}$ may be written as $\hat{P} = |\psi\rangle\,\langle\psi|$.

Conversely, it is easy to show that for any operator of the form $|\psi\rangle\,\langle\psi|$, the three conditions of Def. 6 hold valid. Thus there is a one-to-one correspondence between normalized vectors and projectors onto a direction. In order to reinforce this notion, in the following we shall employ the notation $P_\psi \equiv |\psi\rangle\,\langle\psi|$. Now the postulate of the first paragraph can be restated in a different way:

Postulate 1 *Each microstate corresponds to a projector onto a direction on some Hilbert space $\mathcal{H}$.*

Whereas in classical mechanics, ordinary functions are sufficient to describe a measurement, things are again more difficult in quantum mechanics:

Postulate 2 *Every observable A corresponds to some Hermitean operator $\hat{A}$ on $\mathcal{H}$. The possible values obtained by a measurement of the observable are given by the eigenvalues of $\hat{A}$.*

As the operator $\hat{A}$ corresponding to the observable A is Hermitean, there must be a complete set $(\left|\alpha_1\right\rangle, \left|\alpha_2\right\rangle, \ldots, \left|\alpha_k\right\rangle \ldots)$ of orthonormal eigenvectors $\left|\alpha_i\right\rangle$ with eigenvalues α_i. Postulate 2 has left open the question how probable it is to obtain some given value α_i if a measurement of an observable is performed on a microstate. However, there is a simple answer to that question:

Postulate 3 *Let $\hat{P}_\psi$ be a microstate and $\hat{A}$ the operator corresponding to some observable A. Furthermore, let $\{\left|\alpha_1\right\rangle, \left|\alpha_2\right\rangle, \ldots\}$ be a complete set of orthonormal eigenvectors of $\hat{A}$. Then if a measurement of A is performed on the given microstate, the probability of obtaining the value α_i is given by $\left|\left\langle\alpha_i\right|\left.\psi\right\rangle\right|^2$.*

The importance of Postulates 2 and 3 shouldn't be underestimated. Together they imply that the outcome of a measurement is *not* fully determined by a microstate. Loosely speaking, we might say that in quantum mechanics generally even nature itself cannot predict the result of measuring an observable.

Hence, for quantum mechanical systems an ensemble treatment is called for at the level of microstates. We therefore turn to the computation of the mean of an observable (which in this respect often is referred to as the *expectation value*). To this end, let h denote the dimension of $\mathcal{H}$ (which may be infinite). Then for a microstate P_ψ

$$\langle A\rangle = \sum_{i=1}^{h} \left|\left\langle\alpha_i\right|\left.\psi\right\rangle\right|^2 \alpha_i = \sum_{i=1}^{h} \alpha_i \left\langle\psi\right|\left.\alpha_i\right\rangle \left\langle\alpha_i\right|\left.\psi\right\rangle = \left\langle\psi\right| \left\{ \sum_{i=1}^{h} \alpha_i \left|\alpha_i\right\rangle \left\langle\alpha_i\right| \right\} \left|\psi\right\rangle$$

But the term in the curled brackets is just the operator $\hat{A}$, and we get the well-known result

$$\langle A\rangle = \left\langle\psi\right| \hat{A} \left|\psi\right\rangle . \tag{2.6}$$

We may rewrite this result in another form by noting that for a complete set $\{|\alpha\rangle\}$ of orthonormal vectors we get

$$\begin{aligned}\langle\psi|\,\hat{A}\,|\psi\rangle &= \langle\psi|\underbrace{\left\{\sum_{\alpha}|\alpha\rangle\,\langle\alpha|\right\}}_{1}\hat{A}\,|\psi\rangle = \sum_{\alpha}\langle\psi|\,\alpha\rangle\,\langle\alpha|\,\hat{A}\,|\psi\rangle \\ &= \sum_{\alpha}\langle\alpha|\,\hat{A}\,|\psi\rangle\,\langle\psi|\,\alpha\rangle = \sum_{\alpha}\langle\alpha|\left\{\hat{A}\hat{P}_{\psi}\right\}|\alpha\rangle\,,\end{aligned} \tag{2.7}$$

where $\sum_{\alpha}$ denotes the sum over all vectors $|\alpha\rangle$. As the right hand side of Eq. 2.7 is just a trace, together with Eq. 2.6 we obtain

$$\langle A\rangle = \operatorname{tr}\hat{A}\hat{P}_{\psi}.$$

Similarly to the case of classical mechanics, also in quantum mechanics it is impossible to prepare some specific microstate. We therefore again have to resort to investigating a statistical ensemble. Here, in order to avoid rather technical problems, let us take for granted that the preparation procedure used to generate our ensemble ρ allows a *countable* set $M = \{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle, \ldots\}$ of possible outcomes only. The ensemble considered shall consist of $\mathcal{N}_{\rho}$ systems, $E = \{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_{\mathcal{N}_{\rho}}\rangle\}$, where for all i $|\phi_i\rangle \in M$ and where $|\psi_{\gamma}\rangle$ occurs $\mathcal{N}_{\gamma}$ times in E. Then the mean of an observable A is given by (cf. Eqs. 2.2, 2.6 and 2.7)

$$\langle A\rangle_{\rho} = \frac{1}{\mathcal{N}_{\rho}}\sum_{i=1}^{\mathcal{N}_{\rho}}\langle\phi_i|\,\hat{A}\,|\phi_i\rangle = \frac{1}{\mathcal{N}_{\rho}}\sum_{|\psi_{\gamma}\rangle\in M}\mathcal{N}_{\gamma}\,\langle\psi_{\gamma}|\,\hat{A}\,|\psi_{\gamma}\rangle$$

If we define w_{γ} as $w_{\gamma} \equiv \mathcal{N}_{\gamma}/\mathcal{N}_{\rho}$, we thus get

$$\begin{aligned}\langle A\rangle_{\rho} &= \sum_{|\psi_{\gamma}\rangle\in M} w_{\gamma}\,\langle\psi_{\gamma}|\,\hat{A}\,|\psi_{\gamma}\rangle = \sum_{|\psi_{\gamma}\rangle\in M} w_{\gamma}\operatorname{tr}(\hat{A}\hat{P}_{\psi_{\gamma}}) \\ &= \operatorname{tr}\left(\left\{\sum_{\psi_{\gamma}\in M} w_{\gamma}\hat{P}_{\psi_{\gamma}}\right\}\hat{A}\right).\end{aligned} \tag{2.8}$$

As the operator appearing in the curled brackets is of considerable importance, we give it a name of its own:

Definition 7 *The* density matrix *(or* state operator*) of the ensemble just decribed is given by*

$$\hat{\rho} \equiv \sum_{\psi_\gamma \in M} w_\gamma \hat{P}_{\psi_\gamma}.$$

With the density matrix, we may cast Eq. 2.8 in a compact form:

$$\langle A \rangle_\rho = \mathrm{tr}(\hat{\rho}\hat{A}).$$

The density matrix contains the whole statistics of any observable A and may thus be considedered as a macrostate. To see this, we write $\hat{A}$ in terms of its eigenvectors $|\alpha_i\rangle$ and eigenvalues α_i, i.e. $\hat{A} = \sum_{i=1}^{h} \alpha_i |\alpha_i\rangle \langle\alpha_i|$ with $h = \dim \mathcal{H}$. Then a function f of $\hat{A}$ can be defined by

$$f(\hat{A}) \equiv \sum_{i=1}^{h} f(\alpha_i) |\alpha_i\rangle \langle\alpha_i| .$$

In particular, for the Dirac δ-function we get

$$\delta(a - \hat{A}) = \sum_{i=1}^{h} \delta(a - \alpha_i) |\alpha_i\rangle \langle\alpha_i| . \tag{2.9}$$

Now the expectation value of some function f of the observable A for a microstate $|\psi\rangle$ has the value

$$\langle f(A) \rangle = \sum_{i=1}^{h} |\langle\psi| \alpha_i\rangle|^2 f(\alpha_i) = \langle\psi| \left\{ \sum_{i=1}^{h} f(\alpha_i) |\alpha_i\rangle \langle\alpha_i| \right\} |\psi\rangle = \langle\psi| f(\hat{A}) |\psi\rangle$$

Hence from an argument completely analogous to that of Eq. 2.8 we may conclude that in an ensemble ρ the average of $f(A)$ is given by

$$\langle f(A) \rangle_\rho = \mathrm{tr}(\hat{\rho} f(\hat{A})).$$

But then from Def. 5 and Eq. 2.9 we obtain for the probability density $p_{\rho,A}$ of A

$$p_{\rho,A}(a) = \langle \delta(a - A) \rangle_\rho = \mathrm{tr} \left(\sum_{i=1}^{h} \delta(a - \alpha_i) |\alpha_i\rangle \langle\alpha_i| \hat{\rho} \right) .$$

As $p_{\rho,A}$ completely describes the statistical behavior of A, it is thus clear that $\hat{\rho}$ indeed fully determines the statistics of the ensemble ρ.

We finish this section by stating three basic properties of density matrices:

(i) A density matrix is Hermitean, $\rho = \rho^\dagger$.

(ii) For a density matrix $\hat{\rho} = \sum_\gamma w_\gamma P_{\psi_\gamma}$ all w_γ lie in the range from 0 to 1.

(iii) All density matrices $\hat{\rho}$ are non-negative, i.e.

$$\forall \left|\psi\right\rangle \in \mathcal{H} : \; \left\langle\psi\right| \hat{\rho} \left|\psi\right\rangle \geqslant 0.$$

The proof is left an exercise for the reader.

2.3 Probability theory

2.3.1 Kolmogorov's axioms

Intuitively, the probability of some event E is the relative frequency with which it occurs, i.e.

$$\text{probability of event } E \equiv \frac{\text{number of occurences of } E}{\text{number of trials}}.$$

But if the number of trials is small, there will be a large statistical uncertainty in the actual number of occurences of E. This should be mirrored in the definition of the probability, and hence it is tempting to resort to the limit of an infinite number of trials:

$$\text{probability of event } E \equiv \lim_{\text{number } N \text{ of trials} \longrightarrow \infty} \frac{\text{number of occurences of } E}{N}$$

Evidently, this approach *assumes* that this limit converges, and a rather skeptical reader might ask whether that is necessarily the case. Furthermore, things get slightly cumbersome if we are not dealing with a limited number of events (such as the number of spots on the uppermost face of a die, say) but rather with a continuum of events (such as the lifetime of a bulb).

Therefore we will adopt a mathematically-oriented approach to probability theory (cf. [5]). To this end, let us first introduce two straightforward definitions:

Definition 8 *An* experiment *is any operation the outcome of which is not completely determined.*

Definition 9 *The* sample space Ω *of an experiment is the set of its possible outcomes.*

From the plethora of possible examples we just choose the following three:

(i) One of the simplest experiments is throwing a coin and looking which side comes up. Here, the sample space is given by

$$\Omega = \{\text{heads come up}, \text{tails come up}\}.$$

(ii) If the experiment consists of rolling a die and counting the number of spots on its uppermost face, the sample space can be written as

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

(iii) The lifetime of a single radioactive nucleus is undetermined, so that its measurement constitutes an experiment. It has $\Omega = \mathbb{R}_0^+$ as its sample space.

Note that Ω doesn't have to be a set of numbers. Indeed, we might have written the sample space of example (ii) more explicitly as

$$\begin{aligned} \Omega = \{&\text{The number of spots on the uppermost face is 1.},\\ &\ldots, \text{The number of spots on the uppermost face is 6.}\}. \end{aligned}$$

If $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k, \ldots\}$ is countable, one may assign to each $\omega_k \in \Omega$ a number P_k with $0 \leqslant P_k \leqslant 1$, so that $\sum_{\omega_i \in \Omega} P_i = 1$. These numbers P_i are referred to as the *probabilities* of the experiment under consideration.

In order to cover the general case of a continuous Ω we need some further prerequisites. We start wth the concept of events:

Definition 10 *A subset of a sample space* Ω *is called an* event. *We say that* an event occurs *if the outcome of the experiment is an element of* E.

Perhaps the notion of an event as a subset of sample space seems somewhat counterintuitive. But it makes sense: Consider, for example, the rolling of a die. Here perhaps we might be interested not in the actual number of spots on its uppermost face, but rather in whether this number is odd. "The number is odd." may then qualify as an event, and as this statement corresponds to the subset $\{1, 3, 5\}$ of Ω, all is well.

Now for a moment things will get rather technical. We turn to the definition of a σ-algebra:

Definition 11 *A non-empty system $\mathcal{S}$ of subsets of a set M is called a σ-algebra, if it fulfils the following conditions:*

(i) M and the empty set $\varnothing$ are elements of $\mathcal{S}$, $M \in \mathcal{S}$ and $\varnothing \in \mathcal{S}$.

(ii) If A is member of $\mathcal{S}$, so is its complement:

$$A \in \mathcal{S} \;\Rightarrow\; M \backslash A \in \mathcal{S}$$

(iii) Let $A_1, A_2, \ldots, A_k, \ldots \in \mathcal{S}$. Then their union and intersection are also elements of $\mathcal{S}$:

$$A_i \in \mathcal{S}\ \forall i \in \mathbb{N} \;\Rightarrow\; \bigcup_{i=1}^{\infty} A_i \in \mathcal{S} \textit{ and } \bigcap_{i=1}^{\infty} A_i \in \mathcal{S}$$

One may construct a σ-algebra from a countable set by forming all unions of its subsets. Similarly, a σ-algebra is generated from an interval $[a, b]$ by forming the unions and intersections of all closed subintervals.

What do we need Def. 11 for? To give an answer to this question, we must first note that for a given experiment the set of all events constitutes a σ-algebra, as all the required conditions follow from basic logic. But the existence of a σ-algebra is sufficient for the definition of a measure. Hence we can introduce a measure on the set of all events, and that is precisely what we'll do in a moment.

However, before doing so, let us state that intuitively any probability should have three fundamental properties:

(i) Any probability is non-negative.

(ii) The probability that something (no matter what) happens in an experiment is one.

(iii) The probability that either of two completely distinct events happens is the sum of the probabilities for the two events. To put in a nutshell: Probability is additive.

These properties, which are so profound that one could scarcely do without, are the rationale for the general definition of probability:

Definition 12 (Kolmogorov's axioms) *Let $\mathcal{S}$ be the set of all events of some sample space Ω. Then a function $P: \mathcal{S} \longrightarrow \mathbb{R}$ is called a* probability function, *if it is a normalized measure, i.e. if the following conditions are fulfilled:*

(i) $P(A) \geqslant 0 \; \forall A \in \mathcal{S}$

(ii) $P(\Omega) = 1$

(iii) Let $\{A_i\}$ *be a countable set of events with* $A_n \cup A_m = \varnothing$ *for* $n \neq m$. *Then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Now the reason for our discussion of σ-algebras should be clear: As the set of all events constitutes a σ-algebra, as the existence of a σ-algebra allows the definition of measure, and as the probability function may be regarded as a (normalized) measure, we can rest assured that a probability function really exists.

Maybe the reader has wondered why we have stressed that the probability of the sample space must be one, while completely ignoring the fact that the probability of the empty set should vanish (after all, the probability that nothing happens in an experiment is zero). The reason for this restraint becomes evident from the following theorem:

Theorem 1 *Let $A \in \Omega$ be an event and P a probability function. Then $P(\Omega \backslash A) = 1 - P(A)$.*

Proof: The proof is straightforward: We know that $\Omega = A \cup (\Omega \backslash A)$ with $A \cap (\Omega \backslash A) = \varnothing$. Hence we can conclude from conditions (2) and (3) of Def. 12 that

$$1 = P(\Omega) = P(A) + P(\Omega \backslash A),$$

which is the desired result. *(QED)*

In particular, because of $\varnothing = \Omega \backslash \Omega$, Theorem 1 implies that

$$P(\varnothing) = 1 - P(\Omega) = 1 - 1 = 0.$$

So there was no reson for bothering about the probability of the empty set, as we get this as a bonus, anyway.

Figure 2.5: Principle of a measurement. By means of a measuring device every element of the sample space Ω is related to some number.

2.3.2 Random variables

As was pointed out in Sect. 2.3.1, the elements of a sample space Ω need not be numbers, and the same then must be true for events. However, in a real experiment we are interested in obtaining some numerical value. Of course, this aim is achieved by the measuring device used. Fig. 2.5 illustrates the situation.

Mathematically speaking, a measurement M may thus be regarded as mapping the sample space Ω into $\mathbb{R}$, so that we may describe it by means of an ordinary function

$$M:\ \Omega \ni \omega \longrightarrow M(\omega) \in \mathbb{R}$$

Any such function is called a *stochastic* or *random variable*. One should be careful not to confuse it with one of its values. In order to avoid this pitfall, in the following *we shall denote random variables by upper case letters and their values by the corresponding lower case letters.*

Now the distribution function of an observable A at some value a is just the probability that A has a value less than a. However, the state of a (random) system (or its preparation) may be viewed as an experiment, and we might think of a measuring device which faithfully records the values of the observable to be considered. Hence, it is evident that any observable gives rise to a corresponding random variable, which we shall denote by the same letter. Concerning the the distribution function F_A this implies that we may write

$$F_A(a) = P(\{\omega \in \Omega \,|\, A(\omega) \leqslant a\})$$

As above, the probability density p_A of a random variable (or observable) is defined just as the derivative of the distribution:

$$p_A(a) \equiv \frac{\mathrm{d}F_A(a)}{\mathrm{d}a} = \frac{\mathrm{d}}{\mathrm{d}a} P(\{\omega \in \Omega \,|\, A(\omega) \leqslant a\})$$

Now imagine we are not interested in the values x of some random variable X but rather in a real-valued one-to-one function f of these values, i.e. we are interested in the distribution of $f(x)$. Clearly, as f is of the form $f:\ \mathbb{R} \longrightarrow \mathbb{R}$, the function

$$Y \equiv f \circ X:\ \Omega \ni \omega \longrightarrow y = f(X(\omega)) \in \mathbb{R}$$

is a random variable, the probability density of which describes the desired distribution of the values $y = f(x)$. Its distribution function can be written as

$$F_Y(y) = P(\{\omega \in \Omega \,|\, Y(\omega) \leqslant y\}). \tag{2.10}$$

We have to distinguish between two cases: As it is one-to-one, f may be either monotonously increasing or monotonously decreasing. In the former case we obtain from Eq. 2.10

$$F_Y(y) = P(\{\omega \in \Omega \,|\, X(\omega) \leqslant f^{-1}(y)\}) = F_X(f^{-1}(y)),$$

whereas using Theorem 1 in the latter case we get

$$\begin{aligned} F_Y(y) &= P(\{\omega \in \Omega \,|\, X(\omega) \geqslant f^{-1}(y)\}) = 1 - P(\{\omega \in \Omega \,|\, X(\omega) \leqslant f^{-1}(y)\}) \\ &= 1 - F_X(f^{-1}(y)). \end{aligned}$$

Differentiating both sides of these equations, taking into account the chain rule and the definition of the probability density, we arrive at

$$p_Y(y) = p_X(f^{-1}(y)) \left| \frac{\mathrm{d}f(y)}{\mathrm{d}y} \right|^{-1}, \tag{2.11}$$

where we employed the relations $\mathrm{d}(f^{-1})/\mathrm{d}y = (\mathrm{d}f/\mathrm{d}y)^{-1}$ and (for a function which decreases monotonously) $\mathrm{d}f/\mathrm{d}y = -|\mathrm{d}f/\mathrm{d}y|$. Fortunately, the generalization of random variables to random vectors is straightforward. We just give the relevant definitions and a theorem corresponding to Eq. 2.11:

Definition 13 *An N-dimensional* random vector $\boldsymbol{X}$ *is a function of the form*

$$\boldsymbol{X}:\ \Omega \ni \omega \longrightarrow \boldsymbol{X}(\omega) \in \mathbb{R}^N.$$

Definition 14 *The* distribution function $F_{\boldsymbol{X}}$ *of an N-dimensional random vector $\boldsymbol{X}$ is given by*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \equiv P(\{\omega \in \Omega \,|\, X_1(\omega) \leqslant x_1, \ldots, X_N(\omega) \leqslant x_N\}).$$

The corresponding probability density $p_{\boldsymbol{X}}$ *is defined as*

$$p_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \frac{\partial^N F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial x_1 \ldots \partial x_N}.$$

Theorem 2 *Let $\boldsymbol{X}$ and f be an N-dimensional random vector and an invertible transformation, respectively Then the probability distribution of the random vector $\boldsymbol{Y} \equiv f(\boldsymbol{X}) \equiv f \circ \boldsymbol{X}(\boldsymbol{x})$ can be computed from*

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = p_{\boldsymbol{X}}(f^{-1}(\boldsymbol{y})) \left| \det \frac{\partial(y_1, \ldots, y_N)}{\partial(x_1, \ldots, x_N)} \right| .$$

To keep things as simple as possible, in the following we shall limit ourselves to the one-dimensional case again. Before going on, though, let us remark on two things:

(i) There can be rather non-trivial sets of probability 0. (Note, for example, that $\mathbb{Q}$ has the measure 0 in $\mathbb{R}$.)

(ii) As usual, the discrete case can be recovered from the continuous one. The probability density of a discrete random variable is of the form

$$p_X(x) = \sum_\alpha P_\alpha(x - x_\alpha),$$

where the P_α are the probabilities of the x_α.

We may now define the mean of random variables:

Definition 15 *Let X be a random variable. Then the* mean *or* average *of X is given by*

$$\langle X \rangle \equiv \int p_X(x) x \, \mathrm{d}x.$$

As the function of a random variable is itself a random variable, its mean must be given by the preceding definition. However, we may rewrite it in a far more suitable way:

Theorem 3 *Let X be a random variable and f a real-valued function. Then the mean of $f(X)$ has the value*

$$\langle f(X) \rangle = \int p_X(x) f(x) \mathrm{d}x.$$

Proof: We divide $\mathbb{R}$ into disjoint intervals I on which f is either strictly monotonous or constant. Let us first turn to the strictly monotonous case. Here, we can assume that f is one-to-one, and by using the inverse of the transformation law of Theorem 2 we obtain

$$\int_{f(I)} p_{Y,x\in I}(y)y\,\mathrm{d}y = \int_I p_{Y,x\in I}(f(x))\left|\frac{\mathrm{d}f}{\mathrm{d}x}\right|\mathrm{d}x = \int_I p_X(x)f(x)\mathrm{d}x,$$

where $p_{Y,x\in I}$ denotes the probability density of the random variable $Y \equiv f(X)$ under the constraint $X(\omega)\in I$. Now for all $y\notin f(I)$ $p_{Y,x\in I}(y)=0$, so that we may write

$$\int_{\mathbb{R}} p_{Y,x\in I}(y)y\,\mathrm{d}y = \int_I p_X(x)f(x)\mathrm{d}x.$$

On the other hand, for a constant function $f(x)\equiv y_0$ on I it follows that

$$\int_{f(I)} p_{Y,x\in I}(y)y\,\mathrm{d}y = y_0\cdot P(\{\omega\in\Omega\,|\,f(X(\omega))=y_0\wedge X(\omega)\in I\}).$$

But the condition $f(X(\omega))=y_0$ is fulfilled by definition, and of course $p_{Y,x\in I}(y)$ vanishes for all $y\notin I$. Hence again we get

$$\begin{aligned}\int_{\mathbb{R}} p_{Y,x\in I}(y)y\,\mathrm{d}y &= \int_{f(I)} p_{Y,x\in I}(y)y\,\mathrm{d}y = y_0\cdot P(\{\omega\in\Omega\,|\,X(\omega)\in I\})\\ &= y_0\int_I p_X(x)\mathrm{d}x = \int_I p_X(x)f(x)\mathrm{d}x.\end{aligned}$$

A summation over all intervals I yields

$$\begin{aligned}\langle f(X)\rangle = \langle Y\rangle &= \int_{\mathbb{R}} p_Y(y)y\,\mathrm{d}y = \int_{\mathbb{R}}\left\{\sum_I p_{Y,x\in I}(y)\right\}y\,\mathrm{d}y\\ &= \sum_I\int_{\mathbb{R}} p_{Y,x\in I}(y)y\,\mathrm{d}y = \sum_I\int_I p_X(x)f(x)\mathrm{d}x = \int_{\mathbb{R}} p_X(x)f(x)\mathrm{d}x,\end{aligned}$$

which is the desired result. *(QED)*

Note that instead of $\langle f(X)\rangle$ one often simply writes $\langle f(x)\rangle$. For further reference, we need to generalize Theorem 3 to the case of random vectors.

Theorem 4 *Let* $\boldsymbol{X}$ *be an* N*-dimensional random vector and* f *some real-valued function. Then the mean of* $f(X)$ *has the value*

$$\langle f(\boldsymbol{X})\rangle = \int p_{\boldsymbol{X}}(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}^N x.$$

Proof: We give a hand-waving argument showing the correctness of the proposition. To this end, let us divide $\mathbb{R}^N$ into intervals Δ_i centered on vectors $\boldsymbol{x}^{(i)}$,

$$\Delta_i \equiv \{\boldsymbol{x} \,|\, x_1^{(i)} - \delta/2 \leqslant x_1 \leqslant x_1^{(i)} + \delta/2, \ldots, x_N^{(i)} - \delta/2 \leqslant x_N \leqslant x_N^{(i)} + \delta/2\}.$$

Here, assume that the set of Δ_i's is countable. If δ is sufficiently small, we may approximate f by the "step function" $\bar{f}$ defined as

$$\bar{f}(\boldsymbol{x}) \equiv f(\boldsymbol{x}_i) \text{ for the } i \text{ with } \boldsymbol{x} \in \Delta_i,$$

which leads to

$$\begin{aligned} F &\equiv \int_{\mathbb{R}^N} f(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}^N x \approx \int_{\mathbb{R}^N} \bar{f}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}^N x = \sum_i \int_{\Delta_i} \bar{f}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}^N x \\ &= \sum_i f(\boldsymbol{x}_i) \int_{\Delta_i} p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}^N x = \sum_i f(\boldsymbol{x}_i) \cdot P(\{\omega \in \Omega \,|\, \boldsymbol{X}(\omega) \in \Delta_i\}). \end{aligned}$$

The $f(\boldsymbol{x}_i)$ give rise to the set $\{f_j \,|\, \exists i :\ f_j = f(\boldsymbol{x}_i)\}$, where $f_i \neq f_j$ for $i \neq j$ is assumed. With this set of values of f we obtain

$$F = \sum_j f_j \cdot \sum_{\substack{i \\ \bar{f}(\boldsymbol{x}) = f_j}} P(\{\omega \in \Omega \,|\, X(\omega) \in \Delta_i\}),$$

and due to the fact that the $P(\{\omega \in \Omega \,|\, X(\omega) \in \Delta_i\})$ are piecewise disjoint and together cover the whole of $\mathbb{R}^N$, we may conclude that

$$\begin{aligned} F &= \sum_j f_j P(\{\omega \in \Omega \,|\, \bar{f}(\omega) = f_j\}) = \int y \cdot p_{\bar{f}(\boldsymbol{X})}(y) \mathrm{d}y \approx \int y \cdot p_{f(\boldsymbol{X})}(y) \mathrm{d}y \\ &= \langle f(\boldsymbol{X})\rangle \end{aligned}$$

Hence in the limit of vanishing Δ_i's the proposition follows. *(QED)*

As an important example of functions of random variables we introduce the moments:

Definition 16 *The n-th* moment μ_n *of a random variable X is defined as*

$$\mu_n \equiv \langle X^n \rangle \, .$$

The moments can be obtained from differentiating what is known as the characteristic function:

Definition 17 *The* characteristic function χ_X *of a random variable X is the Fourier transform of the probability density,*

$$\chi_X(k) \equiv \int e^{-ikx} p_X(x) \mathrm{d}x = \left\langle e^{-ikX} \right\rangle .$$

Alternatively, the probability density may be viewed as the inverse Fourier transform of the characteristic function,

$$p_X(x) = \int \frac{\mathrm{d}k}{2\pi} e^{ikx} \chi_X(k).$$

χ_X *is a* generating function *for the moments:*

$$\left. \frac{\partial^n \chi_X(k)}{\partial k^n} \right|_{k=0} = (-i)^n \langle X^n \rangle$$

If we want to, we may express the probability density of a random variable by means of the delta function. This is shown by the following theorem.

Theorem 5 *Let X and f be a random variable and a real-valued function, respectively. Then*

$$p_{f(X)}(f(x)) = \langle \delta(f(x) - f(X)) \rangle \, .$$

Proof: Let $y \equiv f(x)$. Then from Def. 17 we get

$$\begin{aligned} p_{f(X)}(f(x)) = p_{f(X)}(y) &= \int \frac{\mathrm{d}k}{2\pi} e^{iky} \chi_{f(X)}(k) = \int \frac{\mathrm{d}k}{2\pi} e^{iky} \left\langle e^{-ikf(X)} \right\rangle \\ &= \int \frac{\mathrm{d}k}{2\pi} e^{iky} \int \mathrm{d}x \, p_X(x) e^{-ikf(x)} = \int \mathrm{d}x \, p_X(x) \int \frac{\mathrm{d}k}{2\pi} e^{iky} e^{-ikf(x)} \\ &= \int \mathrm{d}x \, p_X(x) \delta(y - f(x)) = \langle \delta(y - f(X)) \rangle \, , \end{aligned}$$

where in the last step we used Theorem 3. Replacing y by $f(x)$ on the right hand side yields the proposition. *(QED)*

The *variance* σ_X^2 of a random variable X is defined as usual, i.e.

$$\sigma_X^2 \equiv \left\langle (X - \langle X \rangle)^2 \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2 .$$

Its square root σ_X is called *standard deviation.* Independent of the specific experiment and random variable considered, the variance is a measure of the "spread" of the corresponding probability density. This statement is put in a quantitative form by the Chebychev inequality.

Theorem 6 (Chebychev Inequality) *If X is a random variable with mean $\langle X \rangle$ and variance σ_X^2, then the relation*

$$P(|X - \langle X \rangle| \geqslant \epsilon) \leqslant \frac{\sigma_X^2}{\epsilon^2}$$

holds valid.

It should be clear that in the Chebychev inequality $P(|X - \langle X \rangle| \geqslant \epsilon)$ is just an abbreviation for $P(\{\omega \in \Omega \,|\, |X(\omega) - \langle X \rangle| \geqslant \epsilon\})$.

Proof: Consider the function $\Theta(|X(\omega) - \langle X \rangle| - \epsilon)$ of $\omega \in \Omega$. Evidently, it is non-zero if and only if $|X(\omega) - \langle X \rangle| \geqslant \epsilon$, i.e. if and only if $\sigma_X^2/\epsilon^2 \geqslant 1$. Therefore we can conclude that

$$\forall \omega \in \Omega : \; \Theta(|X(\omega) - \langle X \rangle| - \epsilon) = 1 \cdot \Theta(|X(\omega) - \langle X \rangle| - \epsilon) \leqslant \frac{\sigma_X^2}{\epsilon^2},$$

where the final step follows from the fact that the Heaviside function is less equal 1. But the mean of the Heaviside function $\Theta(g(X))$ is just the probability of the event $\{\omega \in \Omega | g(X(\omega)) \geqslant 0\}$. Thus, as σ_X^2/ϵ^2 is just a number, we get

$$P(|X - \langle X \rangle| \geqslant \epsilon) = \langle \Theta(|X - \langle X \rangle| - \epsilon) \rangle \leqslant \left\langle \frac{\sigma_X^2}{\epsilon^2} \right\rangle = \frac{\sigma_X^2}{\epsilon^2},$$

which is the Chebychev inequality. *(QED)*

Sometimes one is not interested in the overall probability of some event A_1, but rather in the probability of that event given that some other event A_2 occurs. This is known as a conditional probability. For example, a student may wonder not so much whether he'll pass an exam, but may be concerned with the question whether he'll fail *if he doesn't revise the subject properly.* Naively, we suspect that such a conditional probability should fulfill the requirements

(i) that it is a probability,

(ii) that if two events A and B are mutually exclusive (or, as the more mathematically-minded might put it, if $A \cap B = \varnothing$), the conditional probability of A given B is zero, *and*

(iii) that for events A_1, A_2 which are subsets of B, the ratio of the conditional probabilities of A_1 and A_2 given B should equal that of the respective "normal" probabilities.

This motivates the following definition:

Definition 18 *Consider an experiment with sample space* Ω *and a probability* P*. Let* $\mathcal{S}$ *denote the set of all events, and let* $B \in \mathcal{S}$*. Then the function*

$$P(\cdot|B): \ \mathcal{S} \ni A \longrightarrow P(A|B) \in \mathbb{R}$$

is called a conditional probability *if*

(i) it is a probability,

(ii) $A \cap B = \varnothing \ \Rightarrow \ P(A|B) = 0$, and

(iii) $\forall A_1, A_2 \subset B: \ \dfrac{P(A_1|B)}{P(A_2|B)} = \dfrac{P(A_1)}{P(A_2)}$.

Loosely speaking, the conditional probability should be the "normal" probability normalized for the constraining event B. This is made more precise by the following theorem.

Theorem 7 *Let* P *be a probability, and let* B *be an event. Then to* P *and* B *there corresponds a unique conditional probability* $P(\cdot|B)$*, which is given by*

$$\forall A \subset \Omega: \ P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{2.12}$$

Proof: Let us first show that Eq. 2.12 indeed defines a conditional probability: By construction, $P(\cdot|B)$ is a real-valued function on the set of all events. Evidently, $P(A|B) \geqslant 0$ for all events A, and

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

In addition, for a set $\{A_1, A_2, \ldots, A_k, \ldots\}$ of pairwise disjoint events

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)}$$
$$= \sum_{i=1}^{\infty} P(A_i|B),$$

where in the last-but-one step condition (iii) of Def. 12 has been used. Hence all the conditions of Def. 12 are fulfilled, and $P(\cdot|B)$ must be a probability. Of course, if A and B are disjoint, $P(A|B)$ vanishes,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\varnothing)}{P(B)} = 0.$$

Finally, we obtain for two events $A_1, A_2 \subset B$

$$\frac{P(A_1|B)}{P(A_2|B)} = \frac{P(A_1 \cap B)}{P(B)} \cdot \frac{P(B)}{P(A_2 \cap B)} = \frac{P(A_1)}{P(A_2)},$$

so that we have verified all the requirements of Def. 18. Thus $P(\cdot|B)$ as given by Eq. 2.12 is in fact a conditional probability.

It remains to be shown that the conditional probability is uniquely defined by Eq. 2.12. However, if $P(\cdot|B)$ is a conditional probability, we may write

$$P(A|B) = P((A \cap B) \cup (A \backslash (A \cap B))|B) = P(A \cap B) + P(A \backslash (A \cap B)),$$

where we have made use of the fact that $P(\cdot|B)$ is a probability. As B and $A \backslash (A \cap B)$ are disjoint by definition, the second term of the right hand side must vanish, and we get

$$P(A|B) = P(A \cap B|B).$$

In particular, it follows that $1 = P(\Omega|B) = P(\Omega \cap B|B) = P(B|B)$. Using this result together with condition (iii) of Def. 18, we obtain

$$P(A|B) = \frac{P(A \cap B|B)}{1} = \frac{P(A \cap B|B)}{P(B|B)} = \frac{P(A \cap B)}{P(B)},$$

which implies the uniqueness of the conditional probability. *(QED)*

For some random variable X_1 and an event B, we may compute the probability density $p_{X_1}(\cdot|B)$ of X_1 corresponding to the conditional probability $P(\cdot|B)$ by means of the standard definition,

$$p_{X_1}(x_1|B) = \frac{\mathrm{d}}{\mathrm{d}x_1} P(\{\omega \in \Omega | X(\omega) \leqslant x_1\}|B).$$

Here, normally we are interested in the probability density of X_1 given some other random variable X_2 has some specific value. In order to derive an expression for this conditional probability density, let us first compute that for a (small) range of values x_2, i.e. let us consider $p_{X_1}(x_1|B_\Delta)$ with

$$B_\Delta \equiv \{\omega \in \Omega \,|\, x_2 \leqslant X_2(\omega) \leqslant x_2 + \Delta\}.$$

We then obtain

$$\begin{aligned} p_{X_1}(x_1|B_\Delta) &= \frac{\mathrm{d}}{\mathrm{d}x_1}\left(\frac{P(\{\omega \in \Omega|X(\omega) \leqslant x_1\} \cap B_\Delta)}{P(B_\Delta)}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}x_1}\left(\frac{F_{X_1,X_2}(x_1, x_2 + \Delta) - F_{X_1,X_2}(x_1, x_2)}{F_{X_2}(x_2 + \Delta) - F_{X_2}(x_2)}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}x_1}\left(\frac{F_{X_1,X_2}(x_1, x_2 + \Delta) - F_{X_1,X_2}(x_1, x_2)}{\Delta} \cdot \right. \\ &\qquad\qquad \left. \cdot \frac{\Delta}{F_{X_2}(x_2 + \Delta) - F_{X_2}(x_2)}\right). \end{aligned}$$

In the limit of $\Delta \longrightarrow 0$ one thus gets

$$\begin{aligned} \lim_{\Delta \longrightarrow 0} p_{X_1}(x_1|B_\Delta) &= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} F_{X_1,X_2}(x_1, x_2) \cdot \frac{\Delta}{\Delta \cdot p_{X_2}(x_2)}\right) \\ &= p_{X_1,X_2}(x_1, x_2) \cdot \frac{1}{p_{X_2}(x_2)}. \end{aligned}$$

This motivates the following definition:

Definition 19 *Let X_1, X_2 be random variables and let p_{X_1,X_2}, p_{X_2} the probability densities of the random vector (X_1, X_2) and of X_2, respectively. Then the* conditional probability density $p_{X_1}(\cdot|x_2)$ *is defined as*

$$p_{X_1}(x_1|x_2) \equiv \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_2}(x_2)}.$$

As shown above, the conditional probability density may be viewed as the limit

$$p_{X_1}(x_1|x_2) = \lim_{\Delta \longrightarrow 0} p_{X_1}(x_1 \,|\, \{\omega \in \Omega | x_2 \leqslant X_2(\omega) \leqslant x_2 + \Delta\}).$$

Having introduced the concept of a conditional probability density, we may now discuss the statistical dependence of random variables.

Definition 20 *Two random variables X_1, X_2 are called statistically independent, if the conditional probability density $p_{X_1}(x_1|x_2)$ is function of x_1 only.*

Alternatively we may say that two random variables are independent, if their probability density factorizes:

Theorem 8 *Two random variables X_1, X_2 are statistically independent if and only if their probability density p_{X_1,X_2} factorizes, i.e.*

$$p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$$

for all x_1, x_2.

Proof: Let $p_{X_1}(x_1,x_2) \equiv f(x_1)$, i.e. consider two statistically independent random variables. Then

$$p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1|x_2)p_{X_2}(x_2) = f(x_1)p_{X_2}(x_2),$$

and integration of both sides over $\mathbb{R}$ with respect to x_2 yields

$$p_{X_1}(x_1) = \int \mathrm{d}x_2\, p_{X_1,X_2}(x_1,x_2) = \int \mathrm{d}x_2\, f(x_1)p_{X_2}(x_2) = f(x_1),$$

so that

$$p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2).$$

Conversely, if $p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$ we obtain

$$p_{X_1}(x_1|x_2) = \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_2}(x_2)} = \frac{p_{X_1}(x_1)p_{X_2}(x_2)}{p_{X_2}(x_2)} = p_{X_1}(x_1),$$

so that $p_{X_1}(x_1|x_2)$ doesn't depend on x_2. *(QED)*

We should mention an important corollary of Theorem 8:

Theorem 9 *The mean of the product of two independent random variables X_1, X_2 is equal to the product of the means of the random variables,*

$$\langle X_1 X_2 \rangle = \langle X_1 \rangle \langle X_2 \rangle .$$

Proof: The proposition follows directly from the fact that for independent variables the probability density factorizes:

$$\begin{aligned}\langle X_1 X_2 \rangle &= \int \mathrm{d}x_1 \int \mathrm{d}x_2\, p_{X_1,X_2}(x_1,x_2) = \int \mathrm{d}x_1\, p_{X_1}(x_1) \int \mathrm{d}x_2\, p_{X_2}(x_2) \\ &= \langle X_1 \rangle \langle X_2 \rangle\end{aligned}$$

(QED)

The degree of statistical dependence of two random variables may be described by means of their covariance or their correlation coefficient.

Definition 21 *Let X_1, X_2 be random variables. Then the* covariance $\sigma_{X_1X_2}$ *or* $\mathrm{cov}(X_1,X_2)$ *of X_1 and X_2 is given by*

$$\mathrm{cov}(X_1,X_2) \equiv \langle (X_1 - \langle X_1 \rangle)(X_2 - \langle X_2 \rangle) \rangle = \langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle\,,$$

whereas the correlation coefficient $\mathrm{cor}(X_1,X_2)$ *is defined as*

$$\mathrm{cor}(X_1,X_2) \equiv \frac{\mathrm{cov}(X_1,X_2)}{\sigma_{X_1}\sigma_{X_2}},$$

where σ denotes the square root of the variance.

It is left as an exercise for the reader to show that the absolute value of the correlation coefficient is always less equal 1.

2.3.3 Law of large numbers. Central limit theorem

So far, we have been concerned with single measurements for some given experiment. However, it is the lot of physicists that they have to measure the quantities of interest more than once. At first sight, this seems to be a piece of bad news, as within the framework of our theory of probability a measurement (as known as a random variable) is a given function and as such must be fixed. Accordingly, in order to describe a repetition of measurements, we have to invoke more than one experiment. But such an extension is not provided for by the formalism developed in the preceding sections. So, do we have to embark on looking for a more general approach to probability?

Fortunately enough, the answer is no. We are spared the trouble by the observation that *a set of (several) experiments is itself a (single) experiment.*

Indeed, consider N experiments E_k with sample spaces Ω_k. Then all the E_k together may be regarded as an experiment E with the sample space

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_N = \{(\omega_1, \ldots, \omega_N) \,|\, \omega_1 \in \Omega_1 \wedge \ldots \wedge \omega_N \in \Omega_N\}.$$

The corresponding set of all events may be expressed as

$$\mathcal{S} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_N$$

This gives us the possibility to make sense of the concept of repeated measurements on the same experiment: Let $\tilde{X}_k$ be a random variable for the experiment E_k. Then by means of

$$\forall A_i \subset \Omega_i : \; X((A_1, \ldots, A_N)) \equiv \tilde{X}_k(A_k)$$

we can define a corresponding random variable X_k of the experiment E. Note that if some experiments E_i are identical (i.e. if they have the same sample space and probability), the $\tilde{X}_i$ may be the same. On the other hand, for $i \neq j$ the corresponding random variables X_i, X_j are different from each other, and in a moment, we'll see that they are even statistically independent. But, as we haven't introduced this term for more than two dimensions yet, we should first define it:

Definition 22 *N random variables are called* statistically independent*, if their joint probability density factorizes, i.e. if*

$$p_{X_1,\ldots,X_N}(x_1, \ldots, x_N) = \prod_{i=1}^{N} p_{X_i}(x_i).$$

The reader should convince themselves that this is consistent with Def. 20 for the two-dimensional case. Now consider the random variables introduced above. As they belong to completely independent (though possibly identical) experiments, the condition

$$P(\{\omega \in \Omega | X_1 \leqslant x_1\} \cap \ldots \cap \{\omega \in \Omega \,|\, X_N \leqslant x_N\}) = \prod_{i=1}^{N} P(\{\omega \in \Omega \,|\, X_i \leqslant x_i\})$$

should hold valid. But this implies that the distribution and hence the probability density factorizes. In particular, this is true for a repeated measurement, so that we have reached a conclusion of utmost importance: *The* *random variables corresponding to a repeated measurement are statistically independent.* Incidentally, for N independent random variables one obtains (analogously to the case of two dimensions)

$$\begin{aligned}\langle X_1 \dots X_N \rangle &= \int \mathrm{d}x_1 \dots \int \mathrm{d}x_N \, p_{X_1,\dots,X_N}(x_1,\dots,x_N) \\ &= \int \mathrm{d}x_1 \, p_{X_1}(x_1) \cdot \ldots \cdot \int \mathrm{d}x_N = \langle X_1 \rangle \dots \langle X_N \rangle \, ,\end{aligned}$$

and thus *the mean of the product of independent random variables is equal to the product of the means of the random variables.* Now let X_1, …, X_N be independent random variables. Then the function $\bar{X}$ defined as

$$\bar{X}_N \equiv \frac{1}{N} \sum_{i=1}^{N} X_i$$

is itself a random variable, the mean of which evidently has the value

$$\left\langle \bar{X}_N \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \langle X_i \rangle \, .$$

Its variance can be expressed as

$$\sigma^2_{\bar{X}_N} = \left\langle \bar{X}^2_N \right\rangle - \left\langle \bar{X}_N \right\rangle^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle).$$

Assume that the X_i are statistically independent. Then $\langle X_i X_j \rangle = \langle X_i \rangle \langle X_j \rangle$ for $i \neq j$ and thus

$$\sigma^2_{\bar{X}_N} = \frac{1}{N^2} \sum_{i=1}^{N} (\left\langle X_i^2 \right\rangle - \langle X_i \rangle^2) = \frac{1}{N^2} \sum_{i=1}^{N} \sigma^2_{X_i}.$$

If the variance is the same for all random variables, $\sigma^2_{X_i} \equiv \sigma^2_X$, we get

$$\sigma^2_{\bar{X}_N} = \frac{1}{N} \sigma^2_X$$

and from the Chebychev inequality it follows that

$$P(|\bar{x}_N - \left\langle \bar{X}_N \right\rangle| \geq \epsilon) \leq \frac{\sigma^2_{\bar{X}_N}}{\epsilon^2} = \frac{\sigma^2_X}{N \epsilon^2}.$$

Hence we have proved that in the limit of large N the probability of measuring a value of $\bar{X}_N$ outside the 1σ interval centered on $\langle X_N \rangle$ (i.e. outside the interval $[\langle \bar{X}_N \rangle - \sigma^2_{\bar{X}_N}, \langle \bar{X}_N \rangle + \sigma^2_{\bar{X}_N}]$) becomes arbitrarily small. This is the essential content of the law of large numbers.

Theorem 10 *Let $X_1, \ldots, X_N$ be N independent identically distributed random variables, and define*

$$\bar{X}_N \equiv \frac{1}{N} \sum_{i=1}^{N} X_i.$$

Then for any arbitrary (but fixed) $\epsilon > 0$ and $\langle X_i \rangle \equiv \langle X \rangle$

$$\lim_{N \longrightarrow \infty} P(|\bar{x}_N - \langle X \rangle| \leqslant \epsilon) = 0.$$

Obviously the law of large numbers constitutes the rationale of measuring quantities more than once, as, loosely speaking, *if you carry out sufficiently many identical measurements, the mean of your measured value will be very probably very close to the mean of the respective quantity.* Note that we made no assumption concerning the form of the distribution of the random variable under consideration.

This result may seem sufficiently surprising; but we can do even better. To see this, let $X_1, \ldots, X_N$ again be N independent and identically distributed random variables. Then the probability density factorizes (cf. Def. 22),

$$p_{X_1,\ldots,X_N}(x_1, \ldots, x_N) = p_{X_1}(x_1) \ldots p_{X_N}(x_N). \tag{2.13}$$

Now consider the random variable $\bar{X}_N \equiv \sum_{i=1}^{N} X_i$. Eventually, we will be interested in the limit of its probability density $p_{\bar{X}_N}$ for an infinite number of variables N, but for the moment, we make no assumption concerning N. We start from the simple observation that the probability density of a random variable may be regarded as the mean of a delta function (which can readily be deduced from Theorem 4), use Eq. 2.13 and employ the inverse Fourier transform of the delta function:

$$\begin{aligned}
p_{\bar{X}_N}(\bar{x}) &= \left\langle \delta(\bar{x} - \bar{X}_N) \right\rangle \\
&= \int \mathrm{d}x_1 \dots \mathrm{d}x_N \, p_{X_1}(x_1) \dots p_{X_N}(x_N) \delta\left(\bar{x} - \frac{1}{N}\sum_{i=1}^{N} x_i\right) \\
&= \int \frac{\mathrm{d}\lambda}{2\pi} \int \mathrm{d}x_1 \dots \mathrm{d}x_N \, p_{X_1}(x_1) \dots p_{X_N}(x_N) e^{i\lambda\left(\bar{x} - \frac{1}{N}\sum_{i=1}^{N} x_i\right)} \\
&= \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda\bar{x}} \int \mathrm{d}x_1 \, p_{X_1}(x_1) e^{-i\frac{\lambda}{N}x_1} \dots \int \mathrm{d}x_N \, p_{X_N}(x_N) e^{-i\frac{\lambda}{N}x_N}
\end{aligned}$$

Now as the X_i are identically distributed, their probability density must be the same each. We write $p_{X_i} \equiv p$ and, using the characteristic function χ, we get

$$\begin{aligned}
p_{\bar{X}_N}(\bar{x}) &= \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda\bar{x}} \left(\int \mathrm{d}x \, e^{-i\frac{\lambda}{N}x} p(x) \right)^N = \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda\bar{x}} \chi\left(\frac{\lambda}{N}\right)^N \\
&= \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda\bar{x} + N \ln \chi\left(\frac{\lambda}{N}\right)}.
\end{aligned}$$

To proceed further, we have to *assume* that the characteristeric function can be expanded up to second order in a Taylor series about 0, i.e.

$$\chi(k) = \chi(0) + \left.\frac{\mathrm{d}\chi(k)}{\mathrm{d}k}\right|_{k=0} k + \frac{1}{2} \left.\frac{\mathrm{d}^2\chi(k)}{\mathrm{d}k^2}\right|_{k=0} k^2 + O(k^3)$$

and thus (as χ is a generating function and as $\chi(0) = 1$, cf. Def. 17)

$$\chi(k) = 1 - i \left\langle X \right\rangle k - \frac{1}{2} \left\langle X^2 \right\rangle k^2 + O(k^3),$$

where $\langle X^m \rangle$ here and in the following denotes the mean of X_i^m. Hence the statement that χ can be expanded up to second order may also be put in the form that the first two moments of the random variables exist. We will comment on this assumption below, but for the moment, let us take it for granted. Together with the Taylor expansion of $\ln(z)$ about $y = 1$,

$$\ln(y) = (y - 1) - \frac{1}{2}(y - 1)^2 + O((y - 1)^3),$$

we then obtain

$$\begin{aligned}\ln\left(\chi\left(\frac{\lambda}{N}\right)\right) &= \ln\left(1 - i\frac{\lambda}{N}\left\langle X\right\rangle - \frac{1}{2}\frac{\lambda^2}{N^2}\left\langle X^2\right\rangle + O\left(\frac{\lambda^3}{N^3}\right)\right)\\ &= -i\frac{\lambda}{N}\left\langle X\right\rangle - \frac{\left\langle X^2\right\rangle - \left\langle X\right\rangle^2}{2N^2}\lambda^2 + O\left(\frac{\lambda^3}{N^3}\right)\\ &= -i\frac{\lambda}{N}\left\langle X\right\rangle - \frac{\sigma_X^2}{2N^2}\lambda^2 + O\left(\frac{\lambda^3}{N^3}\right)\end{aligned}$$

Hence $p_{\bar{X}_N}$ may be expressed as

$$p_{\bar{X}_N}(\bar{x}) = \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda(\bar{x}-\langle X\rangle)} e^{-\lambda^2\frac{\sigma_X^2}{2N} + O\left(\frac{\lambda^3}{N^3}\right)}.$$

For $N \gg 1$ the term $O\left(\frac{\lambda^3}{N^3}\right)$ can be neglected, so that

$$p_{\bar{X}_N}(\bar{x}) \xrightarrow{N\gg 1} \int \frac{\mathrm{d}\lambda}{2\pi} e^{i\lambda(\bar{x}-\langle X\rangle)} e^{-\lambda^2\frac{\sigma_X^2}{2N}}. \tag{2.14}$$

Now for all positive a and b

$$\begin{aligned}\int_{-\infty}^{+\infty} \mathrm{d}z\, e^{-az^2+ibz} &= \int_{-\infty}^{+\infty} \frac{\mathrm{d}u}{\sqrt{a}} e^{-u^2+\frac{ib}{\sqrt{a}}u} = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} \mathrm{d}u\, e^{-\left(u-\frac{ib}{2\sqrt{a}}\right)^2+\frac{-b^2}{4a}}\\ &= \frac{1}{\sqrt{a}} e^{\frac{-b^2}{4a}} \int_{-\infty}^{+\infty} \mathrm{d}u\, e^{-\left(u-\frac{ib}{2\sqrt{a}}\right)^2} = \frac{1}{\sqrt{a}} e^{\frac{-b^2}{4a}} \int_{-\infty-\frac{ib}{2\sqrt{a}}}^{+\infty-\frac{ib}{2\sqrt{a}}} \mathrm{d}v\, e^{-v^2}\end{aligned}$$

The path of integration of the integral on the right hand side runs parallel to the real axis. However, from the calculus of residues we may conclude that instead we may integrate along the real axis itself, and thus we obtain[1]

$$\int_{-\infty}^{+\infty} \mathrm{d}z\, e^{-az^2+ibz} = \frac{1}{\sqrt{a}} e^{\frac{-b^2}{4a}} \int_{-\infty}^{+\infty} \mathrm{d}v\, e^{-v^2} = \sqrt{\frac{\pi}{a}} e^{\frac{-b^2}{4a}} \tag{2.15}$$

---

[1] We employ the well-known relation $I \equiv \int_{\infty}^{+\infty} e^{-x^2}\mathrm{d}x = \sqrt{\pi}$, which follows directly from

$$I^2 = \int_{-\infty}^{+\infty} \mathrm{d}x \int_{-\infty}^{+\infty} \mathrm{d}y\, e^{-(x^2+y^2)} = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} \mathrm{d}r\, re^{-r^2} = \pi,$$

where r and ϕ are polar coordinates.

Applying Eq. 2.15 to Eq. 2.14 finally yields

$$p_{\bar{X}_N}(\bar{x}) \overset{N\gg 1}{\longrightarrow} \frac{1}{\sqrt{2\pi N^{-1}\sigma_X^2}} \exp\left(-\frac{(\bar{x}-\langle X\rangle)^2}{2N^{-1}\sigma_X^2}\right),$$

Hence we have proved the following remarkable proposition:

Theorem 11 *Let $X_1, \ldots, X_N$ be N independent identically distributed random variables, the first and second moment of which exist. Let $\langle X\rangle$ and σ_X^2 denote the mean and the variance of the X_i, respectively. Then in the limit of large N the probability distribution $p_{\bar{X}_N}$ of the random variable $\bar{X}_N \equiv \frac{1}{N}\sum_{i=1}^N X_i$ approaches a Gaussian distribution with mean $\langle X\rangle$ and variance σ_X^2/N:*

$$p_{\bar{X}_N}(\bar{x}) \overset{N\gg 1}{\longrightarrow} \frac{1}{\sqrt{2\pi N^{-1}\sigma_X^2}} \exp\left(-\frac{(\bar{x}-\langle X\rangle)^2}{2N^{-1}\sigma_X^2}\right). \tag{2.16}$$

Let us add two remarks concerning the applicability of this theorem:

(i) While the condition that the first two moments of the random variable under consideration actually exist may seem slightly technical, one should bear in mind that it is not fulfilled automatically. As an example, consider the innocent-looking Cauchy distribution

$$p(x) \propto \frac{1}{x^2+a^2}.$$

Here, neither the first nor the second moment exists, and Theorem 11 cannot be applied. In particular, the mean of the Cauchy distribution is infinite, while of course the probability of large values of x is small. Hence in this case the mean value is *not* a good measure of what value to expect for the random variable.

(ii) There is an important example of dependent random variables: Usually, in a computer-based random number generator, a random number (corresponding to a random variable) is constructed from previous ones. Hence, quite some care has to be taken. The reader is referred to [8] for a discussion on this topic.

Figure 2.6: Probability densities $p_{\bar{X}_N}(x)$ as given by Eq. 2.16. While all curves are centered at the same value, their width decreases with increasing N. In the limit of infinite N one gets a delta function, $p_{\bar{X}_n}(\bar{x}) = \delta(\bar{x} - \langle X \rangle)$.

The content of Theorem 11 may be stated in various ways, which we will now discuss in turn. (We adopt the notation of this theorem and assume its conditions to be fulfilled.)

- As is illustrated in Fig. 2.6, Eq. 2.16 implies that (in the limit of large N) the probability density of $\bar{X}_N$ is a Gaussian distribution which independent of N is centered at $\langle X \rangle$ and the width of which decreases with increasing N. In the limit of infinite N one obtains a delta function.
- The probability density p_{S_N} of the random variable

  $$S_N \equiv \sum_{i=1}^{N} X_i$$

  can be obtained from that of $\bar{X}_N$ by a simple transformation according to Eq. 2.11. One gets

  $$p_{S_N}(\bar{s}) \stackrel{N \gg 1}{\longrightarrow} \frac{1}{\sqrt{2\pi N \sigma_X^2}} \exp\left(-\frac{(\bar{s} - N\langle X \rangle)^2}{2N\sigma_X^2}\right), \tag{2.17}$$

  so that the probability densities are centered at $N\langle X \rangle$. Their width increases with increasing N. This is shown in Fig. 2.7. However, the relative fluctuations are given by

  $$\left(\frac{\sigma_{S_N}^2}{\langle S_N \rangle^2}\right)^{1/2} = \left(\frac{N\sigma_X^2}{N^2\langle X \rangle^2}\right)^{1/2} = O\left(\frac{1}{\sqrt{N}}\right),$$

  and thus they decrease for increasing N.
- Finally, considering the random variable

  $$Z_N \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - \langle X \rangle) = \sqrt{N}(\bar{X}_N - \langle X \rangle)$$

Figure 2.7: Probability densities p_{S_N} according to Eq. 2.17. If N increases, the width of the curves increases, and their centre moves to larger values.

and applying Eq. 2.11, one obtains

$$p_{Z_N}(z) \xrightarrow{N \gg 1} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{z^2}{2\sigma_X^2}\right).$$

Hence the limiting probability density doesn't depend on N any longer.

At least from a mathematical point of view, of the three random variables $\bar{X}_N$, S_N, and Z_N, Z_N is the most appropriate, as it is the only one which has a well-defined probability density in the limit $N \longrightarrow \infty$. That this probability is (as shown above) of Gaussian form is the proposition of what is known as the central limit theorem.

Theorem 12 (Central Limit Theorem) *Let X_1, ..., X_N be N independent identically distributed random variables, the first two moments of which exist. Furthermore, let $\langle X \rangle$ and σ_X^2 denote the mean and the variance of the X_i, respectively, and define*

$$Z_N \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - \langle X \rangle).$$

Then in the limit of infinite N the probability density p_{Z_N} becomes a Gaussian distribution of width σ_X^2 centered at zero, i.e.

$$\lim_{N \longrightarrow \infty} p_{Z_N}(z) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{z^2}{2\sigma_X^2}\right).$$

When embarking on our discussion of probability theory, we started with mentioning the intuitive notion of probabilities as relative frequencies. Using the results proved in the present section, we can now elaborate a little bit on this concept. To this end, let us consider a discrete sample space, $\Omega = \{k \,|\, k \in \mathbb{N}\}$, where the various k have the probabilities P_k. Formally, we may extend Ω to $\mathbb{R}$ and use

$$P:\ \Omega = \mathbb{R} \supset I \longrightarrow P(I) \equiv \sum_{k \in I \cap \mathbb{N}} P_k$$

as the probability. Then the identity function (i.e. $X: \; x \longrightarrow x$) constitutes a random variable, the distribution of which is given by

$$F_X(x) = P(\{x' \,|\, x' \leqslant x\}) = \sum_{k \leqslant x'} P_k \Theta(x-k),$$

so that the corresponding probability density may be written as

$$p_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \sum_k P_k \delta(x-k).$$

Now consider the function

$$n_l(x) \equiv \left\{ \begin{array}{ll} 1 & (x = l) \\ 0 & (\text{otherwise}) \end{array} \right. .$$

Clearly, $n_l(X)$ denotes the random variable describing the act of obtaining the element l of the sample space. Its mean and variance are given by

$$\langle n_l(X) \rangle = \int \mathrm{d}x \, n_l(x) p_X(x) = \int \mathrm{d}x \, n_l(x) \sum_k P_k \delta(x-k) = \sum_k P_k n_l(k) = P_l$$

and

$$\begin{aligned} \sigma^2_{n_l(X)} &= \left\langle (n_l(X) - \langle n_l(X) \rangle)^2 \right\rangle = \int \mathrm{d}x \, p_X(x) (n_l(x) - P_l)^2 \\ &= \int \mathrm{d}x \left(\sum_k P_k \delta(x-k) \right) (n_l(x) - P_l)^2 = \sum_k P_k (n_l(k) - P_l)^2 \\ &= P_l (1 - P_l)^2, \end{aligned}$$

respectively. Now consider N independent random variables $n_l^{(1)}(X), \ldots, n_l^{(N)}(X)$ distributed according to $n_l(X)$ and define the random variable

$$\bar{n}_{N,l}(X) = \frac{1}{N} \sum_{i=1}^{N} n_l^{(i)}(X),$$

which corresponds to the relative frequency of the occurrence of ω_l. Then from the law of large numbers we can conclude that for all positive ϵ

$$\lim_{N \longrightarrow \infty} P(|\bar{n}_{N,l} - P_l| \geqslant \epsilon) = 0.$$

Thus in the limit of inifinite N, i.e. in the limit of infinitely many repeated measurements, the probability that the relative frequency of an event observed is not very close to the probability of this event gets exceedingly small. So one can indeed make sense of the intuitive notion that probabilities are relative frequencies in infinitely repeated experiments.

Chapter 3

Ensembles of Equilibrium

Having revised the necessary concepts, we can now turn to the discussion of statistical mechanics. Here, we will first concentrate on *systems in equilibrium*, and hence we should ask how such systems can be prepared.

Empirically, one finds the simple prescription illustrated in Fig. 3.1: You start with some arbitrary system and wait until it has become stationary, i.e. until no (macroscopic) changes occur any longer.[1] The system then will have reached a state of equilibrium.

The simplicity of the recipe shouldn't fool the reader. While everyday experience clearly suggsts that any system left alone must eventually reach equilibrium, it proves hard to explain why this should be the case. Within the scope of this chapter, however, we will just take the existence of equilibria for granted.

3.1 Time Development of Macrostates

3.1.1 Liouville Equation

Let us reconsider systems of classical mechanics, as illustrated by Fig. 3.2. In the case of a single microstate the whole dynamics can be described by

---

[1] Actually, in the meantime you might help yourself to a decent cup of coffee, say.

Figure 3.1: How to prepare a system in equilibrium.

Figure 3.2: Dynamics in phase space. *Left:* A microstate $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ moves along some trajectory in phase space. *Right:* The constituent points of a macrostate move along non-intersecting trajectories.

a single trajectory in phase space, which means that assuming the system considered consists of N (point) particles in d dimensions, the phase space density $\rho(\boldsymbol{p}, \boldsymbol{q}, t)$ can be expressed as

$$\rho(\boldsymbol{p}, \boldsymbol{q}, t) = \delta^{d\cdot N}(\boldsymbol{p} - \boldsymbol{p}_\mathrm{t}(\boldsymbol{p}_0, \boldsymbol{q}_0, t))\delta^{d\cdot N}(\boldsymbol{q} - \boldsymbol{q}_\mathrm{t}(\boldsymbol{p}_0, \boldsymbol{q}_0, t)),$$

where $\boldsymbol{p}_0$ ($\boldsymbol{q}_0$) denotes the initial momentum (position) of all the particles at an initial time t_0. Evidently, the functions $\boldsymbol{p}_\mathrm{t}$ and $\boldsymbol{q}_\mathrm{t}$, which describe the motion of the microstate along its trajectory in phase space, must be given by the canonical equations, i.e.

$$\frac{\mathrm{d}q_{\mathrm{t},i}}{\mathrm{d}t} = \frac{\partial H}{\partial p_i} \quad , \quad \frac{\mathrm{d}p_{\mathrm{t},i}}{\mathrm{d}t} = -\frac{\partial H}{\partial q_i}. \tag{3.1}$$

Let us now turn to an arbitrary macrostate and consider the differential volume element $\mathrm{d}^{d\cdot N}p_0\mathrm{d}^{d\cdot N}q_0$ located at $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ at the time t_0. At some later time t this element will have been mapped onto the volume element $\mathrm{d}^{d\cdot N}p\,\mathrm{d}^{d\cdot N}q$, where $\boldsymbol{p}$ and $\boldsymbol{q}$ are given by $\boldsymbol{p} = \boldsymbol{p}_\mathrm{t}(\boldsymbol{p}_0, \boldsymbol{q}_0, t)$ and $\boldsymbol{q} = \boldsymbol{q}_\mathrm{t}(\boldsymbol{p}_0, \boldsymbol{q}_0, t)$, respectively. However, as different trajectories in phase space must remain separate and as the number of systems in an ensemble is a constant, we know that

$$\rho(\boldsymbol{p}_0, \boldsymbol{q}_0, t_0)|\mathrm{d}^{d\cdot N}p_0\mathrm{d}^{d\cdot N}q_0| = \rho(\boldsymbol{p}, \boldsymbol{q}, t)|\mathrm{d}^{d\cdot N}p\,\mathrm{d}^{d\cdot N}q|,$$

and thus we can infer that

$$\rho(\boldsymbol{p}_0, \boldsymbol{q}_0, t_0) = \left|\frac{\partial(\boldsymbol{p}_\mathrm{t}, \boldsymbol{q}_\mathrm{t})}{\partial(\boldsymbol{p}_0, \boldsymbol{q}_0)}\right| \rho(\boldsymbol{p}, \boldsymbol{q}, t). \tag{3.2}$$

On the other hand we may write $\rho(\boldsymbol{p}, \boldsymbol{q}, t)$ as

$$\begin{aligned} \rho(\boldsymbol{p}, \boldsymbol{q}, t) &= \int \rho(\boldsymbol{u}, \boldsymbol{v}, t)\delta^{d\cdot N}(\boldsymbol{p} - \boldsymbol{u})\delta^{d\cdot N}(\boldsymbol{q} - \boldsymbol{v})\mathrm{d}^{d\cdot N}u\,\mathrm{d}^{d\cdot N}v \\ &= \int \rho(\boldsymbol{u}, \boldsymbol{v}, t)\delta^{d\cdot N}(\boldsymbol{p} - \boldsymbol{u}(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}))\delta^{d\cdot N}(\boldsymbol{q} - \boldsymbol{v}(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}))\cdot \\ &\qquad \cdot \left|\frac{\partial(\boldsymbol{u}, \boldsymbol{v})}{\partial(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})}\right| \mathrm{d}^{d\cdot N}\tilde{u}\,\mathrm{d}^{d\cdot N}\tilde{v}. \end{aligned} \tag{3.3}$$

In particular, we might assume that in the last step of Eq. 3.3 the coordinate transformation from $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})$ to $(\boldsymbol{u}, \boldsymbol{v})$ is given by $(\boldsymbol{p}_{\mathrm{t}}, \boldsymbol{q}_{\mathrm{t}})$. Denoting $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ by $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$, respectively, we see from Eqs. 3.3 and 3.2 that

$$\begin{aligned}\rho(\boldsymbol{p}, \boldsymbol{q}, t) &= \int \rho(\boldsymbol{u}, \boldsymbol{v}, t)\delta^{d\cdot N}(\boldsymbol{p} - \boldsymbol{u}(\boldsymbol{p}_0, \boldsymbol{q}_0))\delta^{d\cdot N}(\boldsymbol{q} - \boldsymbol{v}(\boldsymbol{p}_0, \boldsymbol{q}_0))\cdot \\ &\qquad\qquad \cdot \left|\frac{\partial(\boldsymbol{p}_{\mathrm{t}}, \boldsymbol{q}_{\mathrm{t}})}{\partial(\boldsymbol{p}_0, \boldsymbol{q}_0)}\right| \mathrm{d}^{d\cdot N}p_0 \, \mathrm{d}^{d\cdot N}q_0 \\ &= \int \rho(\boldsymbol{p}_0, \boldsymbol{q}_0, t_0)\delta^{d\cdot N}(\boldsymbol{p} - \boldsymbol{p}_{\mathrm{t}}(\boldsymbol{p}_0, \boldsymbol{q}_0, t))\cdot \\ &\qquad\qquad \cdot \delta^{d\cdot N}(\boldsymbol{q} - \boldsymbol{q}_{\mathrm{t}}(\boldsymbol{p}_0, \boldsymbol{q}_0, t))\mathrm{d}^{d\cdot N}p_0 \, \mathrm{d}^{d\cdot N}q_0.\end{aligned} \tag{3.4}$$

Differentiating both sides of this equation with respect to time and keeping in mind the chain rule and the relation

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x' - x) = -\frac{\mathrm{d}}{\mathrm{d}x'}f(x' - x)$$

we get the time development of the macrostate:

$$\begin{aligned}\frac{\partial\rho(\boldsymbol{p}, \boldsymbol{q}, t)}{\partial t} &= \int \mathrm{d}^{d\cdot N}p_0 \, \mathrm{d}^{d\cdot N}q_0 \rho(\boldsymbol{p}_0, \boldsymbol{q}_0, t_0)\cdot \\ &\qquad \cdot \sum_{i=1}^{d\cdot N}\left[\dot{p}_{\mathrm{t},i}(\boldsymbol{p}_0, \boldsymbol{q}_0, t)\left(-\frac{\partial}{\partial p_i}\right) + \dot{q}_{\mathrm{t},i}(\boldsymbol{p}_0, \boldsymbol{q}_0, t)\left(-\frac{\partial}{\partial q_i}\right)\right]\cdot \\ &\qquad\qquad \cdot \delta(\boldsymbol{p} - \boldsymbol{p}_{\mathrm{t}}(\boldsymbol{p}_0, \boldsymbol{q}_0, t))\delta(\boldsymbol{p} - \boldsymbol{p}_{\mathrm{t}}(\boldsymbol{p}_0, \boldsymbol{q}_0, t))\end{aligned}$$

$\dot{p}_{\mathrm{t},i}$ and $\dot{q}_{\mathrm{t},i}$ are given by the canonical equations (Eq. 3.1). In addition, due to the delta functions these functions only contribute if $p_{\mathrm{t},i} = p_i$ and $q_{\mathrm{t},i} = q_i$, so that no dependence on $\boldsymbol{p}_0$ or $\boldsymbol{q}_0$ is left. Hence we may draw these terms in front of the integral, and we obtain

$$\begin{aligned}\frac{\partial\rho(\boldsymbol{p}, \boldsymbol{q}, t)}{\partial t} &= \sum_{i=1}^{d\cdot N}\left[\frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i}\right]\cdot \\ &\qquad \cdot \int \rho(\boldsymbol{p}_0, \boldsymbol{q}_0, t_0)\delta(\boldsymbol{p} - \boldsymbol{p}_{\mathrm{t}}(\boldsymbol{p}_0, \boldsymbol{q}_0, t))\cdot \\ &\qquad\qquad \cdot \delta(\boldsymbol{q} - \boldsymbol{q}_{\mathrm{t}}(\boldsymbol{p}_0, \boldsymbol{q}_0, t))\mathrm{d}^{d\cdot N}p_0 \, \mathrm{d}^{d\cdot N}q_0.\end{aligned}$$

According to Eq. 3.4, the integral on the right hand side is just the phase space density. Therefore, using the Poisson bracket (Eq. 2.1) we arrive at Liouville's theorem.

Theorem 13 (Liouville's Theorem) *The time development of the density* $\rho(\boldsymbol{p}, \boldsymbol{q}, t)$ *in phase space is given by*

$$\frac{\rho(t)}{\partial t} = -\{H, \rho\}. \tag{3.5}$$

We add two brief remarks:

(i) The phase space velocity $\tilde{\boldsymbol{v}} = (\dot{\boldsymbol{p}}, \dot{\boldsymbol{q}})$ has a vanishing divergence, as can be seen from the canonical equations:

$$\nabla\tilde{\boldsymbol{v}} = \sum_{i=1}^{d\cdot N}\left(\frac{\partial\dot{p}_i}{\partial p_i} + \frac{\partial\dot{q}_i}{\partial q_i}\right) = \sum_{i=1}^{d\cdot N}\left(-\frac{\partial^2 H}{\partial p_i\partial q_i} + \frac{\partial^2 H}{\partial q_i\partial p_i}\right) = 0.$$

This result can be used to show that Eq. 3.5 has the form of a continuity equation,

$$\begin{aligned}\frac{\partial\rho}{\partial t} &= -\{H, \rho\} = -\sum_{i=1}^{d\cdot N}\left(\frac{\partial H}{\partial p_i}\frac{\partial\rho}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial\rho}{\partial p_i}\right)\\ &= -\sum_{i=1}^{d\cdot N}\left(\dot{q}_i\frac{\partial\rho}{\partial q_i} + \dot{p}_i\frac{\partial\rho}{\partial p_i}\right) = -\nabla\rho\cdot\tilde{\boldsymbol{v}} = -(\nabla\rho\cdot\tilde{\boldsymbol{v}} + \rho\cdot\nabla\tilde{\boldsymbol{v}})\\ &= -\nabla(\rho\tilde{\boldsymbol{v}}),\end{aligned}$$

which mirrors the fact that the number of ensemble points is a conserved quantity.

(ii) Consider the phase space density $\rho(\boldsymbol{p}(t), \boldsymbol{q}(t), t)$ along some trajectory. From Liouville's theorem and the canonical equations one gets

$$\begin{aligned}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= \frac{\partial\rho}{\partial t} + \sum_{i=1}^{d\cdot N}\left(\frac{\partial\rho}{\partial p_i}\dot{p}_i + \frac{\partial\rho}{\partial q_i}\dot{q}_i\right) = \frac{\partial\rho}{\partial t} + \sum_{i=1}^{d\cdot N}\left(-\frac{\partial\rho}{\partial p_i}\frac{\partial H}{\partial q_i} + \frac{\partial\rho}{\partial q_i}\frac{\partial H}{\partial p_i}\right)\\ &= \frac{\partial\rho}{\partial t} + \{H, \rho\} = 0.\end{aligned}$$

Hence for an ensemble point the phase space density is a constant.

3.1.2 von Neumann Equation

After the somewhat technical derivation of Liouville's theorem in the previous section, the reader might feel glad that, amazingly, the derivation of the corresponding equation in quantum mechanics is somewhat easier.

Indeed, consider an arbitrary quantum mechanical macrostate, i.e. a density matrix $\hat{\rho}$, in a Hilbert space with a countable basis. From the discussion in Sect. 2.2 we know that $\hat{\rho}$ must be Hermitean, so that there is a basis $\{|\alpha_i\rangle\}$ in which it is of diagonal form,

$$\hat{\rho} = \sum_{i=1}^{\infty} \tilde{w}_i \, |\alpha_i\rangle \, \langle\alpha_i| \, ,$$

where the $\tilde{w}_i$ are real constants. Of course, we can express the vectors $|\alpha_i\rangle$ as linear combinations of the eigenvectors $|E_i\rangle$ of the Hamiltonian $\hat{H}$. Hence we may write $\hat{\rho}$ as

$$\hat{\rho} = \sum_{i,j=1}^{\infty} w_{ij} \, |E_i\rangle \, \langle E_j| \, ,$$

where again the w_{ij} denote some constants. The time evolution of the density matrix is obtained by differentiating both sides of this equation with respect to time:

$$\frac{\partial\hat{\rho}}{\partial t} = \sum_{i,j=1}^{\infty} \Big(w_{ij} |\dot{E}_i\rangle \, \langle E_j| + |E_i\rangle \, \langle \dot{E}_j| \Big) \, .$$

Now $|\dot{E}_k\rangle$ and $\langle\dot{E}_k|$ are given by the Schrödinger equation and its complex conjugate:

$$i\hbar|\dot{E}_k\rangle = i\hbar\frac{\partial}{\partial t} \, |E_k\rangle = \hat{H} \, |E_k\rangle \tag{3.6}$$

$$-i\hbar\langle\dot{E}_k| = -i\hbar\frac{\partial}{\partial t} \, \langle E_k| = \langle E_k| \, \hat{H} \tag{3.7}$$

Thus we obtain

$$i\hbar\frac{\partial\hat{\rho}}{\partial t} = \sum_{i,j} w_{ij} \left(\hat{H} \, |E_i\rangle \, \langle E_j| - |E_i\rangle \, \langle E_j| \, \hat{H} \right) = \hat{H}\hat{\rho} - \hat{\rho}\hat{H},$$

and hence we have proved what is known as the von Neumann equation.

Theorem 14 (von Neumann Equation) *The time development of a density matrix $\hat{\rho}$ is given by*

$$i\hbar\frac{\partial\hat{\rho}}{\partial t} = \left[\hat{H}, \hat{\rho}\right],$$

where $\hat{H}$ denotes the Hamiltonian.

But we may even go a step further: As the $|E_i\rangle$ are eigenvectors of the Hamiltonian, the relations

$$\begin{aligned}\hat{H}\,|E_k\rangle &= E_i\,|E_k\rangle \\ \langle E_k|\,\hat{H} &= \langle E_k|\,E_k\end{aligned}$$

hold valid, and the von Neumann equation becomes

$$i\hbar\frac{\partial\hat{\rho}}{\partial t} = \sum_{i,j=1}^{\infty} w_{ij}(E_i - E_j)\,|E_i\rangle\,\langle E_j|\,.$$

If the macrostate under consideration is stationary, the left hand side of this equation vanishes. This implies that for all $E_i \neq E_j$ the statistical weights w_{ij} must be zero as well, and we get the following nice result:

Theorem 15 *The density matrix of a stationary macrostate has a diagonal representation in terms of the eigenstates of the Hamiltonian.*

In the next section we will see that this theorem is of considerable importance.

3.2 Microcanonical ensemble

In principle, all statistical mechanics is included in Liouville's theorem and the von Neumann equation, as they can be used to compute the time development of any given system. Alas, as noted already in the introductory paragraph of this chapter, in practice such a calculation is far from possible. We therefore need an additional postulate, which, while not being able to prove it, we will at least make plausible.

To this end, we start by noting that it was shown at the end of Sect. 3.1.1 that the phase space density ρ is constant along trajectories in phase space. Hence if we assume that the macrostate under consideration is stationary, at least along trajectories we might express ρ in terms of constants of motion.

Figure 3.3: Particles moving in opposite directions along a line. While it is true that the phase space density along the trajectories in phase space can be written in terms of a conserved quantity (namely the momentum), its functional form depends on the trajectory chosen.

However, in general the functional form may be different for different trajectories. As an example consider the simple case of a stream of non-interacting particles moving with some fixed speed in opposite directions along a line (Fig. 3.3). Here along the particle path, the phase space density is a sum of two delta functions of the momentum, whereas everywhere else it vanishes identically. But now let us concentrate on equlibrium systems. These are undoubtedly stationary, and thus along some given trajectory their phase space density will be some function of conserved quantities. But there is more to equilibria than the mere absence of time dependence, which after all applies to the fluid depicted in Fig. 3.3 as well. Loosely speaking, we might say that the main property of an equilibrium state must be something like "having as few distinct features as possible." We therefore make the bold *assumption* that *for equilibrium states the functional form of the phase space density is the same everywhere.*

At first sight, this doesn't help much, as there are infinitely many constants of motion. However, we may narrow down the number of possible quantities appearing in the phase space density by splitting the system under consideration into two subsystems with phase space densities ρ_1 and ρ_2. Then the phase density ρ of the total system must equal the product of ρ_1 and ρ_2, so that

$$\ln \rho = \ln \rho_1 + \ln \rho_2,$$

which implies that $\ln \rho$ is an additive quantity and as such should depend on a linear combination of additive constants of motion only. However, in mechanics there are just seven independent additive constants of motion – energy, the three components of momentum and the three components of angular momentum.

Although this has considerably improved the situation, we haven't finished yet. But before going on we need to clarify what kind of systems we are about to discuss. Indeed, to keep things straightforward, we shall assume for the moment that the system(s) under consideration are finite and closed,

Figure 3.4: Microcanonical ensemble. A system is called a microcanonical ensemble if it is in an equilibrium state and if its energy E, volume V and particle number N are fixed.

which means that there is no change in energy, volume or particle number. Such systems are known as microcanonical ensembles.

Definition 23 *An equilibrium system with fixed energy, volume, and particle number is called a* microcanonical ensemble.

The notion of a microcanonical ensemble is illustrated in Fig. 3.4. Now we know from the Noether theorem that the conservation of energy arises from time invariance, that of momentum from translational invariance and that of angular momentum from rotational invariance. Thinking of a microcanonical system as a box with well-insulated walls, we note that it is invariant under neither translations in space nor rotations. Hence we may suspect that neither momentum nor angular momentum constitutes a constant of motion. Concerning the phase space density, this leaves us just with a dependence on energy.

In the argument put forward, the Liouville theorem was invoked only for showing that the phase space density is a constant of motion. However, due to the von Neumann equation, the same can be said of the statistical weights of the density matrix. As the rest of the argument remains valid in quantum mechanics as well, we see that the conclusion reached is true independent of whether the system under consideration is classical or quantum mechanical.

It should be stressed that the previous was not intended as a formal proof but as a handwaving argument which has led to the following hypothesis:

Hypothesis 1 *The phase space density of a classical microcanonical ensemble depends on the total energy only. The same is true for the statistical weights of density matrix corresponding to a microcanonical ensemble in quantum mechanics.*

Now there may be classical systems such as spin configurations treated in a classical way (see below) or dipoles in an electric field which are not really covered by the phase space formalism advocated so far. However, the

Figure 3.5: States with energies in the range $[E - \Delta, E]$.

quantum mechanical content of the preceding hypothesis suggests a generalization, which may be regarded as the fundamental hypothesis of statistical physics:

Hypothesis 2 *In a microcanonical ensemble the probability of some microstate depends on the total energy only.*

We put a nice box around this hypothesis in order to emphasize its importance, as basically the whole of statistical physics follows from this innocent-looking result. Concerning the phase space density ρ_{mc} of a classical microcanonical ensemble we obtain from it the relation

$$\rho_{\text{mc}}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{\omega(E)} \delta(E - H(\boldsymbol{p}, \boldsymbol{q})), \tag{3.8}$$

where E denotes the energy, H the Hamiltonian, and where $\omega(E)$ constitutes a normalization factor to be discussed below in more detail.

So far we have taken the energy to be fixed precisely, which is of course a somewhat dubious assumption, as in real life there are always limited precision and unavoidable disturbances. We shall therefore allow the energy to lie in a small range $[E - \Delta, E]$, as is illustrated in Fig. 3.5. Then the correspunding probability density $\rho_{\text{mc}}^{\Delta}$ will be given by[2]

$$\rho_{\text{mc}}^{\Delta}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{\omega_{\Delta}} \left(\Theta(E - H(\boldsymbol{p}, \boldsymbol{q})) - \Theta(E - \Delta - H(\boldsymbol{p}, \boldsymbol{q})) \right), \tag{3.9}$$

where for small Δ we can approximate the normalization factor ω_{Δ} by $\omega_{\Delta} \approx \Delta \cdot \omega(E) \cdot$. In addition, we may expand $\Theta(E - \Delta - H(\boldsymbol{p}, \boldsymbol{q}))$ to first order, i.e.

$$\Theta(E - \Delta - H(\boldsymbol{p}, \boldsymbol{q})) \approx \Theta(E - H(\boldsymbol{p}, \boldsymbol{q})) - \Delta \cdot \delta(E - H(\boldsymbol{p}, \boldsymbol{q})).$$

---

[2] Strictly speaking, Hypothesis 2 doesn't cover this case, as it applies to states of constant energy only. However, assuming that $\omega(E)$ is a continuous function and that Δ is sufficiently small, we may assume that the normalization factor is a constant.

Accordingly, in the limit of a vanishing Δ we recover the microcanonical phase space density,

$$\lim_{\Delta \longrightarrow 0} \rho_{\mathrm{mc}}^{\Delta}(\boldsymbol{p}, \boldsymbol{q}) = \lim_{\Delta \longrightarrow 0} \frac{\Delta \cdot \delta(E - H(\boldsymbol{p}, \boldsymbol{q}))}{\Delta \cdot \omega(E)} = \rho_{\mathrm{mc}}(\boldsymbol{p}, \boldsymbol{q}).$$

This justifies the use of $\rho_{\mathrm{mc}}^{\Delta}$. Actually, at this stage let us remark that we will see later that almost all of the phase space volume with energies E' less than some given value E is located in an (almost) arbitrarily thin shell at the hypersurface corresponding to the energy E.

Looking at the formal definition given in Sect. 2.3.2 one notes that positions and momenta are not random variables as they have a dimension instead of being just real numbers. Hence strictly speaking the phase space density ρ_{mc} is not a probability density. However, if one rescales ρ_{mc} (and thus ω) by means of

$$\begin{aligned} \rho_{\mathrm{mc}} &\longrightarrow (2\pi\hbar)^{d \cdot N} \rho_{\mathrm{mc}} \\ \omega &\longrightarrow \frac{1}{(2\pi\hbar)^{d \cdot N}} \omega \end{aligned} \tag{3.10}$$

the outcome will be dimensionless and qualifies as a probability density. The reason for choosing $(2\pi\hbar)^{d \cdot N}$ as the rescaling constant will become clear later; for the moment, suffice it to say that it is motivated by quantum mechanics. In the following, ρ_{mc} and ω shall denote these rescaled density and normalization factor, respectively. Of course, the differential volume element of phase space needs to be rescaled accordingly,

$$\mathrm{d}^{d \cdot N} p \, \mathrm{d}^{d \cdot N} q \longrightarrow \frac{\mathrm{d}^{d \cdot N} p \, \mathrm{d}^{d \cdot N} q}{(2\pi\hbar)^{d \cdot N}}.$$

From a classical point of view, different particles are always distinguishable, even though they may belong to the same species. Hence classical physics asserts that we might for example label two electrons “red” and “green”, and interchanging the red and green electron would matter.

But classical mechanics is wrong. In fact, it is shown in quantum mechanics that particles belonging to the same species are completely indistinguishable. Hence given some arbitrary state, if you can reach another state simply by pairwise interchange of particle positions and momenta, the new state is identical to the original one. This is illustrated in Fig. 3.6.

To be more precise, we should have said that the wave function of a set of particles of the same kind is either completely symmetric or completely

Figure 3.6: Indistinguishable particles. If the particles under consideration are indistinguishable, the two states shown (obtained from each other by exchanging two particles) are the same.

antisymmetric. Indeed, the former is true for bosons, whereas the latter holds valid for fermions. We will discuss this in detail in a later chapter.

However, the property of being indistinguishable must carry over into the realm of classical physics, so that we must take phase space points differing by a particle interchange to be the same. Hence, as there are $N!$ ways of arranging N particles, the phase space volume must be smaller by a factor of $1/N!$ than expected naively.

Using Eq. 3.8 (and bearing in mind that now we deal with the corresponding rescaled quantities), we get for the normalization condition of the the probability density that

$$1 \overset{!}{=} \frac{1}{N!}\int \frac{\mathrm{d}^{d\cdot N}p\,\mathrm{d}^{d\cdot N}q}{(2\pi\hbar)^{d\cdot N}}\rho_{\mathrm{mc}}(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{\omega(E)}\int \frac{1}{N!}\frac{\mathrm{d}^{d\cdot N}p\,\mathrm{d}^{d\cdot N}q}{(2\pi\hbar)^{d\cdot N}}\delta(E - H(\boldsymbol{p},\boldsymbol{q})). \tag{3.11}$$

From now on, we shall employ an abbreviating notation for the differential element involved in the integratetion over phase space.

Definition 24 *The differential element* $\mathrm{d}\Gamma$ *is defined as*

$$\mathrm{d}\Gamma \equiv \frac{1}{N!}\frac{\mathrm{d}^{d\cdot N}p\mathrm{d}^{d\cdot N}q}{(2\pi\hbar)^{d\cdot N}}.$$

So finally, we obtain

$$\omega(E) = \int \delta(E - H(\boldsymbol{p},\boldsymbol{q}))\mathrm{d}\Gamma. \tag{3.12}$$

This quantity is called *microcanonical partition function.*

3.3 Microcanonical density matrix and level density

Consider a discrete stationary quantum mechanical system. As we have seen in Sect. 3.1.2, the corresponding density matrix $\hat{\rho}$ may be expressed as

$$\hat{\rho} = \sum_i w_i \left|E_i\right\rangle \left\langle E_i\right| ,$$

where here and in the following i runs over all elements of a basis of orthonormal energy eigenvectors $|E_i\rangle$, which we will denote by B. If the system under consideration happens to be a microcanonical ensemble, we know from the previous section that the statistical weights depend on the energy only, so that for a given energy E we get

$$\hat{\rho}_{\mathrm{mc}} = \sum_{E_i=E} \frac{1}{\omega(E)} \left|E_i\right\rangle \left\langle E_i\right| .$$

The normalization factor $\omega(E)$ can be obtained from the normalization condition $1 = \operatorname{tr}\hat{\rho} = \sum_i w_i$:

$$\omega(E) = \sum_{E_i=E} 1 = \text{number of states } |E_i\rangle \text{ in } B \text{ with } E_i = E.$$

Again, $w(E)$ is known as the partition function. Contrary to the case of classical mechanics it obviously is dimensionless, and no rescaling is necessary. But this time another technical difficulty arises: As the energy spectrum of the system under consideration normally is discrete, the partition function will be different from zero for a discrete set of energies only.

We solve this problem by allowing the energy to lie in a range of width Δ, where Δ is assumed to be small. Then the density matrix becomes[3]

$$\hat{\rho}_{\mathrm{mc}} = \sum_{E-\Delta\leqslant E_i\leqslant E} \frac{1}{\omega_{\Delta}(E)} \left|E_i\right\rangle \left\langle E_i\right| ,$$

and the normalization condition yields

$$\omega_{\Delta}(E) = \sum_{E-\Delta\leqslant E_i\leqslant E} 1 = \text{number of states } |E_i\rangle \text{ in } B \text{ with } E_i \in [E-\Delta, E].$$

---

[3] The content of footnote 2 on page 58 is valid here as well.

Figure 3.7: $\mathcal{N}(E)$ and $g(E)$. *Left:* Example of $\mathcal{N}(E)$ for a discrete energy spectrum. *Right:* Corresponding level density $g(E)$.

Now let us denote with $\mathcal{N}(E)$ the number of states in B with energies less equal E. Then the *level density* $g(E)$ is defined as

$$g(E) \equiv \frac{\mathrm{d}\mathcal{N}(E)}{\mathrm{d}E}.$$

As illustrated by Fig. 3.7, for a discrete system $g(E)$ must be a sum of delta functions. But in everyday life the spacing betweeen neighboring energy levels is extremely small, so that we may replace $g(E)$ by a continuous function $\bar{g}(E)$. Then we may express ω_Δ in terms of $\bar{g}$ by means of

$$\omega_\Delta = \mathcal{N}(E) - \mathcal{N}(E - \Delta) = \int_{E-\Delta}^{E} g(E')\mathrm{d}E' \approx \bar{g}(E) \cdot \Delta.$$

We now have all the prerequisites necessary for discussing the partition function of the ideal gas.

3.4 Example: Counting states of the ideal gas

As we have seen in the previous section, in order to figure out the value of the partition function of a quantum mechanical system, one basically has to count energy eigenstates. This doesn't mean, however, that one has to count *all* eigenstates, a caveat which gets relevant if one considers degenerate energies: If there are two or more linear independent energy eigenvectors, there must be an uncountable number of eigenstates, but nonetheless when "counting the states" we just count the states that belong to a given basis.

This may seem a somewhat biased approach, as after all we might have chosen a different basis and thus would have counted different states. True indeed. But of course the result doesn't depend on our particular choice of basis vectors, as we are interested in the number of vectors only.

Concerning the classical case, there is no such problem, as here rather than counting a discrete set of eigenstates we integrate over the continuous phase space. Still, we shall call this procedure counting of the states as well.

We will now procede to carrying out the appropriate counting of the states of an ideal gas for both the classical and the quantum mechanical case. Before doing this, however, we need to consider the volume and surface area of high-dimensional spheres.

3.4.1 Of high-dimensional apples

From the area of a circle and the volume of the sphere it is straightforward to guess that the volume of a sphere in N dimensions with radius R should be proportional to R^N. The necessary prefactors, on the other hand, elude simple guesswork. They are given by the following theorem.

Theorem 16 *Let $N \geqslant 2$. Then the volume $V_N(R)$ and surface area $A_N(R)$ of an N-dimensional sphere with radius R are given by*

$$V_N(R) = \frac{\pi^{N/2}}{\frac{N}{2}\Gamma\left(\frac{N}{2}\right)} R^N \tag{3.13}$$

$$A_N(R) = \frac{2\pi^{N/2}}{\Gamma\left(\frac{N}{2}\right)} R^{N-1} \tag{3.14}$$

Proof: We may assume without loss of generality that the center of the sphere S is located at the origin of our coordinate system. Then S is just the set

$$S = \{\boldsymbol{x} \in \mathbb{R}^N \mid |\boldsymbol{x}| \leqslant R\},$$

and hence it makes sense to introduce spherical coordinates $r = |\boldsymbol{x}|$, ϕ_1, ..., ϕ_{N-1}, which are related to cartesian coordinates x_1, ..., x_N via the coordinate transformations

$$\begin{aligned} x_1 &= r \cdot f_1(\phi_1, \ldots, \phi_{N-1}) \\ &\vdots \\ x_N &= r \cdot f_N(\phi_1, \ldots, \phi_{N-1}) \end{aligned}$$

In principle we might give explicit formulae for the functions f_i, but as it turns out this is both (the bad news) tedious and (the good news) superfluous. But even without knowing the exact form of the coordinate transformation we can state that

$$\begin{aligned} \det \frac{\partial(x_1, \ldots, x_N)}{\partial(r, \phi_1, \ldots, \phi_{N-1})} &= \begin{vmatrix} f_1 & r\partial_{\phi_1} f_1 & \ldots & r\partial_{\phi_{N-1}} f_1 \\ f_2 & r\partial_{\phi_1} f_2 & \ldots & r\partial_{\phi_{N-1}} f_2 \\ \cdots & \cdots & \cdots & \cdots \\ f_N & r\partial_{\phi_1} f_N & \ldots & r\partial_{\phi_{N-1}} f_N \end{vmatrix} \\ &= r^{N-1} g(\phi_1, \ldots, \phi_{N-1}). \end{aligned}$$

where the (unknown) function g does not depend on the radius r. Therefore, if we use the notation

$$\mathrm{d}\Omega \equiv g(\phi_1, \ldots, \phi_{N-1})\mathrm{d}\phi_1 \ldots \mathrm{d}\phi_{N-1},$$

we may write the volume integral of the sphere as

$$\begin{aligned} V_N(R) &= \int_{|\boldsymbol{x}| \leqslant R} \mathrm{d}x_1 \ldots \mathrm{d}x_N \\ &= \int_0^R \mathrm{d}r \int \mathrm{d}\phi_1 \ldots \mathrm{d}\phi_{N-1} r^{N-1} g(\phi_1, \ldots, \phi_{N-1}) \\ &= \int \mathrm{d}\Omega \cdot \int_0^R \mathrm{d}r\, r^{N-1}. \end{aligned} \tag{3.15}$$

We won't try to compute $\int \mathrm{d}\Omega$ by brute force. Instead, we investigate a seemingly completely unrelated integral, namely $\int_{\mathbb{R}^N} \mathrm{d}^N x\, e^{-|\boldsymbol{x}|^2}$. In spherical coordinates this becomes

$$\int_{\mathbb{R}^N} \mathrm{d}^N x\, e^{-|\boldsymbol{x}|^2} = \int_0^\infty \mathrm{d}r\, r^{N-1} \int \mathrm{d}\Omega\, e^{-r^2} = \int_0^\infty \mathrm{d}r\, r^{N-1} e^{-r^2} \cdot \int \mathrm{d}\Omega,$$

where the radial part can be computed by means of the transformation $u = r^2$ and the definition

$$\Gamma(s) \equiv \int_0^\infty x^{s-1} e^{-x} \mathrm{d}x$$

of the Gamma function:

$$\int_0^\infty \mathrm{d}r\, r^{N-1} e^{-r^2} = \int_0^\infty \frac{\mathrm{d}u}{2\sqrt{u}} u^{\frac{N-1}{2}} e^{-u} = \frac{1}{2}\Gamma\left(\frac{N}{2}\right)$$

Hence we get

$$\int_{\mathbb{R}^N} \mathrm{d}^N x\, e^{-|\boldsymbol{x}|^2} = \frac{1}{2}\Gamma\left(\frac{N}{2}\right) \cdot \int \mathrm{d}\Omega \tag{3.16}$$

On the other hand, we may compute $\int_{\mathbb{R}^N} \mathrm{d}^N x\, e^{-|\boldsymbol{x}|^2}$ in cartesian coordinates (cf. Footnote 1 on page 43):

$$\begin{aligned} \int_{\mathbb{R}^N} \mathrm{d}^N x\, e^{-|\boldsymbol{x}|^2} &= \int \mathrm{d}x_1\, e^{-x_1^2} \int \mathrm{d}x_2\, e^{-x_2^2} \ldots \int \mathrm{d}x_N\, e^{-x_N^2} \\ &= \left(\int \mathrm{d}x\, e^{-x^2}\right)^N = \sqrt{\pi}^N. \end{aligned} \tag{3.17}$$

Taking together Eqs. 3.16 and 3.17 we find that

$$\int \mathrm{d}\Omega = \frac{2\pi^{N/2}}{\Gamma\left(\frac{N}{2}\right)}.$$

Inserting this result into Eq. 3.15 finally yields the volume of the sphere,

$$V_N(R) = \frac{2\pi^{N/2}}{\Gamma\left(N2\right)} \cdot \int_0^R \mathrm{d}r\, r^{N-1} = \frac{\pi^{N/2}}{\frac{N}{2}\Gamma\left(\frac{N}{2}\right)} R^N,$$

and we get the corresponding surface area by a mere differentiation,

$$A_N(R) = \frac{\mathrm{d}V_N(R)}{dR} = \frac{2\pi^{N/2}}{\Gamma\left(\frac{N}{2}\right)} R^{N-1}.$$

Thus V_N and A_N are indeed given by Eqs. 3.13 and 3.14. *(QED)*

Now consider the volume $V_{\Delta,N}(R)$ of a spherical shell comprising the radii from $R - \Delta$ to R. Clearly

$$V_{\Delta,N}(R) = V_N(R) - V_N(R-\Delta) = \frac{\pi^{N/2}}{\frac{N}{2}\Gamma\left(\frac{N}{2}\right)}(R^N - (R-\Delta)^N),$$

and therefore the ratio $V_{\Delta,N}/V_N$ has the value

$$\frac{V_{\Delta,N}(R)}{V_N(R)} = \frac{R^N - (R-\Delta)^N}{R^N} = 1 - \left(1 - \frac{\Delta}{R}\right)^N.$$

But $1 - \Delta/R$ is less than 1, so that

$$\lim_{N \longrightarrow \infty} \frac{V_{\Delta,N}(R)}{V_N(R)} = 1.$$

Thus for high-dimensional spheres almost all of the volume is to be found near the surface. Concerning statistical physics, this result implies that the overwhelming majority of states with energies less than some energy E are located very close to the hypersurface of constant energy E. Concerning fruits, it gives rise to a useful rule of thumb, with which we conclude this section: *Never peel a high dimensional apple!*[4]

---

[4] The reason being, of course, that almost nothing of the apple would remain. – Rumor has it that, by sheer coincidence, in the case of three-dimensional apples the same is true for both vitamins and pesticides alike. We leave it as an exercise for the reader to figure out whether or not to peel such apples.

3.4.2 Classical case

The Hamiltonian of a classical ideal, non-interacting gas consisting of N particles is simply given by

$$H(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m},$$

where the mass m is assumed to be the same for all particles. Accordingly, from Eq. 3.12 and Def. 24 we obtain the relation

$$\omega(E) = \int \mathrm{d}\Gamma\, \delta(E - H(\boldsymbol{p},\boldsymbol{q})) = \frac{1}{N!}\int \frac{\mathrm{d}^{3N}p\, \mathrm{d}^{3N}q}{(2\pi\hbar)^{3N}}\, \delta\left(E - \sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m}\right)$$

for the microcanonical partition function. As the integrand is independent of the particle positions, the integration over all position coordinates yields just the N-th power of the volume V of the container in which the gas is contained, i.e.

$$\omega(E) = \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}} \cdot \int \mathrm{d}^{3N}p\, \delta\left(E - \sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m}\right).$$

The integration over the momenta may be simplified by introducing spherical coordinates. Denoting $|\boldsymbol{p}|$ by p, we the get

$$\begin{aligned}\omega(E) &= \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\int \mathrm{d}\Omega \int_0^\infty \mathrm{d}p\, p^{3N-1}\delta\left(E - \frac{p^2}{2m}\right)\\ &= \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\int \mathrm{d}\Omega \int_0^\infty \mathrm{d}p\, p^{3N-1}\frac{m}{\sqrt{2mE}}\cdot\\ &\qquad\qquad \cdot\left\{\delta(\sqrt{2mE} - p) + \delta(-\sqrt{2mE} - p)\right\},\end{aligned}$$

where in the last step we employed the well-known transformation rule

$$\delta(f(x)) = \sum_{f(x_i)=0}\frac{1}{|f'(x)|}\delta(x_i - x)$$

for delta functions. Evidently, $\delta(-\sqrt{2mE}-p)$ vanishes for all non-negative p, so that together with the notation $R \equiv \sqrt{2mE}$ and the fact that the delta function constitutes the derivative of the Heaviside function, it follows that

$$\begin{aligned}
\omega(E) &= \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\frac{m}{\sqrt{2mE}}\int \mathrm{d}\Omega \int_0^\infty \mathrm{d}p\, p^{3N-1}\delta(R-p) \\
&= \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\frac{m}{\sqrt{2mE}}\frac{\mathrm{d}}{\mathrm{d}R}\int \mathrm{d}\Omega \int_0^\infty \mathrm{d}p\, p^{3N-1}\Theta(R-p) \\
&= \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\frac{m}{\sqrt{2mE}}\frac{\mathrm{d}V_{3N}(R)}{\mathrm{d}R} = \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\frac{m}{\sqrt{2mE}}A_{3N}(R) \\
&= \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\frac{m}{\sqrt{2mE}}A_{3N}(\sqrt{2mE}).
\end{aligned}$$

Here $V_{3N}(R)$ and $A_{3N}(R)$ denote the volume and the surface area of the $3N$-dimensional sphere of radius R, respectively. Hence from Eq. 3.14 we can see that

$$\begin{aligned}
\omega(E) &= \frac{1}{N!}\frac{V^N}{(2\pi\hbar)^{3N}}\frac{m}{\sqrt{2mE}}\frac{2\pi^{3N/2}}{\Gamma\left(\frac{3N}{2}\right)}\sqrt{2mE}^{3N-1} \\
&= \frac{1}{N!\Gamma\left(\frac{3N}{2}\right)}\frac{1}{E}\left[V\left(\frac{mE}{2\pi\hbar^2}\right)^{\frac{3}{2}}\right]^N .
\end{aligned}$$

When considering a small energy interval $[E-\Delta, E]$ rather than a completely fixed energy E, the partion function becomes

$$\begin{aligned}
\omega_\Delta(E) &= \frac{1}{N!}\int \frac{\mathrm{d}^{3N}p\,\mathrm{d}^{3N}q}{(2\pi\hbar^2)^{3N}}\left[\Theta\left(E-\sum_{i=1}^{3N}p_i^2\right)-\Theta\left(\sum_{i=1}^{3N}p_i^2-(E-\Delta)\right)\right] \\
&= \frac{1}{N!}\frac{V^N}{(2\pi\hbar^2)^{3N}}\int \mathrm{d}^{3N}p\Theta(\sqrt{2mE}-|\boldsymbol{p}|) \\
&\qquad - \frac{1}{N!}\frac{V^N}{(2\pi\hbar^2)^{3N}}\int \mathrm{d}^{3N}p\Theta(|\boldsymbol{p}|-\sqrt{2m(E-\Delta)}) \\
&= V_{3N}(\sqrt{2mE}) - V_{3N}(\sqrt{2m(E-\Delta)}) \\
&= \frac{1}{N!\cdot\frac{N}{2}\Gamma\left(\frac{3N}{2}\right)}\left[V\left(\frac{1}{2\pi\hbar^2}\right)^{\frac{3}{2}}\right]^N\left\{(2mE)^{3N/2}-(2m[E-\Delta])^{3N/2}\right\},
\end{aligned} \tag{3.18}$$

where in the last step Eq. 3.13 was used.

Figure 3.8: Momentum quantum numbers of a single particle. The quantum numbers form a rectangular lattice in momentum space with $2\pi\hbar/a_1$, $2\pi\hbar/a_2$, and $2\pi\hbar/a_3$ as its lattice constants.

3.5 Quantum mechanical case

Let us now consider the quantum mechanical analogue of the ideal gas discussed in the previous section, with the additional simplification that we take the constituent particles to be distinguishable. Then the wave function ψ_i of an energy eigenstate for the i-th particle satisfies the Schrödinger equation

$$-\frac{\hbar^2}{2m}\Delta\psi_i(\boldsymbol{x}) = E_i\psi_i(\boldsymbol{x}),$$

where $\boldsymbol{x}$ and E_i denote the position in ordinary three-dimensional space and the particle energy, respectively. We assume that the gas is in a rectangular box with side lengths a_1, a_2, and a_3, and that we have periodic boundary conditions, so that

$$\psi(x_1, x_2, x_3) = \psi_i(x_1 + a_1, x_2, x_3) = \psi_i(x_1, x_2 + a_2, x_3) = \psi_i(x_1, x_2, x_3 + a_3).$$

Then, as the reader should check for themselves, the ψ_i are given by

$$\psi_i(\boldsymbol{x}) = \frac{1}{\sqrt{V}}e^{\frac{i}{\hbar}\boldsymbol{p}_i\boldsymbol{x}},$$

where V is the volume and where

$$\boldsymbol{p}_i \in \left\{2\pi\hbar\left(\frac{n_1}{a_1}, \frac{n_2}{a_2}, \frac{n_3}{a_3}\right) \,\middle|\, n_i \in \mathbb{Z}\right\}. \tag{3.19}$$

Hence the quantum numbers of the particle momentum form a lattice in momentum space, as is illustrated by Fig. 3.8. But now the wave functions ψ of the energy eigenstates of the N particles can be written in the form

$$\psi(\boldsymbol{x}) = \prod_{i=1}^{N}\psi_i(\boldsymbol{x}) = \prod_{i=1}^{N}\frac{1}{\sqrt{V}}e^{\frac{i}{\hbar}\boldsymbol{p}_i\boldsymbol{x}},$$

where of course all the $\boldsymbol{p}_i$ must satisfy Eq. 3.19, so that altogether the momentum quantum numbers of the ideal gas may be viewed as a lattice L in

Figure 3.9: Counting states. In order to obtain the partition function, one has to count the number of points of the lattice formed by the momentum quantum numbers that lie in the spherical shell for energies between $E-\Delta$ and E. For simplicity only two dimensions are shown.

the $3N$-dimensional momentum space. Hence we might say the state density in momentum space is just

$$1\text{ state}\Bigg/\left[\left(\frac{2\pi\hbar}{a_1}\right)\left(\frac{2\pi\hbar}{a_2}\right)\left(\frac{2\pi\hbar}{a_3}\right)\right]^N = 1\text{ state}\Bigg/\frac{(2\pi\hbar)^{3N}}{V^N}\,. \tag{3.20}$$

From the discussion of Sect. 3.3 it is clear that the partition function $\omega_\Delta(E)$ equals the number of energy eigenstates with energies between $E-\Delta$ and E. Hence we obtain the partition function by counting the number of lattice points of L that lie in a spherical shell with radii between $\sqrt{2m(E-\Delta)}$ and $\sqrt{2mE}$ (cf. Fig. 3.9).

Counting this number will in general be rather difficult. However, in the limit of a large box (i.e. for $2\pi\hbar/a_k \ll 1$), the error made by assuming that the state density is continuous will be negligibly small. Thus we may compute the number of states by multiplying the density given in Eq. 3.20 with the volume of the spherical shell,

$$\begin{aligned}\omega_\Delta(E) &= \frac{V^N}{(2\pi\hbar)^{3N}}\left[V_{3N}(\sqrt{2mE}) - V_{3N}(\sqrt{2m(E-\Delta)})\right]\\ &= \frac{1}{\frac{3N}{2}\Gamma\left(\frac{3N}{2}\right)}\left[V\left(\frac{1}{2\pi\hbar^2}\right)^{\frac{3}{2}}\right]^N\left\{(2mE)^{3N/2} - (2m[E-\Delta])^{3N/2}\right\}.\end{aligned} \tag{3.21}$$

But (apart from the factor $1/N!$ missing as we have considered distinguishable particles) this is exactly the result we got at the end of Sect. 3.4.2. Accordingly, the choice for the scaling factor advocated in Eq. 3.10 has been a wise one, as any other value for this factor would have resulted in an inconsistency between the classical and the quantum mechanical treatment.

Figure 3.10: Derivation of Stirling's approximation. $\ln N!$ may be viewed as the area of rectangles whose height is given by $\ln n$, where n runs from 1 to N. This area can be compared to that of the natural logarithm.

3.6 Stirling's approximation

In statistical physics, factors involving some factorial crop up on a regular basis, and as we shall see throughout the following sections, it is frequently necessary to take their logarithm. We might therefore ask whether there is some approximation which simplifies such logarithmic factors.

Due to the large numbers involved, at least for macroscopic systems such an approximation does in fact exist. It is generally known as Stirling's approximation and can be stated in the following way:

Theorem 17 (Stirling's approximation) *Let N be a large positive integer. Then the logarithm of $N!$ may be approximated by*

$$\ln N! \approx N \ln N - N + O(\ln N).$$

Proof: The logarithm of $N!$ may be written as

$$\ln N! = \ln \left(\prod_{n=1}^{N} n \right) = \sum_{n=1}^{N} \ln n,$$

where the sum on the right hand side can be regarded as the area of the N rectangles of side-lengths 1 and n with $n = 1, 2, \ldots, N$. But then from Fig. 3.10 we see that $\ln N!$ must fulfil the two relations

$$\begin{aligned} \ln N! &\geqslant \int_1^N \ln x \, \mathrm{d}x = [x \ln x - x]_1^N = N \ln N - N + 1 \\ \ln N! &\leqslant \int_1^{N+1} \ln x \, \mathrm{d}x = [x \ln x - x]_1^{N+1} = (N+1)\ln(N+1) - (N+1) + 1 \end{aligned}$$

Hence we get

$$N \ln N - N + 1 \leqslant \ln N! \leqslant N \ln N - N + N \ln \frac{N+1}{N} + \ln(N+1). \quad (3.22)$$

For large N we may expand $\ln((N+1)/N)$ to first order, so that

$$\begin{aligned} N\ln\frac{N+1}{N} &= N\left\{\frac{N+1}{N} - 1 + O\left(\left(\frac{N+1}{N} - 1\right)^2\right)\right\} \\ &= N\left\{\frac{1}{N} + O\left(\frac{1}{N^2}\right)\right\} = 1 + O(N^{-1}). \end{aligned}$$

Inserting this result into Eq. 3.22, we obtain

$$N\ln N - N + O(1) \leqslant \ln N! \leqslant N\ln N - N + O(N^{-1}) + O(\ln N),$$

from which we can derive $\ln N! = N\ln N - N + O(\ln N)$. *(QED)*

The reader is strongly encouraged to memorize Stirling's approximation. Perhaps this task can be accomplished more easily, if they note that it just asserts that $\ln N$ is approximately equal to the value of the anti-derivative of the natural logarithm at N.

3.7 Macroscopic limit and thermostatic entropy

Evidently, the precise form of the distribution function of a given system depends on the geometry, the boundary conditions, and so forth. However, these effects will vanish for macroscopic systems and thus can be discarded. Hence for such systems the bulk behavior is governed by the substance under consideration only, which clearly forms the basis for the bulk properties known from thermodynamics (such as pressure or energy).

Broadly speaking, concerning these bulk properties there are two main classes: Either a quantity depends on the amount of matter present or it doesn't. But let us be a little bit more precise about this. An *extensive* quantity is proportional to the particle number N, i.e. it is of the order $O(N)$. Examples would be the volume V, the energy E, and of course N itself. From a formal point of view, extensive quantities pose a problem, as in the limit of infinitely many particles they will diverge and hence possess no well-defined limit.

One gets rid of this annoyance by considering *specific* quantities, which are the ratio of extensive quantities. Here, as both the nominator and the denominator scale like N, the existence of a proper limit is ensured. For example, instead of V one may consider the specific volume $v = V/N$ or the (particle) density $n = v^{-1} = N/V$, and instead of E one may discuss the energy per particle, $e = E/N$, or the energy density $\epsilon = E/V$.

More generally, quantities which are independent of the amount of matter are called *intensive*. Thus it ought to be clear that while every specific must be an intensive quantity, the converse cannot be correct (temperature may serve as a counter-example).

As an additional bit of terminology we note that the process of letting the extensive quantities go towards infinity so that the specific quantities have a well-defined limit is called *macroscopic* or *thermodynamic limit*. It is precisely this limit that enables us to derive formulae describing the everyday world of macroscopic objects, as we shall see in due course.

We may now introduce the first quantity known from thermodynamics, which is also a further example of an extensive quantity.

Definition 25 *The* (thermostatic) entropy S *of a microcanonical ensemble is defined as*

$$S(E) = k_{\mathrm{B}} \ln \omega_{\Delta}(E), \tag{3.23}$$

where k_{B} *and* ω_{Δ} *are an appropriately chosen constant*[5] *and the partition function for the case of slightly variable energies, respectively.*

Before making a few comments on this definition, as an example we compute the entropy of the ideal gas. Here, Eq. 3.21 together with a prefactor $1/N!$ for indistinguishable particles yields

$$\omega_{\Delta}(E) = \frac{1}{\frac{3N}{2}\Gamma\left(\frac{3N}{2}\right)} \left[V \left(\frac{1}{2\pi\hbar^2}\right)^{\frac{3}{2}}\right]^N \left\{(2mE)^{3N/2} - (2m[E-\Delta])^{3N/2}\right\}.$$

But for a sufficiently small Δ we may use the approximation

$$(2mE)^{3N/2} - (2m[E-\Delta])^{\frac{3N}{2}} \approx \frac{3N}{2} E^{\frac{3N}{2}-1}\Delta,$$

and we obtain

$$\omega_{\Delta}(E) = \frac{\Delta}{N!\,\Gamma\left(\frac{3N}{2}\right)} \frac{1}{E} \left[V \left(\frac{mE}{2\pi\hbar^2}\right)^{\frac{3}{2}}\right]^N .$$

---

[5] It will turn out to be the Boltzmann constant.

From this we may compute the corresponding entropy:

$$\begin{aligned}
S(E,V) = k_{\mathrm{B}} \ln \omega_\Delta(E) = k_{\mathrm{B}} \ln \left\{ \frac{\Delta}{N!\,\Gamma\left(\frac{3N}{2}\right)} \frac{1}{E} \left[V \left(\frac{mE}{2\pi\hbar^2} \right)^{\frac{3}{2}} \right]^N \right\} \\
= k_{\mathrm{B}} \left\{ \ln \Delta - \ln N! - \ln \Gamma\left(\frac{3N}{2}\right) - \ln E \right. \\
\left. + N \ln V + \frac{3N}{2} \ln E + \frac{3N}{2} \ln \frac{m}{2\pi\hbar^2} \right\}
\end{aligned}$$

As we are interested in the macroscopic limit, we may get rid of the factorial by means of the Stirling approximation,

$$\ln N! \approx N \ln N - N.$$

Similarly, as $\Gamma(N) = (N-1)!$, it follows that for even N

$$\begin{aligned}
\ln \Gamma\left(\frac{3N}{2}\right) = \ln\left(\frac{3N}{2} - 1\right)! \approx \left(\frac{3N}{2} - 1\right) \ln\left(\frac{3N}{2} - 1\right) - \left(\frac{3N}{2} - 1\right) \\
\approx \left(\frac{3N}{2}\right) \ln\left(\frac{3N}{2}\right) - \left(\frac{3N}{2}\right)
\end{aligned}$$

holds valid. However, due to the conutinuity of the gamma function, the same approximation can be used for odd N as well, so that we get

$$\begin{aligned}
S(E,V) \approx k_{\mathrm{B}} \left\{ \ln \Delta - N \ln N + N - \frac{3N}{2} \ln \frac{3N}{2} + \frac{3N}{2} - \ln E \right. \\
\left. + N \ln V + \frac{3N}{2} \ln E + \frac{3N}{2} \ln \frac{m}{2\pi\hbar^2} \right\}.
\end{aligned}$$

Using the specific quantities $v = V/N$ and $e = E/N$ we thus obtain

$$\begin{aligned}
S(E,V) = k_{\mathrm{B}} \left\{ \ln \Delta - N \ln N + N - \frac{3N}{2} \ln \frac{3}{2} - \frac{3N}{2} \ln N + \frac{3N}{2} - \ln N \right. \\
\left. - \ln e + N \ln N + N \ln v + \frac{3N}{2} \ln N + \frac{3N}{2} \ln e + \frac{3N}{2} \ln \frac{m}{2\pi\hbar^2} \right\} \\
= k_{\mathrm{B}} N \left\{ \frac{5}{2} + \ln \left[v \left(\frac{me}{3\pi\hbar^2} \right)^{3/2} \right] \right\} + k_{\mathrm{B}} \ln \Delta - k_{\mathrm{B}} \ln N - k_{\mathrm{B}} \ln e.
\end{aligned}$$

As $\ln\Delta$, $\ln N$, and $\ln e$ grow more slowly than N, one can see immediately that in the macroscopic limit the entropy of the ideal gas will scale like N, so that it constitutes an extensive quantity. The corresponding *specific entropy* s_N, must then be a specific quantity, and indeed it has a well-defined limit s:

$$\begin{aligned} s(e,v) &= \lim_{N\longrightarrow\infty} s_N\left(\frac{E}{N},\frac{V}{N}\right) = \lim_{N\longrightarrow\infty}\frac{S(Ne,Nv)}{N} \\ &= k_{\mathrm{B}}\left\{\frac{5}{2}+\ln\left[v\left(\frac{me}{3\pi\hbar^2}\right)^{3/2}\right]\right\} + k_{\mathrm{B}}\lim_{N\longrightarrow\infty}\frac{\ln\Delta}{N} \\ &\qquad - k_{\mathrm{B}}\lim_{N\longrightarrow\infty}\frac{\ln N}{N} - k_{\mathrm{B}}\lim_{N\longrightarrow\infty}\frac{\ln e}{N}. \end{aligned}$$

But the last three terms on the right hand side vanish, and thus we arrive at the result

$$s(e,v) = k_{\mathrm{B}}\left\{\frac{5}{2}+\ln\left[v\left(\frac{me}{3\pi\hbar^2}\right)^{3/2}\right]\right\}, \tag{3.24}$$

which depends on specific quantities only. In particular, one should note that there is no dependence on the rather arbitrary parameter Δ any longer.

As promised above, we may now comment on our definition of the entropy:

(i) Eq. 3.23 is *just a definition*, and in fact, in due course we shall meeet other definitions of entropy. Fortunately, all these definitions will be consistent with each other in the case of equilibrium systems.

(ii) From the previous remark it should be clear that the equivalence (or at least consistency) of the thermostatic entropy with the one known from thermodynamics remains to be shown.

(iii) The value of the constant k_{B} hasn't been fixed yet. We will make up for that as soon as we have introduced the concept of temperature.

For the ideal gas, the entropy turned out to be extensive, so that we could introduce a specific entropy in a meaningful way. Now the question is, of course, whether that perhaps was just a happy coincidence. That this is not the case and that rather on the contrary the entropy is extensive for a broad class of Hamiltonians, is asserted by the following theorem:

Theorem 18 *If the Hamiltonian of a microcanonical ensemble with N particles is given by*

$$H(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m} + \sum_{i<j} U(\boldsymbol{q}_i-\boldsymbol{q}_j),$$

where for U the relation

$$U(-\boldsymbol{q}) = U(\boldsymbol{q})$$

holds valid, then

$$s_N(e,v) \equiv \frac{1}{N} S(eN, vN)$$

has a well defined limit

$$s(e,v) = \lim_{N \longrightarrow \infty} s_N(e,v),$$

which doesn't depend on the thickness Δ of the energy shell considered. Hence the specific entropy exists for the given Hamiltonian.

Proof: We don't attempt to give the proof, which is rather technical. Instead, we refer the interested reader to Chapter 3 of [9], where the thermodynamic limit of the entropy is discusssed from an even more general point of view. *(QED)*

3.8 Canonical Ensemble

So far, we have contented ourselves with microcanonical ensembles, i.e. with systems that have a fixed energy, volume, and particle number. However, under many circumstances these conditions will not be present. Imagine, for example, an experiment with no thermal isolation. Here, clearly, an energy flow to or from the surroundings will occur, spoiling energy conservation.

Hence, an extension of our formalism is called for. While this may sound like a piece of bad news, things turn out to surprisingly positive: Rather than rendering the computations more difficult, our generalized formalism will *faciliate* them.

Throughout the following sections we will investigate systems where energy, volume, and particle number may vary. In this section we shall start by allowing energy fluctuations, but keeping both volume and particle number fixed. In the next section, in addition we will allow for a variable volume. Finally, in Sect. 3.13 we will consider systems with varying energy and particle number. In all cases we shall assume that the system to be considered is placed in a very large second system, where both systems together form a microcanonical ensemble, and in addition we make the assumption that we may neglect the interaction energy between the two systems.

Figure 3.11: Two canonical ensembles, i.e. systems with fixed volume and particle number, but variable energy. Together the two systems are assumed to form a microcanonical ensemble.

Actually, the recipe used for obtaining the respective probability densities[6] will be the same for every system under consideration. It may be summarized as follows:

1. Write the probability density of the first system as an integral over that of the combined system, where the integration is carried out over all possible states of the second (large) system.
2. Expand the entropy of the second system, which occurs in the integrand, to first order.
3. Give appropriate names (such as temperature or pressure) to the first-order derivatives of this expansion, and thus obtain the desired probability density.
4. Convince yourself that the name given in the previous step was well-chosen, i.e. identify the defined quantities with quantities known from thermodynamics.

We will elaborate on each of these steps in the discussion to come. But first things first. Therefore, let us focus on systems with varying energy, but fixed volume and particle number. Such systems are called *canonical ensembles.* We assume that there are two canonical ensembles, which together constitute a microcanonical ensemble (cf. Fig. 3.11). The interaction energy between the two systems is taken to be negligible, which is motivated by the fact that whereas the energies of the two systems are extensive quantities, the interaction just depends on the common surface area. Hence the Hamiltonian of the composite system is given by

$$H = H_1 + H_2 + H_{12} \approx H_1 + H_2, \tag{3.25}$$

---

[6] Henceforth we shall speak of probability densities, even though dimensional quantities are considered.

where H_1, H_2, and H_{12} denote the Hamiltonian of the first system, the second system, and the interaction, respectively. Analogously, for the corresponding energies we obtain

$$E = E_1 + E_2 + E_{12} \approx E_1 + E_2.$$

Before going on, we need to give a more precise meaning to the notion of a "very large system". This is accomplished by the definition of reservoirs:

Definition 26 *Let there be two systems 1 and 2 with extensive quantities* $A^{(1)}$, $A^{(2)}$, …, $A^{(M)}$. *Then we call system 2 a* reservoir of the $A^{(k)}$ *if*

- *(i) system 2 is in equilibrium,*
- *(ii) the relation*
  $$A^{(k)} \equiv (N_1 + N_2)a^{(k)} \approx N_2 a^{(k)}$$
  holds valid for all $A^{(k)}$ *(i.e. system 1 can be neglected as compared to system 2), and*
- *(iii) the specific entropy of system 2 can be approximated as*

$$\begin{aligned} s_2 &\left(\frac{A^{(1)} - A_1^{(1)}}{N_2}, \ldots, \frac{A^{(M)} - A_1^{(M)}}{N_2}\right) \\ &\approx s_2\left(\frac{A^{(1)}}{N_2}, \ldots, \frac{A^{(M)}}{N_2}\right) - \sum_{k=1}^{M} \frac{\partial s_2}{\partial(A^{(k)}/N_2)}\left(\frac{A^{(1)}}{N_2}, \ldots, \frac{A^{(M)}}{N_2}\right) \cdot \frac{A_1^{(k)}}{N_2} \\ &\qquad + O\left(\frac{1}{N_2^2}\right). \end{aligned}$$

In particular, if one of the $A^{(k)}$ *is the energy, system 2 is known as a* heat bath.

The reader might have an objection concerning this definition: We introduced the entropy in the context of microcanonical ensembles, and whereas it is true that the total system is a microcanonical ensemble, its two subsystems definitely are not. So does it make sense to speak of the entropy of system 2?

The answer is yes. To see this, the reader should remember that the entropy basically was defined as a logarithm of a sum or integral over accessible states, and hence is fixed by the Hamiltonian of the system under consideration. Or to put it more bluntly, entropy is just a definition which we can apply to any system which possesses a Hamiltonian. (Whether the application is a meaningful one, is a completely diffferent question.)

Actually, condition (3) of Def. 26 may be restated in terms of the entropy of system 2:

$$\begin{aligned} S_2(A^{(1)}-A_1^{(1)},\ldots,A^{(M)}-A_1^{(M)}) &= N_2 s_2\left(\frac{A^{(1)}-A_1^{(1)}}{N_2},\ldots,\frac{A^{(M)}-A_1^{(M)}}{N_2}\right) \\ &\approx N_2\left\{s_2\left(\frac{A^{(1)}}{N_2},\ldots,\frac{A^{(M)}}{N_2}\right)\right. \\ &\qquad \left. -\sum_{k=1}^{M}\frac{\partial s_2}{\partial(A^{(k)}/N_2)}\left(\frac{A^{(1)}}{N_2},\ldots,\frac{A^{(M)}}{N_2}\right)\cdot\frac{A_1^{(k)}}{N_2}+O\left(\frac{1}{N_2^2}\right)\right\} \\ &= S_2(A^{(1)},\ldots,A^{(M)})-\sum_{k=1}^{M}\frac{\partial S_2}{\partial A^{(k)}}(A^{(1)},\ldots,A^{(M)})\cdot A_1^{(k)}+O(N_2^{-1}). \end{aligned} \tag{3.26}$$

We can now use the recipe stated above in order to investigate an equilibrium system with variable energy, but fixed volume and particle number, which is placed in a heat bath. Such systems are called *canonical ensembles.* We'll go through the recipe step by step for a system with a non-countable set of states. However, everything remains correct for systems with discrete states, as the reader should check for themselves.

step 1:

Let system 1 and 2 be the system under consideration and the heat bath, respectively. As we assume the combined system is a microcanonical ensemble, we may write down its phase space density ρ immediately,

$$\rho(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{\omega(E)}\delta(E-H(\boldsymbol{p},\boldsymbol{q})),$$

and employing Eq. 3.25 we obtain

$$\rho(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{\omega(E)}\delta(E-H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)-H_2(\boldsymbol{p}_2,\boldsymbol{q}_2)). \tag{3.27}$$

Now one gets the corresponding phase space density ρ_1 of system 1 by integrating ρ over the phase space Γ_2 of system 2,

$$\rho_1(\boldsymbol{p}_1,\boldsymbol{q}_1) = \int \mathrm{d}\Gamma_2\rho(\boldsymbol{p},\boldsymbol{q}) = \int \mathrm{d}\Gamma_2\frac{1}{\omega(E)}\delta(E-H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)-H_2(\boldsymbol{p}_2,\boldsymbol{q}_2)).$$

But $1/\omega(E)$ is a constant and hence may be drawn in front of the integral. Thus it follows that

$$\begin{aligned}\rho_1(\boldsymbol{p}_1, \boldsymbol{q}_1) &= \frac{1}{\omega(E)} \int \mathrm{d}\Gamma_2\, \delta(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1) - H_2(\boldsymbol{p}_2, \boldsymbol{q}_2)) \\ &= \frac{1}{\omega(E)} \cdot \omega_2(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1)),\end{aligned}$$

where in the last step we have just used the definition of the microcanonical partition function, Eq. 3.12. Together with Def. 25 this leads to

$$\rho_1(\boldsymbol{p}_1, \boldsymbol{q}_1) = \frac{1}{\omega(E)} \cdot \exp\left(-\frac{1}{k_\mathrm{B}} S_2(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1))\right). \tag{3.28}$$

step 2:
The total energy E may be regarded as the value of the Hamiltonian of the combined system. Because system 2 constitutes a heat bath, we may therefore use Eq. 3.26 to expand the entropy occuring in Eq. 3.28:

$$S_2(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1)) \approx S_2(E) - \frac{\partial S_2(E)}{\partial E} \cdot H_1(\boldsymbol{p}_1, \boldsymbol{q}_1) + O(N_2^{-1}).$$

In the thermodynamic limit we obtain

$$\rho_1(\boldsymbol{p}_1, \boldsymbol{q}_1) = \frac{1}{\omega(E)} \cdot \exp\left(-\frac{1}{k_\mathrm{B}} \left\{S_2(E) - \frac{\partial S_2(E)}{\partial E} \cdot H_1(\boldsymbol{p}_1, \boldsymbol{q}_1)\right\}\right). \tag{3.29}$$

step 3:
Let us now give the first-order derivative of Eq. 3.29 a nice and decent name.

Definition 27 *The* temperature T *of a heat bath with entropy S and energy E is given by*

$$\frac{1}{T} \equiv \frac{\partial S(E)}{\partial E}.$$

This definition is not particularly intuitive, but we shall show in the next step that it makes sense, and in addition we will generalize the concept of temperature to systems which aren't heat baths.

At any rate, inserting the temperature into Eq. 3.29 yields

$$\begin{aligned}\rho_1(\boldsymbol{p}_1,\boldsymbol{q}_1) &= \frac{1}{\omega(E)}\cdot\exp\left(-\frac{S_2(E)}{k_\mathrm{B}}-\frac{1}{k_\mathrm{B}T}\cdot H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)\right)\\ &= \frac{e^{-S_2(E)/k_\mathrm{B}}}{\omega(E)}\cdot e^{-H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)/k_\mathrm{B}T}.\end{aligned}$$

Here $e^{-S_2(E)/k_\mathrm{B}}/\omega(E)$ is a constant, the value of which we need not compute, as we know that ρ_1 must be normalized, $\int \mathrm{d}\Gamma_1\,\rho_1 = 1$. Hence we have proved the following theorem:

Theorem 19 *The phase space density ρ of a canonical ensemble with a continuum of states is given by*

$$\rho(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{Z(T)}e^{-H(\boldsymbol{p},\boldsymbol{q})/k_\mathrm{B}T},$$

where T denotes the temperature of the surrounding heat bath and

$$Z(T) \equiv \int \mathrm{d}\Gamma\, e^{-H(\boldsymbol{p},\boldsymbol{q})/k_\mathrm{B}T}$$

is the canonical partition function*.*

As stated above, the reasoning which led to the preceding theorem remains valid for discrete states as well. We therefore get

Theorem 20 *Consider a canonical ensemble with discrete microstates s_i, $i\in\mathbb{N}$. Then the probability $P(s_k)$ of the state s_k has the value*

$$P(s_k) = \frac{1}{Z(T)}e^{-H(s_k)/k_\mathrm{B}T},$$

where T is the temperature of the surrounding heat bath, $H(s_k)$ denotes the value of the Hamiltonian for s_k, and

$$Z(T) \equiv \sum_{s_i} e^{-H(s_i)/k_\mathrm{B}T}$$

is the canonical partition function*.*

In Theorems 19 and 20 we have made our first acquaintance with the factor $1/k_{\mathrm{B}}T$. It is of such importance that introducing a special notation for it is an extremely reasonable idea.

Definition 28 *The factor $1/k_{\mathrm{B}}T$ is abbreviated as β,*

$$\beta \equiv \frac{1}{k_{\mathrm{B}}T}.$$

step 4:
It remains to be shown that the temperature as defined by Def. 27 is consistent with the notion of temperature known from phenomenological thermodynamics. To do this, we start by checking that all is well for the case of a heat bath consisting of an ideal gas. Here from Eq. 3.7 we may infer that the entropy can be written as

$$S(E) = \frac{3}{2}Nk_{\mathrm{B}}\ln E + \text{terms not depending on } E,$$

so that the temperature is given by

$$\frac{1}{T} = \frac{\partial S(E)}{\partial E} = \frac{3}{2}\frac{Nk_{\mathrm{B}}}{E}. \tag{3.30}$$

Hence we recover the well-known relation $E = \frac{3}{2}Nk_{\mathrm{B}}T$ between inner energy and temperature of the ideal gas. This clearly makes our choice for the definition a plausible one.

Now let us take a more general point of view. So far, we have defined temperature only for heat baths; but we may extend the definition to cover other systems in equilibrium as well:

Definition 29 *Consider an equilibrium system with an entropy S, and let $\bar{E}$ denote its most probable energy. Then the temperature of the system is defined as the inverse of the derivative of S with respect to the energy E at $\bar{E}$,*

$$\frac{1}{T} \equiv \frac{\partial S(\bar{E})}{\partial E}.$$

Note that this is consistent with Def. 27, as in the case of a heat bath, the energy can be assumed to remain constant and thus by definition may be regarded as the most probable energy of the system. In addition, it should be pointed out that as a derivative of an extensive with respect to another extensive quantity, temperature must be an intensive quantity.

From phenomenological thermodynamics we know that if two systems are in equilibrium they have the same temperature, and we will now proceed to show that this holds valid for our definition of temperature.

To this end consider a microcanonical ensemble composed of two canonical ensembles 1 and 2, none of which needs to be a heat bath. Then the probability density $\rho_1(E_1)$ of the energy of system 1 is just the probability density of the Hamiltonian of that system,

$$\rho_1(E_1) = \rho_{H_1}(E_1) = \int \mathrm{d}h_1 \rho_{H_1}(h_1)\delta(h_1 - E_1) = \langle \delta(H_1 - E_1) \rangle \tag{3.31}$$

Assume we are dealing with indistinguishable particles in d dimensions. Then at first sight, the differential "phase space volume element" of the combined system should be

$$\mathrm{d}\Gamma = \frac{1}{(N_1 + N_2)!} \frac{\mathrm{d}^{d\cdot N_1} p_1 \, \mathrm{d}^{d\cdot N_1} q_1 \, \mathrm{d}^{d\cdot N_2} p_2 \, \mathrm{d}^{d\cdot N_2} q_2}{(2\pi\hbar)^{d\cdot(N_1+N_2)}}, \quad \text{(wrong!)}$$

because after all we have $(N_1 + N_2)!$ indistinguishable particles. But the factorial in this expression arose from the fact that states which are obtained by merely exchanging particles may be identified with each other and thus must be counted once only. However, when discussing particles in two separate systems, one normally doesn't consider particle exchanges between the systems, so that the necessity of identifying states connected with each other via such exchanges does not arise.

Hence, as particle exchanges within the individual systems still have to be taken care of, the phase space volume element to be used is given by

$$\mathrm{d}\Gamma_1 \, \mathrm{d}\Gamma_2 = \frac{1}{N_1! N_2!} \frac{\mathrm{d}^{d\cdot N_1} p_1 \, \mathrm{d}^{d\cdot N_1} q_1 \, \mathrm{d}^{d\cdot N_2} p_2 \, \mathrm{d}^{d\cdot N_2} q_2}{(2\pi\hbar)^{d\cdot(N_1+N_2)}}. \tag{3.32}$$

But then from Eq. 3.31, Eq. 3.27, and Theorem 4 we can infer that $\rho_1(E_1)$ may be written as

$$\begin{aligned}\rho_1(E_1) = \langle\delta(H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)-E_1)\rangle &= \int \mathrm{d}\Gamma_1\,\mathrm{d}\Gamma_2\,\rho(\boldsymbol{p},\boldsymbol{q})\delta(H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)-E_1)\\ &= \int \mathrm{d}\Gamma_1\,\mathrm{d}\Gamma_2\,\frac{1}{\omega(E)}\delta(E-H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)-H_2(\boldsymbol{p}_2,\boldsymbol{q}_2))\delta(H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)-E_1)\\ &= \int \mathrm{d}\Gamma_1\,\mathrm{d}\Gamma_2\,\frac{1}{\omega(E)}\delta(E-E_1-H_2(\boldsymbol{p}_2,\boldsymbol{q}_2))\delta(H_1(\boldsymbol{p}_1,\boldsymbol{q}_1)-E_1),\end{aligned}$$

where in the last step we used the fact that $\delta(H_1-E_1)$ is non-zero for $E_1=H_1$ only. As the total energy E is a constant, we may draw the factor $1/\omega(E)$ in front of the integral, and together with the definition of the microcanonical partition function we obtain

$$\begin{aligned}\rho_1(E_1) &= \frac{1}{\omega(E)}\int \mathrm{d}\Gamma_1\,\delta(E_1-H_1(\boldsymbol{p}_1,\boldsymbol{q}_1))\int \mathrm{d}\Gamma_2\,\delta(E-E_1-H_2(\boldsymbol{p}_2,\boldsymbol{q}_2))\\ &= \frac{1}{\omega(E)}\omega_1(E_1)\omega_2(E-E_1).\end{aligned}$$

Accordingly, the derivative of $\rho_1(E_1)$ is given by

$$\rho_1'(E_1) = \frac{1}{\omega(E)}\left(\omega_1'(E_1)\omega_2(E-E_1)-\omega_1(E_1)\omega_2'(E-E_1)\right).$$

The most probable energy $\bar{E}_1$ of system 1 obviously must fulfill the relation $\rho_1'(\bar{E}_1)=0$, so that

$$\frac{\omega_1'(\bar{E}_1)}{\omega_1(\bar{E}_1)} = \frac{\omega_2'(E-\bar{E}_1)}{\omega_2(E-\bar{E}_1)}.$$

Due to energy conservation $E=E_1+E_2$ and thus $\bar{E}_2=E-\bar{E}_1$. Hence together with the formula $(\ln\omega)'=\omega'/\omega$ and the definition of the entropy we get

$$\frac{\partial S_1(\bar{E}_1)}{\partial \bar{E}_1} = k_{\mathrm{B}}\frac{\partial \ln\omega_1(\bar{E}_1)}{\partial \bar{E}_1} = k_{\mathrm{B}}\frac{\omega_1'(\bar{E}_1)}{\omega_1(\bar{E}_1)} = k_{\mathrm{B}}\frac{\omega_2'(\bar{E}_2)}{\omega_2(\bar{E}_2)} = \frac{\partial S_2(\bar{E}_2)}{\partial \bar{E}_2},$$

and hence from Def. 29 we may conclude:

Theorem 21 *Two canonical ensembles which are in equilibrium with each other (so that together they constitute a microcanonical ensemble) have the same temperature.*

Figure 3.12: One-dimensional array of non-interacting spins in a homogeneous magnetic field B. While the spins may point in any direction, only the components parallel to $\boldsymbol{B}$ are relevant for the Hamiltonian.

Thus our definition of temperature is a reasonable one, and Eq. 3.30 shows that it coincides with that of the empirical temperature measured with normal thermometers. The Boltzmann constant k_{B} may be fixed by demanding that temperature is measured in Kelvin (i.e. the appropriate SI unit). In that case k_{B} has the value

$$k_{\mathrm{B}} \approx 1.38 \times 10^{-23}\ \mathrm{J/K} \approx 8.62 \times 10^{-5}\ \mathrm{eV/K}.$$

In thermodynamics, instead of k_{B} one often uses the *ideal gas constant* R, which is defined as the product of the Boltzmann constant and Avogadro's number N_{A},[7]

$$R \equiv k_{\mathrm{B}} N_{\mathrm{A}} \approx 1.38 \times 10^{-23}\ \mathrm{J/K} \cdot 6.022 \times 10^{23}\ \mathrm{mol}^{-1} \approx 8.31\ \frac{\mathrm{J}}{\mathrm{K} \cdot \mathrm{mol}}.$$

To conclude this section, we note that if the entropy is a monotonously increasing function of energy, temperature is sctrictly non-negative. While this is true in everyday life, one may devise systems where this condition is *not* met, so that negative temperatures may arise. We will discuss an example for this in the following section.

3.9 Example: spins in a magnetic field

As an example of the formalism introduced in the previous section, we consider a one-dimensional lattice of N non-interacting electron spins in an homogenous external magnetic field B, as depicted in Fig. 3.12. From quantum mechanics we know that for such a system the operator $\hat{\boldsymbol{\mu}}$ corresponding to the magnetic moment is given by

---

[7] The reader might note the first three digits of the numerical factor of R (i.e. 831) are just the mirror image of the first three digits of that of k_{B} (i.e. 138). While this is sheer coincidence, of course, it may nonetheless serve as a mnemonic.

$$\hat{\boldsymbol{\mu}} = -\frac{e}{m_e} \sum_{i=1}^{N} \hat{\boldsymbol{s}}_i,$$

where m_e and e denote the electron mass and charge magnitude, respectively. The spin operator $\hat{\boldsymbol{s}}$ for each electron has the form

$$\hat{\boldsymbol{s}} = \frac{\hbar}{2} \begin{pmatrix} \hat{\sigma}_x & \hat{\sigma}_y & \hat{\sigma}_z \end{pmatrix} \equiv \frac{\hbar}{2} \hat{\boldsymbol{\sigma}}$$

with the well-known Pauli matrices $\hat{\sigma}_{x,y,z}$. Accordingly the system is described by the Hamiltonian

$$\hat{H} = \hat{\boldsymbol{\mu}} \boldsymbol{B} = -\frac{e\hbar}{2m_e} \sum_{i=1}^{N} \hat{\boldsymbol{\sigma}}_i \boldsymbol{B}.$$

Now without loss of generality we may assume that $\boldsymbol{B}$ points in the z direction, $B_z = B$, so that

$$\hat{H} = \frac{e\hbar}{2m_e} \sum_{i=1}^{N} \hat{\sigma}_{z,i} B.$$

As $\hat{\sigma}_z$ has the form

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

so that its eigenvalues are -1 and +1, it is clear that for the case of just one spin, from the point of view of statistical mechanics one may replace the quantum mechanical Hamiltonian $\hat{H} = (e\hbar/2m_e)\hat{\sigma}_z$ by the classical Hamiltonian $H = (e\hbar/2m_e)\sigma$ with $\sigma \in \{-1, 1\}$. However, as all the spin operators $\hat{s}_z$ commute, the transition from one to N spins adds no quantum mechanical nature, and the system containing N spins can be replaced by a classical system possessing the Hamiltonian

$$H = \frac{e\hbar}{2m_e} \sum_{i=1}^{N} \sigma_i, \tag{3.33}$$

where again $\sigma_i \in \{-1, 1\}$. To faciliate the notation, let us define $h \equiv e\hbar/2m_e$. Then it is easy to see that the possible energies E of the system under consideration must lie in the range

$$-Nh \leqslant E \leqslant Nh,$$

which implies that the energy is bounded from above. Hence we should expect some surprise along the way. But first let us introduce the quantities N_+ and N_- denoting the number of spins pointing upward and downard, respectively. Their difference obviously equals the sum M of all σ_i,

$$M \equiv \sum_{i=1}^{N} \sigma_i = N_+ - N_-,$$

and their sum just corresponds to the number of spins, $N_+ + N_- = N$. Now as each spin can adopt one of two states, as the spins are distinguishable, and as there is a one-to-one correspondence between M and N_+, one may conclude from elementary combinatorics that the number $\omega(M)$ of states with a given M has the value

$$\omega(M) = \binom{N}{N_+} = \frac{N!}{N_+!(N - N_+)!}.$$

Using $n_+ \equiv N_+/N$ and employing the Stirling approximation, we thus get

$$\begin{aligned}
\ln \omega(M) &\approx N \ln N - N - N_+ \ln N_+ + N_+ \\
&\qquad\qquad - (N - N_+) \ln(N - N_+) + (N - N_+) \\
&= N \ln N - N - N n_+(\ln N + \ln n_+) + N n_+ \\
&\qquad\qquad - N(1 - n_+)(\ln N + \ln(1 - n_+)) + N(1 - n_+) \\
&= -N(n_+ \ln n_+ + (1 - n_+) \ln(1 - n_+)).
\end{aligned}$$

We may express n_+ by means of N and M as

$$n_+ = \frac{N_+}{N} = \frac{N_+ + N_- + N_+ - N_-}{2N} = \frac{N + M}{2N},$$

and from Eq. 3.33 we may see that the energy E of the system is given by

$$E = h \sum_{i=1}^{N} \sigma_i = h(N_+ - N_-) = hM.$$

Figure 3.13: Entropy per spin as a function of energy per spin for the one-dimensional spin lattice.

But the logarithm of $\omega(M)$ is just the entropy divided by the Boltzmann constant. Hence, putting all together and introducing the energy e per spin, $e \equiv E/N$, we arrive at

$$\begin{aligned}
S(E) &= k_{\mathrm{B}} \ln \omega(M(E)) \\
&= -Nk_{\mathrm{B}} \Bigg(\frac{M(E)+N}{2N} \ln \frac{M(E)+N}{2N} \\
&\qquad\qquad + \left(1 - \frac{M(E)+N}{2N}\right) \ln \left(1 - \frac{M(E)+N}{2N}\right)\Bigg) \\
&= -\frac{1}{2} N k_{\mathrm{B}} \left(\left(1+\frac{e}{h}\right) \ln \left(1+\frac{e}{h}\right) + \left(1-\frac{e}{h}\right) \ln \left(1-\frac{e}{h}\right)\right).
\end{aligned}$$

Evidently the entropy is extensive (as should have been expected), and the entropy s per spin has the value

$$s(e) \equiv \frac{S(E)}{N} = -\frac{1}{2} k_{\mathrm{B}} \left(\left(1+\frac{e}{h}\right) \ln \left(1+\frac{e}{h}\right) + \left(1-\frac{e}{h}\right) \ln \left(1-\frac{e}{h}\right)\right).$$

Now from Def. 29 we can obtain the $\beta = 1/k_{\mathrm{B}}T$ of the system,

$$\beta = \frac{1}{k_{\mathrm{B}}} \frac{\mathrm{d}S(E)}{\mathrm{d}E} = \frac{1}{k_{\mathrm{B}}} \frac{\mathrm{d}s(e)}{\mathrm{d}e} = \frac{1}{2h} \ln \frac{1+e/h}{1-e/h},$$

and thus its temperature,

$$T = \frac{h}{k_{\mathrm{B}}} \ln \frac{1-e/h}{1+e/h}.$$

In Figs. 3.13 and 3.14 the entropy per spin and the temperature are plotted as a function of the energy per spin. One sees that for negative energies the temperature actually becomes negative, a perhaps somewhat unexpected result.

The reader now may wonder whether or not to worry about this. After all, from everyday experience they must have got the impression that temperature is strictly non-negative, after all. There are various replies to this:

Figure 3.14: Temperature as a function of the energy per spin for the one-dimensional spin lattice.

First, one should bear in mind that Def. 29 is, well, just a definition. Hence from a purely formal point of view you might just shrug your shoulders, saying if one makes up strange definitions, one shouldn't wonder about their outcome.

Second, the notion of positive energies is based on heat baths and the like, in which surely the energy is not bounded. However, this condition is not fulfilled in the system just discussed.

Finally, looking at the Boltzmann factor, one notes that a negative energy implies that the probability of some state increases with its energy. Such an inverted population constitutes one of the key ingredients for any laser.

3.10 Example: Maxwell velocity distribution

As a second example of a canonical ensemble let us consider the momentum distribution of a classical gas of N interacting particles of mass m. However, before doing so, we briefly consider a general random vector $\boldsymbol{X}$ with some probability distribution $\rho_{\boldsymbol{X}}(\boldsymbol{x})$. Then from Theorem 4 we obtain for the mean of $\delta(\boldsymbol{x}-\boldsymbol{X})$:

$$\langle\delta(\boldsymbol{x}-\boldsymbol{X})\rangle = \int \mathrm{d}^N x\,\delta(\boldsymbol{x}-\boldsymbol{x}')\rho_{\boldsymbol{X}}(\boldsymbol{x}') = \rho_{\boldsymbol{X}}(\boldsymbol{x}).$$

Thus we have proved the following theorem (cf. Eq. 3.31):

Theorem 22 *Let $\boldsymbol{X}$ be some random variable. Then the probability density $\rho_{\boldsymbol{X}}(\boldsymbol{x})$ of X can be expressed as*

$$\rho_{\boldsymbol{X}}(\boldsymbol{x}) = \langle\delta(\boldsymbol{x}-\boldsymbol{X})\rangle\,.$$

We may now apply this result to the classical gas of interacting particles. The corresponding Hamiltonian is given by

$$H(\boldsymbol{r}_1,\dots,\boldsymbol{r}_N,\boldsymbol{p}_1,\dots,\boldsymbol{p}_N) = \sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m} + \phi(\boldsymbol{r}_1,\dots,\boldsymbol{r}_N),$$

where $\boldsymbol{r}$ and $\boldsymbol{p}$ denote the position and the momentum of the particles, and where ϕ is the interaction potential. As the phase space density ρ_Γ has the form $\rho_\Gamma = e^{-H/k_{\mathrm{B}}T}/Z$, using Theorems 22 and 4 we arrive at

$$\begin{aligned}
\rho_{\boldsymbol{p}_k}(\boldsymbol{p}_k) &= \int \mathrm{d}\Gamma'\, \delta(\boldsymbol{p}_k - \boldsymbol{p}'_k)\rho_\Gamma(\boldsymbol{r}'_1, \ldots, \boldsymbol{r}'_N, \boldsymbol{p}'_1, \ldots, \boldsymbol{p}'_N) \\
&= \frac{1}{Z}\int \frac{\mathrm{d}^3r'_1 \ldots \mathrm{d}^3r'_N \mathrm{d}^3p'_1 \ldots \mathrm{d}^3p'_N}{(2\pi\hbar)^{3N}} \delta(\boldsymbol{p}_k - \boldsymbol{p}'_k)\cdot \\
&\qquad\qquad \cdot \exp\left\{ -\frac{1}{k_{\mathrm{B}}T}\left(\sum_{i=1}^{N} \frac{\boldsymbol{p}'^2_i}{2m} + \phi(\boldsymbol{r}'_1, \ldots, \boldsymbol{r}'_N \right) \right\} \\
&= \frac{1}{Z} e^{-\boldsymbol{p}_k^2/2mk_{\mathrm{B}}T} \int \frac{\prod_{\substack{i=1 \\ i\neq k}}^{N} \mathrm{d}^3p'_i}{(2\pi\hbar)^{3N}} e^{-\boldsymbol{p}'^2_i/2mk_{\mathrm{B}}T} \cdot \\
&\qquad\qquad \cdot \int \mathrm{d}^3r'_1 \ldots \mathrm{d}^3r_N e^{-\phi(\boldsymbol{r}'_1, \ldots, \boldsymbol{r}'_N)/k_{\mathrm{B}}T} \\
&= C e^{-\boldsymbol{p}_k^2/2mk_{\mathrm{B}}T},
\end{aligned}$$

where the constant C follows from the normalization condition for $\rho_{\boldsymbol{p}_k}$, i.e. from

$$1 \overset{!}{=} \int \mathrm{d}^3p\, \rho_{\boldsymbol{p}_k}(\boldsymbol{p}) = C \int \mathrm{d}^3p\, e^{-p^2/2mk_{\mathrm{B}}T}.$$

If we set $K \equiv 1/2mk_{\mathrm{B}}T$, use $e^{-Kp^2} = e^{-K(-p)^2}$, and employ Footnote 1 on page 43, we may compute the integral on the right hand side as follows:

$$\begin{aligned}
\int \mathrm{d}^3p\, e^{-p^2/2mk_{\mathrm{B}}T} &= \int \mathrm{d}^3p\, e^{-Kp^2} = 4\pi \int_0^\infty \mathrm{d}p\, p^2 e^{-Kp^2} \\
&= -4\pi \frac{\partial}{\partial K} \int_0^\infty \mathrm{d}p\, e^{-Kp^2} = -2\pi \frac{\partial}{\partial K} \int_{-\infty}^\infty \mathrm{d}p\, e^{-Kp^2} \\
&= -2\pi \frac{\partial}{\partial K}\left(\sqrt{\frac{\pi}{K}} \right) = \pi^{3/2} K^{-3/2} = (2\pi m k_{\mathrm{B}}T)^{3/2}.
\end{aligned}$$

We thus find that $C = (2\pi m k_{\mathrm{B}}T)^{-3/2}$. Obviously, the distribution doesn't depend on the specific particle chosen. Hence we may drop the subscript $\boldsymbol{p}_k$ and obtain

$$\rho(\boldsymbol{p}) = \frac{1}{(2\pi m k_{\mathrm{B}}T)^{3/2}} e^{-\frac{\boldsymbol{p}^2}{2mk_{\mathrm{B}}T}}.$$

Figure 3.15: Maxwell velocity distribution $\rho(v)$ for a classical electron gas at various temperatures T. The gas is assumed to be sufficiently dilute so that electron-electron interactions can be neglected.

This is known as the *Maxwell momentum distribution.* We leave it for the reader to show that the corresponding distribution of the absolute value of the momentum is given by

$$\rho(p) = 4\pi p^2 \rho(\boldsymbol{p}) = \sqrt{\frac{2}{\pi}} \left(\frac{1}{m k_{\mathrm{B}} T} \right)^{3/2} p^2 e^{-p^2/2mk_{\mathrm{B}}T},$$

so that the distribution of the particle velocity $v = p/m$ must be of the form

$$\rho(v) = \sqrt{\frac{2}{\pi}} \left(\frac{m}{k_{\mathrm{B}} T} \right)^{3/2} * v^2 * e^{-mv^2/2k_{\mathrm{B}}T}.$$

The velocity distribution is shown in Fig. 3.15 for the case of electrons at various temperatures.

3.11 Example: Barometric formula

To conclude our discussion on canonical ensembles, let us turn to the derivation of the barometric formula describing the density of the atmosphere. As (at least in this respect) the curvature of the Earth is negligibly small, we may assume that concerning the potential, there is just a dependence on the height z above the surface of the Earth. Correspondingly, for N particles of mass m we obtain the relation

$$H(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + \sum_{i=1}^{N} mgz_i$$

for the Hamiltonian. Now we treat the surface below and the atmosphere above the N particles under consideration as a heat bath, and in addition we assume that the volume is constrained to some fixed value. Then we have a canonical ensemble, and with the aid of Theorem 22 we get for the height distribution of the k-th particle that

$$\begin{aligned}\rho_{z_k}(z) &= \langle \delta(z - z_k) \rangle \\ &= \frac{1}{Z} \int \mathrm{d}\Lambda \delta(z - z_k) \exp\left\{ -\frac{1}{k_{\mathrm{B}}T} \left(\sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + \sum_{i=1}^{N} mgz_i \right) \right\} \\ &= Ce^{-mgz/k_{\mathrm{B}}T},\end{aligned}$$

where C is a constant the value of which can be computed from the normalization condition

$$1 \overset{!}{=} \int_0^\infty \mathrm{d}z \rho_{z_k}(z) = \int \mathrm{d}z\, Ce^{-mgz/k_{\mathrm{B}}T} = C\frac{k_{\mathrm{B}}T}{mg}.$$

Again, the distribution doesn't depend on the particle, so that we can drop the index $_{z_k}$. We thus obtain

$$\rho(z) = \frac{mg}{k_{\mathrm{B}}T} e^{-mgz/k_{\mathrm{B}}T}.$$

Note that this formula may be used to measure heights by means of a barometer: If the temperature is known, the height is just a function of the difference between the atmospheric pressure at ground level and that at the height to be measured.[8]

3.12 Canonical ensemble with variable volume

So far we have considered canonical ensembles with a fixed volume V only. We now drop this assumption, so that it is reasonable to write the Hamiltonian and thus the entropy as a function of both energy and volume.

This raises the question whether introducing volume as an additional coordinate is merely an act of convenience or of necessity. The answer is

---

[8] Niels Bohr proposed various other methods for measuring the height of a building with a barometer. For example, one might drop the barometer from the top and measure the time of flight. Alternatively, and perhaps more reliably, one might give it as a present to the caretaker in return for the desired piece of information.

Figure 3.16: Canonical ensembles with variable volume, i.e. systems where energy and volume may vary, but where the particle number remains fixed. The two systems together are supposed to form a microcanonical ensemble.

somewhat ambiguous. On the one hand, if you really compute the Hamiltonian in terms of the positions and moments of *all* particles, there is no need for an additional coordinate. On the other hand, for example, in the case of a gas contained in a box with a moveable piston this would mean that we have to provide a complete microscopic description of the piston, which is of course hardly possible.

Hence we introduce a volume-dependence of the Hamiltonian, if the state of some part of the system (such as the piston mentioned in the previous paragraph) can be fully described by specifying the volume of the system. Thus it usually constitutes the part of the Hamiltonian due to work done by or on the system. We will come back to this point at the beginning of Chapter **??**.

A system with fixed particle number but variable energy and volume is called *canonical ensemble with variable volume*. Similarly to the case of canonical ensembles, we consider two canonical ensembles with variable volume (henceforth called 'system 1' and 'system 2'), which are in contact with each other and together form a microcanonical ensemble of energy E and volume V (cf. Fig. 3.16). The interaction energy between these two systems is assumed to be negligible, i.e. the Hamiltonian of the composite systems is given by

$$H = H_1 + H_2 + H_{12} \approx H_1 + H_2,$$

so that in addition

$$E = E_1 + E_2 + E_{12} \approx E_1 + E_2.$$

Let us take system 2 to be a heat bath and volume reservoir. Then we know from Def. 26 and Eq. 3.26 that the energies, volumes, particle numbers, and entropies involved must fulfill the following conditions:

$$\begin{aligned} E &\equiv (N_1 + N_2)e \approx N_2 e \\ V &\equiv (N_1 + N_2)v \approx N_2 v \\ S_2(E - E_1, V - V_1) &\approx S_2(E, V) - \frac{\partial S_2}{\partial E}(E, V) \cdot E_1 - \frac{\partial S_2}{\partial V}(E, V) \cdot V_1 \\ &\qquad + O(N_2^{-1}) \end{aligned} \tag{3.34}$$

In order to obtain the probability density ρ_1 of system 1, we may now again go through the four steps of the recipe advocated in Sect. 3.8:

step 1:

As stated above, there is nothing wrong with using additional variables. We may therefore consider the probability density of both the phase space coordinates and the two volumes V_1 and V_2. Furthermore, ρ_1 can be written as an integral of the probability density ρ of the combined systems over all possible states of system 2, so that we get

$$\rho_1(\boldsymbol{p}_1, \boldsymbol{q_1}, V_1) = \int \mathrm{d}\Gamma_2 \int \mathrm{d}\, V_2 \rho(\boldsymbol{p}_1, \boldsymbol{q}_2, \boldsymbol{p}_2, \boldsymbol{q}_2, V_1, V_2). \tag{3.35}$$

Here $\rho(\boldsymbol{p}_1, \boldsymbol{q}_1, \boldsymbol{p}_2, \boldsymbol{q}_2, V_1, V_2)$ is given by

$$\rho(\boldsymbol{p}_1, \boldsymbol{q}_1, \boldsymbol{p}_2, \boldsymbol{q}_2, V_1, V_2) = \rho(\boldsymbol{p}_1, \boldsymbol{q}_1, \boldsymbol{p}_2, \boldsymbol{q}_2) \cdot \delta(V - V_1 - V_2),$$

which follows from the fact that the probability must be normalized and that it must vanish for any V_1 and V_2 which do not add up to V.

For the probability density of system 1 we obtain

$$\begin{aligned} \rho(\boldsymbol{p}_1, \boldsymbol{q}_1, \boldsymbol{p}_2, \boldsymbol{q}_2, V_1, V_2) = \frac{1}{\omega(E, V)} \delta(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) - H_2(\boldsymbol{p}_2, \boldsymbol{q}_2, V_2)) \cdot \\ \cdot \delta(V - V_1 - V_2), \end{aligned} \tag{3.36}$$

where we have employed Eq. 3.12. The delta function $\delta(V - V_1 - V_2)$ had to be added in order to ensure that the total volume remains constant. Inserting Eq. 3.36 into Eq. 3.35 yields

$$\begin{aligned}
\rho_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) &= \frac{1}{\omega(E,V)} \int \mathrm{d}\Gamma_2 \int \mathrm{d}V_2\, \delta(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) - H_2(\boldsymbol{p}_2, \boldsymbol{q}_2, V_2)) \\
&\qquad \cdot \delta(V - V_1 - V_2) \\
&= \frac{1}{\omega(E,V)} \int \mathrm{d}\Gamma_2 \delta(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) - H_2(\boldsymbol{p}_2, \boldsymbol{q}_2, V - V_2) \\
&= \frac{1}{\omega(E,V)}\, \omega_2(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1), V - V_1) \\
&= \frac{1}{\omega(E,V)} \exp\left\{ \frac{k_\mathrm{B} \ln \omega_2(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1), V - V_1)}{k_\mathrm{B}} \right\} \\
&= \frac{1}{\omega(E,V)} \exp\left\{ \frac{S_2(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1), V - V_1)}{k_\mathrm{B}} \right\}.
\end{aligned}$$

step 2:
The entropy S_2 may be expanded by means of Eq. 3.34. One then gets

$$\begin{aligned}
\rho_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) &= \frac{1}{\omega(E,V)} \exp\left\{ \frac{1}{k_\mathrm{B}} \left[S_2(E,V) - \frac{\partial S_2}{\partial E}(E,V) \cdot H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) \right.\right. \\
&\qquad \left.\left. - \frac{\partial S_2}{\partial V}(E,V) \cdot V_1 + O(N_2^{-1}) \right] \right\} \\
&= \frac{e^{S_2(E,V)/k_\mathrm{B}}}{\omega(E,V)} \exp\left\{ \frac{1}{k_\mathrm{B}} \left[-\frac{\partial S_2}{\partial E}(E,V) \cdot H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) \right.\right. \\
&\qquad \left.\left. - \frac{\partial S_2}{\partial V}(E,V) \cdot V_1 + O(N_2^{-1}) \right] \right\},
\end{aligned}$$

and thus in the thermodynamic limit

$$\begin{aligned}
\rho_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) &= \frac{e^{S_2(E,V)/k_\mathrm{B}}}{\omega(E,V)} \exp\left\{ \frac{1}{k_\mathrm{B}} \left[-\frac{\partial S_2}{\partial E}(E,V) \cdot H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) \right.\right. \\
&\qquad \left.\left. - \frac{\partial S_2}{\partial V}(E,V) \cdot V_1 \right] \right\}.
\end{aligned} \tag{3.37}$$

step 3:
We know already that the derivative $\partial S/\partial E$ is just the temperature. This leaves us with the task to identify $\partial S/\partial V$ with some thermodynamic quantity.

However, as the dimension of the product of this derivative, temperature, and volume must be that of an energy, it seems (correctly) that pressure divided by temperature might be a good choice.

Definition 30 *The* thermodynamic pressure p *of a heat bath and volume reservoir with entropy S, energy E, and volume V is given by*

$$p \equiv T\frac{\partial S(E,V)}{\partial V}.$$

We were careful to include the word "thermodynamic" in the definition, as it remains to be shown that the quantity p thus introduced is consistent with the pressure known from classical mechanics. We'll worry about this in a moment. First, we use Def. 30 and Eq. 3.37 to obtain

$$\rho_1(\boldsymbol{p}_1, \boldsymbol{q_1}, V_1) = \frac{e^{S_2(E,V)/k_{\rm B}}}{\omega(E,V)} \exp\left\{-\frac{1}{k_{\rm B}T}\left[H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) + pV\right]\right\}.$$

The value of the constant $e^{S_2(E,v)/k_{\rm B}}/\omega(E,V)$ follows from the normalization condition $\int \mathrm{d}\Gamma_1 \int \mathrm{d}V_1\, \rho_1 = 1$. Accordingly we have derived the following theorem:

Theorem 23 *The probability density ρ of a canonical ensemble with variable volume and a continuum of states is given by*

$$\rho(\boldsymbol{p}, \boldsymbol{q}, V) = \frac{1}{\tilde{Z}(T,p)} e^{-(H(\boldsymbol{p},\boldsymbol{q},V)+pV)/k_{\rm B}T},$$

where T and p denote the temperature and pressure of the surrounding heat bath and volume reservoir, and where

$$\tilde{Z}(T,p) \equiv \int \mathrm{d}\Gamma \int \mathrm{d}V\, e^{-(H(\boldsymbol{p},\boldsymbol{q},V)+pV)/k_{\rm B}T}$$

is the canonical partition function with variable volume.

Again, everything remains valid if one considers a system with discrete rather than continuous states. We therefore get

Theorem 24 *Consider a canonical ensemble with variable volume possessing the microstates* (s_i, V), $i \in \mathbb{N}$. *Then the probability* $P(s_k, V)$ *of the state* (s_k, V) *has the value*

$$P(s_k, V) = \frac{1}{\tilde{Z}(T,p)} e^{-(H(s_k,V)+pV)/k_\mathrm{B}T},$$

where T *and* p *are the temperature and pressure of the surrounding heat bath and volume reservoir,* $H(s_k, V)$ *denotes the value of the Hamiltonian for* (s_k, V), *and*

$$\tilde{Z}(T,p) \equiv \sum_{i=1}^{\infty} \int \mathrm{d}V \, e^{-(H(s_k,V)+pV)/k_\mathrm{B}T}$$

is the canonical partition function with variable volume.

step 4:
We still have to figure out the meaning of the thermodynamic pressure. To this end, let us start with the ideal gas. Here, the entropy may be written as

$$S(E,V) = Nk_\mathrm{B} \ln V + \text{terms not depending on } V,$$

so that

$$p = T\frac{\partial S(E,V)}{\partial V} = \frac{Nk_\mathrm{B}}{V}.$$

But this is just the equation of state, if we identify p with ordinary pressure, as known from phenomenological thermodynamics. So at least in the case of ideal gases thermodynamic pressure is what one should have expected from its name.

Can this be said more generally, too? We start finding an answer to this question by means of generalizing our definitions of temperature and (thermodynamic) pressure.

Definition 31 *Consider a canonical ensemble with variable volume in equilibrium with a heat bath and volume reservoir. The temperature* T *and thermodynamic pressure* p *of this system are given by*

$$\frac{1}{T} \equiv \frac{\partial S(\bar{E},\bar{V})}{\partial E}$$

and

$$\frac{p}{T} \equiv \frac{\partial S(\bar{E},\bar{V})}{\partial V},$$

where $\bar{E}$ *and* $\bar{V}$ *are the most probable energy and volume, respectively.*

The reader should convince themselves that this is consistent with our previous definitions 29 and 30.

If two canonical ensembles with variable volume are in equilibrium with each other, their pressures and temperatures are the same. The proof of this assertion basically runs along the lines of that of Theorem 21: As (E_1, V_1) constitutes a random vector, Theorem 22 implies that their probability density ρ_1 is given by

$$\rho_1(\bar{E}_1, \tilde{V}_1) = \Big\langle \delta((\bar{E}_1, \tilde{V}_1) - (E_1, V_1)) \Big\rangle = \Big\langle \delta(\bar{E}_1 - E_1)\delta(\tilde{V}_1 - V_1) \Big\rangle$$

Using Eq. 3.36 we thus obtain (cf. the discussion preceding Eq. 3.32)

$$\begin{aligned}
\rho_1(\bar{E}_1, \tilde{V}_1) &= \int \mathrm{d}\Gamma_1 \mathrm{d}\Gamma_2 \mathrm{d}V_1 \mathrm{d}V_2 \rho(\boldsymbol{p}, \boldsymbol{q}, V_1, V_2)\delta(\bar{E}_1 - E_1)\delta(\tilde{V}_1 - V_1) \\
&= \int \frac{\mathrm{d}\Gamma_1 \mathrm{d}\Gamma_2 \mathrm{d}V_1 \mathrm{d}V_2}{\omega(E, V)} \delta(E - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1) - H_2(\boldsymbol{p}_2, \boldsymbol{q}_2, V_2)) \cdot \\
&\qquad\qquad \cdot \delta(V - V_1 - V_2) \cdot \delta(\bar{E}_1 - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, V_1))\delta(\tilde{V}_1 - V_1) \\
&= \int \mathrm{d}\Gamma_1 \mathrm{d}\Gamma_2 \, \frac{1}{\omega(E, V)} \delta(E - \bar{E}_1 - H_2(\boldsymbol{p}_2, \boldsymbol{q}_2, V - \tilde{V}_1)) \cdot \\
&\qquad\qquad\qquad \cdot \delta(\bar{E}_1 - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, \tilde{V}_1)) \\
&= \frac{1}{\omega(E, V)} \int \mathrm{d}\Gamma_1 \, \delta(\bar{E}_1 - H_1(\boldsymbol{p}_1, \boldsymbol{q}_1, \tilde{V}_1)) \cdot \\
&\qquad\qquad \cdot \int \mathrm{d}\Gamma_2 \delta(E - \bar{E}_1 - H_2(\boldsymbol{p}_2, \boldsymbol{q}_2, V - \tilde{V}_1)) \\
&= \frac{1}{\omega(E, V)} \omega_1(\bar{E}_1, \tilde{V}_1)\omega_2(E - \bar{E}_1, V - \tilde{V}_1).
\end{aligned}$$

For the most probable energy (volume) $\bar{E}_1$ ($\bar{V}_1$) of system 1 the derivative $\partial\rho_1/\partial E$ ($\partial\rho_1/\partial V$) must vanish, which yields the two relations

$$0 \overset{!}{=} \frac{\partial\omega_1(\bar{E}_1, V_1)}{\partial E}\omega_2(E-\bar{E}_1, V-V_1) - \omega_1(\bar{E}_1, V_1)\frac{\partial\omega_2(E-\bar{E}_1, V-V_1)}{\partial E} \tag{3.38}$$

$$0 \overset{!}{=} \frac{\partial\omega_1(E_1, \bar{V}_1)}{\partial V}\omega_2(E-E_1, V-\bar{V}_1) - \omega_1(E_1, \bar{V}_1)\frac{\partial\omega_2(E-E_1, V-\bar{V}_1)}{\partial V} \tag{3.39}$$

and thus

$$\begin{aligned} \frac{\partial S_1(\bar{E}_1, V_1)}{\partial E} &= k_{\mathrm{B}}\frac{\partial\omega_1(\bar{E}_1, V_1)/\partial E}{\omega_1(\bar{E}_1, V_1)} = k_{\mathrm{B}}\frac{\partial\omega_2(E-\bar{E}_1, V-V_1)/\partial E}{\omega_2(E-\bar{E}_1, V-V_1)} \\ &= \frac{\partial S_2(E-\bar{E}_1, V-V_1)}{\partial E} \end{aligned} \tag{3.40}$$

$$\frac{\partial S_1(E_1, \bar{V}_1)}{\partial V} = \frac{\partial S_2(E-E_1, V-\bar{V}_1)}{\partial V} \tag{3.41}$$

As energy and volume are conserved, we may conclude that $E-\bar{E}_1$ and $V-\bar{V}_1$ are the most probable energy and volume of system 2. In addition, we may set $V_1 = \bar{V}_1$ in Eq. 3.40 and $E_1 = \bar{E}_1$ in Eq. 3.41. Then looking at the definitions of temperature and thermodynamic pressure, we see that we have proved the above assertion:

Theorem 25 *If two canonical ensembles with variable volume are in equilibrium with each other, their temperatures and thermodynamic pressures are the same.*

Let us now turn to the example of an ideal gas in a rectangular box with a piston (of cross section A) moving freely in z direction, where there is an outer potential ϕ depending on z only (Fig. 3.17). We assume that the

Figure 3.17: Box with a piston containing an ideal gas.

outer medium can be treated as a heat bath and volume reservoir. Then the Hamiltonian of the system can be written as

$$H(\boldsymbol{p},\boldsymbol{q},V)=H(\boldsymbol{p},\boldsymbol{q},Az)=H_0(\boldsymbol{p},\boldsymbol{q})+\phi(z),$$

and the probability density of the ideal gas is given by

$$\rho(\boldsymbol{p},\boldsymbol{q},V)=\frac{1}{\tilde{Z}(T,p)}e^{-(H_0(\boldsymbol{p},\boldsymbol{q})+\phi(z)+pV)/k_{\mathrm{B}}T}.$$

The probability density of the volume is obtained by integrating over the whole phase space of the system, i.e.

$$\rho(V)=\int \mathrm{d}\Gamma\,\rho(\boldsymbol{p},\boldsymbol{q},V)=\underbrace{\frac{1}{\tilde{Z}(T,p)}\int \mathrm{d}\Gamma\,e^{-H_0(\boldsymbol{p},\boldsymbol{q})/k_{\mathrm{B}}T}}_{\text{independent of }V}\cdot e^{-(\phi(z)+pV)/k_{\mathrm{B}}T}.$$

For a sufficiently large system, we may assume that the volume is identical to the most probable volume (cf. Sect. 3.15), i.e. the derivative of $\rho(V)$ must vanish. Hence

$$0\overset{!}{=}\frac{\partial\rho(V)}{\partial V}=\frac{1}{\tilde{Z}(T,p)}\int \mathrm{d}\Gamma\,e^{-H_0(\boldsymbol{p},\boldsymbol{q})/k_{\mathrm{B}}T}\cdot\frac{1}{k_{\mathrm{B}}T}\left(-\frac{1}{A}\frac{\mathrm{d}\phi(z)}{\mathrm{d}z}-p\right)$$

and therefore

$$p=-\frac{1}{A}\frac{\mathrm{d}\phi(z)}{\mathrm{d}z}.$$

But $-\mathrm{d}\phi/\mathrm{d}z$ is just the force exerted on the piston, so that we obtain

$$\text{thermodynamic pressure}=p=\frac{\text{force}}{\text{area}}=\text{mechanical pressure}.$$

Of course, we have proved the equivalence of thermodynamic and mechanical pressure only for the special case of an ideal gas. However, we know from Theorem 25 that the thermodynamic pressure is the same for all systems in equilibrium with each other. Accordingly, the equivalence holds valid for all canonical ensembles with variable volume.

Theorem 26 *The thermodynamic pressure of a canonical ensemble with variable volume is equivalent to the mechanical pressure exerted on the system.*

Figure 3.18: Grand canonical ensembles. Whereas the volume is assumed to be constant, both energy and particle number may vary. The two systems together constitute a microcanonical ensemble.

$$\begin{array}{rcl} \multicolumn{3}{c}{\text{replacement rules}} \\ \hline V & \longrightarrow & N \\ \displaystyle\int \mathrm{d}V & \longrightarrow & \displaystyle\sum_{N=0}^{\infty} \\ \delta(V-V') & \longrightarrow & \delta_{N,N'} \end{array}$$

Table 3.1: Replacement rules for the transition from the case of a canonical ensemble with variable volume to that of a grand canonical ensemble.

3.13 Grand canonical ensemble

Equilibrium systems with constant volume V but variable energy E and particle number N are known as *grand canonical ensembles*. For their analysis one may again resort to the case of two such systems in contact with each other, which together form a microcanonical ensemble, as shown in Fig. 3.18. We confine ourselves to the case of one kind of particles; the generalization would be straightforward.

Fortunately, the treatment parallels that of canonical ensembles with variable volume, if one uses the three replacements stated in Table 3.1. Hence the only thing that remains to be done is to give the derivative $\partial S/\partial E$ an appropriate name:

Definition 32 *The* chemical potential μ *of a heat bath and particle reservoir of energy E and particle number N is defined as*

$$\mu \equiv -T\frac{\partial S(E,N)}{\partial N}.$$

The reader should note the negative sign in this definition. The analogues of Theorems 23 and 24 read:

Theorem 27 *The probability density* ρ *of a grand canonical ensemble with a continuum of states is given by*

$$\rho(\boldsymbol{p}, \boldsymbol{q}, N) = \frac{1}{\hat{Z}(T, \mu)} e^{-(H(\boldsymbol{p}, \boldsymbol{q}, N) - \mu N)/k_{\rm B}T},$$

where T *and* μ *denote the temperature and chemical potential of the surrounding heat bath and particle reservoir, and where*

$$\hat{Z}(T, \mu) \equiv \sum_{N=0}^{\infty} \int \mathrm{d}\Gamma \, e^{-(H(\boldsymbol{p}, \boldsymbol{q}, N) - \mu N)/k_{\rm B}T}$$

is the grand canonical partition function*.*

Theorem 28 *Consider a grand canonical ensemble with microstates* (s_i, N)*,* $i, N \in \mathbb{N}$*. Then the probability* $P(s_k, N)$ *of the state* (s_k, V) *has the value*

$$P(s_k, N) = \frac{1}{\hat{Z}(T, \mu)} e^{-(H(s_k, N) - \mu N)/k_{\rm B}T},$$

where T *and* μ *are the temperature and chemical potential of the surrounding heat bath and particle reservoir,* $H(s_k, V)$ *denotes the value of the Hamiltonian for* (s_k, N)*, and*

$$\hat{Z}(T, \mu) \equiv \sum_{N=0}^{\infty} \sum_{i=1}^{\infty} e^{-(H(s_i, N) - \mu N)/k_{\rm B}T}$$

is the grand canonical partition function*.*

As usual, we may generalize the concept of the chemical potential (and that of temperature):

Definition 33 *Consider a grand canonical ensemble in equilibrium with a heat bath and particle reservoir. The temperature* T *and chemical potential* μ *of this system are given by*

$$\frac{1}{T} \equiv \frac{\partial S(\bar{E}, \bar{N})}{\partial E}$$

and

$$\mu \equiv T \frac{\partial S(\bar{E}, \bar{N})}{\partial N},$$

where $\bar{E}$ *and* $\bar{N}$ *are the most probable energy and particle number, respectively.*

And of course the following theorem holds valid:

Theorem 29 *If two grand canonical systems are in equilibrium with each other, they have the same temperature and chemical potential.*

Concerning the meaning of the chemical potential, the reader is asked for some patience. It will be discussed at a later stage.

3.14 Quantum mechanical point of view

So far, we have discussed classical systems only. Fortunately, quantum systems behave quite similarly, the differences being of rather technical nature: Whereas in the classical case we deal with real-valued functions, we now have to turn to operators, so that it should come as no surprise that the various distributions have to be stated in form of operators:

Theorem 30 *The density matrices of the canonical, canonical with variable volume, and grand canonical distribution are given by*

$$\begin{aligned}\hat{\rho}_{\mathrm{c}} &= \frac{1}{Z(T,V,N)}e^{-\beta\hat{H}}\\ \hat{\rho}_{\mathrm{c,variable}\ V} &= \frac{1}{\tilde{Z}(T,p,N)}e^{-\beta(\hat{H}+p\hat{V})}\\ \hat{\rho}_{\mathrm{grand}} &= \frac{1}{\hat{Z}(T,V,\mu)}e^{-\beta(\hat{H}-\mu\hat{N})},\end{aligned}$$

where the partition functions are given by

$$\begin{aligned}Z(T,V,N) &\equiv \mathrm{tr}\,\hat{\rho}_{\mathrm{c}}\\ \tilde{Z}(T,p,N) &\equiv \mathrm{tr}\,\hat{\rho}_{\mathrm{c,variable}\ V}\\ \hat{Z}(T,V,\mu) &\equiv \mathrm{tr}\,\hat{\rho}_{\mathrm{grand}}\end{aligned}$$

and where temperature T, pressure p, and chemical potential μ are defined as in the classical case.

Thus at first sight, one would be inclined to say that quantum calculations run along the same lines as their classical counterparts. However one should bear in mind that quantum statistics involve non-commuting operators, which makes things considerably more difficult. We will discuss this important caveat in more detail later on.

Figure 3.19: Typical form of the probability density $\rho(E)$ of the energy E in a canonical ensemble, if it is assumed that the entropy increases with energy.

3.15 Fluctuations

When going through the discussion of the various kinds of ensembles, the reader perhaps has wondered whether there are any differences between them from a practical point of view. After all, from everyday experience we know (or at least seem to know, see below) that the inner energy, the volume, and the particle number of a macroscopic system don't change.

Paraphrasing, we might ask whether in macroscopic systems there are any appreciable fluctuations of the various thermodynamic variables introduced in the preceding sections. If this is the case, distinguishing between the various ensembles is important for observations. Otherwise, it is at least of no observational relevance.

To answer this question let us start with the canonical ensemble, where, as the reader will recall, only the energy E can fluctuate. Indeed, its probability distribution $\rho(E)$ can be computed from Theorems 19 and 22:

$$\begin{aligned}\rho(E) &= \langle\delta(E-H)\rangle = \int \mathrm{d}\Gamma\,\rho(\boldsymbol{p},\boldsymbol{q})\delta(E-h) = \int \mathrm{d}\Gamma\,\frac{e^{-\beta H}}{Z(T)}\delta(E-H)\\ &= \frac{e^{-\beta E}}{Z(T)}\int \mathrm{d}\Gamma\,\delta(E-H) = \frac{e^{-\beta E}}{Z(T)}\omega(E) = \frac{e^{-\beta E}}{Z(T)}e^{S/k_{\mathrm{B}}}\\ &= \frac{1}{Z(T)}e^{-\beta(E-TS)},\end{aligned}$$

where here and in the following we write $Z(T)$ rather than $Z(T,V,N)$. Fig. 3.19 shows the typical form of $\rho(E)$, based on the assumption that the entropy increases with the energy. The maximum of the curve (i.e. the most probable energy $\tilde{E}$) has already been obtained in Sect. 3.8; but let us nonetheless consider it again. Obviously $\tilde{E}$ corresponds to the zero of the derivative of the probability density, i.e

$$0 \overset{!}{=} \left.\frac{\partial\rho}{\partial E}\right|_{\tilde{E}} = -\beta\left(1 - T\left.\frac{\partial S}{\partial E}\right|_{\tilde{E}}\right)e^{-\beta(\tilde{E}-TS)},$$

so that we recover the condition

$$\left.\frac{\partial S}{\partial E}\right|_{\tilde{E}} = \frac{1}{T}$$

known from Def. 27. But of course we are mainly interested in the width of the distribution, that is, in its variance σ_H^2, given by

$$\sigma_H^2 = \left\langle H^2 \right\rangle - \left\langle H \right\rangle^2 . \tag{3.42}$$

We start with the calculation of the mean energy, which at first sight looks slightly cumbersome,

$$\left\langle H \right\rangle = \int \mathrm{d}\Gamma \, \rho_{\mathrm{canon}} H = \int \mathrm{d}\Gamma \, \frac{1}{Z(T)} e^{-\beta H} H \tag{3.43}$$

However, we may now apply a standard trick of statistical mechanics, which is of utmost importance and which will crop up over and over again: *We carry out the calculation in terms of the partition function.* In fact, it is easy to see that we may write Eq. 3.43 as

$$\left\langle H \right\rangle = -\frac{\partial}{\partial \beta} \ln Z(\beta), \tag{3.44}$$

where $Z(\beta)$ stands short for $Z(T = 1/k_{\mathrm{B}}\beta)$. This looks somewhat simpler than Eq. 3.43. And, who knows, perhaps forming the second derivative of $\ln Z$ yields some further insight. We just give it a try:

$$\begin{aligned}
\frac{\partial^2}{\partial \beta^2} \ln Z(\beta) &= -\frac{\partial}{\partial \beta} \left\langle H \right\rangle = -\frac{\partial}{\partial \beta} \int \mathrm{d}\Gamma \, \frac{1}{Z(\beta)} e^{-\beta H} H \\
&= \int \mathrm{d}\Gamma \, \frac{1}{Z(\beta)^2} \frac{\partial Z(\beta)}{\partial \beta} e^{-\beta H} H + \int \mathrm{d}\Gamma \frac{1}{Z(\beta)} H e^{-\beta H} H \\
&= -\left(-\frac{\partial}{\partial \beta} \ln Z(\beta) \right) \int \mathrm{d}\Gamma \frac{1}{Z(\beta)} e^{-\beta H} H + \int \frac{\mathrm{d}\Gamma}{Z(\beta)} e^{-\beta H} H^2 \\
&= -\left\langle H \right\rangle \left\langle H \right\rangle + \left\langle H^2 \right\rangle
\end{aligned} \tag{3.45}$$

But lo behold: Looking at Eq. 3.42 we see this implies that

$$\sigma_H^2 = -\frac{\partial}{\partial \beta} \left\langle H \right\rangle ,$$

which may also be written as

$$\sigma_H^2 = -\frac{\partial T}{\partial \beta}\frac{\partial \langle H \rangle}{\partial T} = k_{\mathrm{B}} T^2 \frac{\partial \langle H \rangle}{\partial T}.$$

Now we know from elementary thermodynamics that $\partial \langle H \rangle / \partial T$ is just the *specific heat* c. Hence we have proved:

Theorem 31 *The mean* $\langle H \rangle$ *of the energy in a canonical ensemble has the value*

$$\langle H \rangle = -\frac{\partial}{\partial \beta} \ln Z(\beta),$$

and the variance is proportional to the specific heat,

$$\sigma_H^2 = k_{\mathrm{B}} T^2 c.$$

In particular, if the specific heat is extensive, so is the variance of the energy. *Then in the thermodynamic limit, the energy fluctuations grow infinite.* So what is wrong about our above-mentioned intuitive notion of there being no energy fluctuations in a macroscopic system?

Nothing, really. When considering fluctuations in everyday life, we are concerned with *relative* fluctuations. And these do vanish in the thermodynamic limit. Indeed,

$$\frac{\sigma_H^2}{\langle H \rangle^2} \sim \frac{O(N)}{O(N^2)} \overset{N\to\infty}{\longrightarrow} 0.$$

As the reader should check for themselves, everything said so far basically remains valid for the case of a canonical ensemble with variable volume and that of a grand canonical ensemble as well. In addition it is straightforward to prove two theorems corresponding to Theorem 31. We begin with the one dealing with volume:

Theorem 32 *In a canonical ensemble with variable volume the mean* $\langle V \rangle$ *of the volume* V *has the value*

$$\langle V \rangle = -\frac{1}{\beta}\frac{\partial}{\partial p} \ln \tilde{Z}(\beta, p),$$

and the variance σ_V^2 *is given by*

$$\sigma_V^2 = k_{\mathrm{B}} T \langle V \rangle \kappa,$$

where the compressibility κ *is defined as*

$$\kappa \equiv -\frac{1}{\langle V\rangle}\frac{\partial \langle V\rangle}{\partial p}.$$

We leave the proof as a simple exercise to the reader. The case of particle number fluctuations in a grand canonical ensemble can be treated completely analogously:

Theorem 33 *In a grand canonical ensemble the mean $\langle N\rangle$ of the particle number N has the value*

$$\langle N\rangle = \frac{1}{\beta}\frac{\partial}{\partial\mu}\ln \hat{Z}(\beta,\mu),$$

and the variance σ_N^2 is given by

$$\sigma_N^2 = k_{\mathrm{B}}T\frac{\partial \langle N\rangle}{\partial\mu}.$$

Again the reader is asked to show this result. Needless to say, the fluctuations of both volume and particle number vanish in the thermodynamic limit:

$$\frac{\sigma_V^2}{\langle V\rangle^2} = \frac{O(N)}{O(N^2)} \overset{N\to\infty}{\longrightarrow} 0$$
$$\frac{\sigma_N^2}{\langle N\rangle^2} = \frac{O(N)}{O(N^2)} \overset{N\to\infty}{\longrightarrow} 0$$

Hence in the thermodynamic limit the fluctuations of the specific energy, specific volume, and particle density vanish. In that sense all the ensembles discussed so far are equivalent.

3.16 Macroscopic Equivalence of Ensembles

We have seen in the previous section that in the macroscopic limit all relative fluctuations vanish, so that the various ensembles are equivalent. However, this doesn't mean that they are the same; after all, they describe different physical situations: For a microcanonical ensemble, energy conservation

Figure 3.20: Approximation of $e^{Nf(x)}$ in the Laplace integral I_N.

holds valid, a canonical ensemble pertains to some system in a heat bath, a canonical ensemble with variable volume can be regarded as a system in a heat bath and volume reservoir, and last but not least a grand canonical ensemble is obtained if one places a system in a heat bath and particle reservoir.

Hence different ensembles are described by different functions known as their respective thermodynamic potentials. In the following we will introduce these potentials and show that they can be transformed into each other in a well-defined way. Hence we provide a mathematical backing to the equivalence of potentials.

3.16.1 A useful integral

Before turning to the thermodynamic potentials, we need to discuss an integral which at first sight may seem slightly pathological. It is given by

$$I_N \equiv \int_0^\infty \mathrm{d}x\, e^{Nf(x)}.$$

As the reader will undoubtedly believe, this integral will have no analytical solution, unless $f(x)$ happens to be rather simple. However, this changes in the limit of infinite N.

To see what is meant by this, let us assume that the (well-behaved) function $f(x)$ has exactly one maximum $\bar{x}$, but no minimum whatsoever. Then for large N only the vicinity around $\bar{x}$ contributes significantly to the value of I_N, as the exponential is much smaller everywhere else. We may therefore expand $f(x)$ to second order around $\bar{x}$ (cf. Fig. 3.20),

$$f(x) \approx f(\bar{x}) + (x-\bar{x})\underbrace{f'(\bar{x})}_{=0} + \frac{1}{2}(x-\bar{x})^2 f''(\bar{x}) = f(\bar{x}) + \frac{1}{2}(x-\bar{x})^2 f''(\bar{x}).$$

Here $f'(\bar{x})$ vanishes, because by definition the maximum of f happens to be at $\bar{x}$. Employing this approximation, we may now compute I_N quite easily:

$$\begin{aligned} I_N &= \int_0^\infty \mathrm{d}x\, e^{Nf(x)} \approx \int_0^\infty \mathrm{d}x\, e^{Nf(\bar{x})+\frac{N}{2}(x-\bar{x})^2 f''(\bar{x})} \\ &= e^{Nf(\bar{x})} \int_0^\infty \mathrm{d}x\, e^{-\frac{1}{2}Nf''(\bar{x})(x-\bar{x})^2} \approx e^{Nf(\bar{x})} \int_{-\infty}^\infty \mathrm{d}x\, e^{-\frac{1}{2}Nf''(\bar{x})(x-\bar{x})^2} \\ &= e^{Nf(\bar{x})} \sqrt{\frac{2\pi}{Nf''(\bar{x})}}. \end{aligned}$$

In the last step we have used the formula for Gaussian integrals (cf. Footnote 1 on page 43). Hence we get

$$\ln I_N \approx Nf(\bar{x}) + \frac{1}{2} \ln \frac{2\pi}{f''(\bar{x})} - \frac{1}{2} \ln N$$

and thus

$$\frac{\ln I_N}{N} = f(\bar{x}) + O\left(\frac{\ln N}{N}\right), \tag{3.46}$$

We may use this relation for computing Laplace transforms asymptotically. To this end, consider a function $g(x)$. The Laplace transform $\mathcal{L}[g]$ of g is given by

$$\mathcal{L}[g](s) \equiv \int_0^\infty \mathrm{d}x\, e^{-sx} g(x).$$

Assuming that g is strictly positive, we can rewrite this as

$$\mathcal{L}[g](s) = \int_0^\infty \mathrm{d}x\, e^{-sx+\ln g(x)},$$

so that if $\ln g(x) - sx$ is of the order $O(N)$ and has exactly one maximum $\bar{x} = \bar{x}(s)$ and no minimum in $\mathbb{R}_0^+$ we see from Eq. 3.46 that in the limit of large N, the Laplace transform may be expressed by means of

$$\frac{\ln \mathcal{L}[g](s)}{N} = \frac{1}{N}(\ln g(\bar{x}(s)) - s\bar{x}(s)) + O\left(\frac{\ln N}{N}\right), \tag{3.47}$$

which will prove extremely helpful in the next section.

3.16.2 Thermodynamic potentials

Now let us turn once again to the partition functions of the ensembles to be considered. We start by noting that the canonical partition function Z as a function of β can be written as

$$\begin{aligned} Z(\beta, V, N) &= \int \mathrm{d}\Gamma\, e^{-\beta H} = \int_0^\infty \mathrm{d}E \int \mathrm{d}\Gamma \delta(E - H(V,N)) e^{-\beta E} \\ &= \int_0^\infty \mathrm{d}E\, e^{-\beta E} \omega(E, V, N). \end{aligned} \tag{3.48}$$

so that it may be regarded as a Laplace transform of the microcanonical partition function. Similarly, the partition function of the canonical ensemble with variable volume constitutes a Laplace transform of the canonical partition function,

$$\begin{aligned} \tilde{Z}(\beta, p, N) &= \int_0^\infty \mathrm{d}V \int \mathrm{d}\Gamma\, e^{-\beta(H(V,N)+pV)} = \int_0^\infty \mathrm{d}V\, e^{-\beta pV} \int \mathrm{d}\Gamma e^{-\beta H(V,N)} \\ &= \int_0^\infty \mathrm{d}V\, e^{-\beta pV} Z(\beta, V, N). \end{aligned} \tag{3.49}$$

Finally, the grand canonical partition function can be expressed in the form

$$\begin{aligned} \hat{Z}(\beta, V, \mu) &= \sum_{N=0}^{\infty} \int \mathrm{d}\Gamma\, e^{-\beta(H-\mu N)} = \sum_{N=0}^{\infty} e^{\beta\mu N} \int \mathrm{d}\Gamma\, e^{-\beta H} \\ &= \sum_{N=0}^{\infty} e^{\beta\mu N} Z(\beta, V, N), \end{aligned} \tag{3.50}$$

where the last term may be viewed as a discrete Laplace transform. For this transform in the thermodynamic limit a relation similar to Eq. 3.47 holds valid: If the particle number is large, in the sum over N the only term contributing significantly is the one for which $\beta\mu N + \ln Z$ reaches its maximum. Hence denoting the corresponding particle number by $\bar{N}$ we get

$$\hat{Z}(\beta, V, \mu) = \sum_{N=0}^{\infty} e^{\beta\mu N + \ln Z(\beta, V, N)} \approx e^{\beta\mu\bar{N} + \ln Z(\beta, V, \bar{N})}$$

and thus

$$\ln \hat{Z}(\beta, V, \mu) = \beta\mu\bar{N} + \ln Z(\beta, V, \bar{N}). \tag{3.51}$$

As the reader should check themselves, βE, βpV, and the logarithms of the partition function are of the order $O(N)$. Therefore from Eqs. 3.48–3.50 and Eqs. 3.47, 3.51 we can deduce immediately that

$$\ln Z(\beta, V, N) = \ln \mathcal{L}\{\omega\}(\beta, V, N) = \ln \omega(\bar{E}, V, N) - \beta\bar{E} + O(\ln N) \tag{3.52}$$

$$\ln \tilde{Z}(\beta, p, N) = \ln \mathcal{L}\{Z\}(\beta, \beta p, N) = \ln Z(\beta, \bar{V}, N) - \beta p\bar{V} + O(\ln N) \tag{3.53}$$

$$\ln \hat{Z}(\beta, V, \mu) = \ln \mathcal{L}_{\text{discrete}}\{Z\}(\beta, V, \beta\mu) = \ln Z(\beta, V, \bar{N}) + \beta\mu\bar{N} + O(\ln N), \tag{3.54}$$

where $\bar{E}$, $\bar{V}$, and $\bar{N}$ are given by the conditions

$$\frac{\partial}{\partial E}(\ln \omega(E, V, N) - \beta E)\Big|_{E=\bar{E}} = 0 \iff \left.\frac{\partial \ln \omega}{\partial E}\right|_{E=\bar{E}} = \beta \tag{3.55}$$

$$\frac{\partial}{\partial V}(\ln Z(\beta, V, N) - \beta pV)\Big|_{V=\bar{V}} = 0 \iff \left.\frac{\partial \ln Z}{\partial V}\right|_{V=\bar{V}} = \beta p \tag{3.56}$$

$$\frac{\partial}{\partial N}(\ln Z(\beta, V, N) + \beta\mu N)\Big|_{N=\bar{N}} = 0 \iff \left.\frac{\partial \ln Z}{\partial N}\right|_{N=\bar{N}} = -\beta\mu. \tag{3.57}$$

From Eq. 3.52 and Theorem 31 we may conclude that

$$\begin{aligned}
\langle E\rangle \equiv \langle H\rangle &= -\frac{\partial}{\partial\beta}\ln Z(\beta, V, N) \\
&= -\frac{\partial}{\partial\beta}(\ln \omega(\bar{E}(\beta), V, N) - \beta\bar{E}(\beta) + O(\ln N)) \\
&= -\left.\frac{\partial \ln \omega}{\partial E}\right|_{E=\bar{E}(\beta)}\frac{\partial\bar{E}}{\partial\beta} + \bar{E}(\beta) + \beta\frac{\partial\bar{E}(\beta)}{\partial\beta} + O(\ln N) \\
&= -\beta\frac{\partial\bar{E}}{\partial\beta} + \bar{E}(\beta) + \beta\frac{\partial\bar{E}(\beta)}{\partial\beta} + O(\ln N),
\end{aligned}$$

where the last step follows from Eq. 3.55. Hence we get the relation

$$\langle E\rangle = \bar{E} + O(\ln N). \tag{3.58}$$

Completely analogously one may prove from Theorems 32, 33 and Eqs. 3.53, 3.54, 3.56, 3.57 the corresponding equations for the average volume and particle number,

$$\langle V\rangle = \bar{V} + O(\ln N) \tag{3.59}$$
$$\langle N\rangle = \bar{N} + O(\ln N). \tag{3.60}$$

As promised above, we may now introduce some thermodynamic potentials:

Definition 34 *In the following, the definitions of various* thermodynamic potentials *are given:*

$S(E,V,N) \equiv k_{\mathrm{B}} \ln \omega(E,V,N)$	*entropy*
$F(T,V,N) \equiv -k_{\mathrm{B}}T \ln Z(T,V,N)$	*(Helmholtz) free energy*
$G(T,p,N) \equiv -k_{\mathrm{B}}T \ln \tilde{Z}(T,p,N)$	*Gibbs free energy* or *free enthalpy*
$\Omega(T,V,\mu) \equiv -k_{\mathrm{B}}T\hat{Z}(T,V,\mu)$	*grand canonical potential*

Of course the definition of the entropy is just the one used so far; it is included for the sake of completeness only.

If one writes Eqs. 3.52–3.54 in terms of the thermodynamic potentials, pays attention to Eqs. 3.58–3.60, and omits any terms of the order $O(\ln N)$ (which is justified in the macroscopic limit, of course), one obtains a set of relations which are of utmost importance for thermodynamics, as we shall see in a later chapter.

Theorem 34 *In the macroscopic limit the following relations between thermodynamic potentials hold valid:*

$$\boxed{\begin{gathered} F(T,V,N) = \langle E\rangle(T,V,N) - TS(\langle E\rangle(T,V,N),V,N) \\ G(T,p,N) = F(T,\langle V\rangle(T,p,N),N) + p\,\langle V\rangle(T,p,N) \\ \Omega(T,V,\mu) = F(T,V,\langle N\rangle(T,V,\mu)) - \mu\,\langle N\rangle(T,V,\mu) \end{gathered}}$$

The reader should note two things: Firstly, in phenomenological thermodynamics one might use the equations of the preceding theorem as the definitions of the various thermodynamic potentials. Then evidently the content of Def. 34 has the status of a theorem which needs to be proved.

Secondly, it is tempting to simplify the relations between the thermodynamic potentials by means of omitting the brackets $\langle\rangle$ and the explicitly given arguments of the mean values, so that one obtains

$$\begin{aligned} F(T,V,N) &= E - TS && \text{(imprecise!)} \\ G(T,p,N) &= F(T,V,N) + pV && \text{(imprecise!)} \\ \Omega(T,N,\mu) &= F(T,V,N) - \mu N && \text{(imprecise!)} \end{aligned}$$

Although it may save on writing, using this form of the relations is not encouraged, at least not at the beginning, as it tends to obscure the variable dependencies involved. In case the reader (understandably) protests that the other form is more difficult to memorize (after all, one has to remember which quantities are mean values), here is a simple remedy: Remembering the set of rules

- *F corresponds to a canonical ensemble and thus depends on T, V, and N.*
- *G corresponds to a canonical ensemble with variable volume and thus depends on T, p, and N.*
- *Ω corresponds to a grand canonical ensemble and thus depends on T, V, and μ.*

is completely sufficient; any quantity appearing on the right hand side of one of the equations in Theorem 34 which is not mentioned in these rules must be a mean value (and depend on the quantities given in the respective rule).

Theorem 34 may be used to obtain various simple formulae which prove useful in thermodynamics. They are collected in the following theorem.

Theorem 35 *The thermodynamic potentials, pressure, volume, and chemical potential fulfill the following relations:*

$$\boxed{\begin{array}{lll}
\dfrac{\partial F}{\partial T} = -S & \dfrac{\partial F}{\partial V} = -p & \dfrac{\partial F}{\partial N} = \mu \\
\dfrac{\partial G}{\partial T} = -S & \dfrac{\partial G}{\partial p} = V & \dfrac{\partial G}{\partial N} = \mu \\
\dfrac{\partial \Omega}{\partial T} = -S & \dfrac{\partial \Omega}{\partial V} = -p & \dfrac{\partial \Omega}{\partial \mu} = -N
\end{array}}$$

Here, all volumes and particle numbers (apart from $\partial/\partial V$ and $\partial/\partial N$) are to be understood as mean values, so that for example F stands short for $F(T, \langle V\rangle, \langle N\rangle)$.

The reader should remember that temperature, pressure, and chemical potential are given by Defs. 29, 31 and 33.

Proof: Let us prove the relation $\partial F/\partial V = -p$. To this end, we write $\langle E\rangle \equiv \bar{E}$, $\langle V\rangle \equiv \bar{V}$, and $\langle N\rangle \equiv \bar{N}$, so that by virtue of Theorem 34 we get

$$\begin{aligned}
\frac{\partial F(T,\bar{V},\bar{N})}{\partial V} &= \frac{\partial}{\partial V}(\bar{E}(T,\bar{V},\bar{N}) - TS(\bar{E}(T,\bar{V},\bar{N}),\bar{V},\bar{N}) \\
&= \frac{\partial \bar{E}(T,\bar{V},\bar{N})}{\partial V} - T\frac{\partial S(\bar{E}(T,\bar{V},\bar{N}),\bar{V},\bar{N})}{\partial E}\frac{\partial \bar{E}(T,\bar{V},\bar{N})}{\partial V} \\
&\quad - T\frac{\partial S(\bar{E}(T,\bar{V},\bar{N}),\bar{V},\bar{N})}{\partial V},
\end{aligned}$$

and hence from Defs. 29 and 31 we see that

$$\begin{aligned}
\frac{\partial F(T,\bar{V},\bar{N})}{\partial V} &= \frac{\partial \bar{E}(T,\bar{V},\bar{N})}{\partial V} - T\cdot\frac{1}{T}\cdot\frac{\partial \bar{E}(T,\bar{V},\bar{N})}{\partial V} - T\cdot\frac{p(T,\bar{V},\bar{N})}{T} \\
&= -p(T,\bar{V},\bar{N}),
\end{aligned}$$

which is just the desired result. The other relations can be proved completely analogously; this is left as an exercise to the reader. *(QED)*

So far, we have put no emphasis on the mathematical structure underlying the concept of thermodynamic potentials. It is given by the Legendre transform, as we shall discover in the next but one subsection. We therefore now turn to briefly revising its main properties.

3.17 Legendre Transform

The reader may know Legendre transforms and their properties from a course on theoretical mechanics, and thus might be tempted to skip this section. Alas, as it seems, the definition of the Legendre transform varies in the literature, so that at least familiarizing oneself with the actual definition we use might be a good idea.

Having said this, let us proceed directly to the introduction of the desired transform:

Definition 35 *Let $f(x)$ be concave or convex, i.e. an at least twice differentiable function with $\mathrm{d}^2 f(x)/\mathrm{d}x^2 \neq 0$ for all x. Then the Legendre transform $\mathcal{L}[f]$ of f is defined as*

$$\mathcal{L}[f](x) = yf'(x) - f(x).$$

Now it can be seen plainly that this definition makes sense even if $f(x)$ is neither convex nor concave, i.e. if its second derivative has zeroes. So why bother about $\mathrm{d}^2 f(x)/\mathrm{d}x \neq 0$? The answer is that alternative definitions require this condition, which perhaps is made clear by the following theorem stating a useful alternative form of the Legendre transform.

Theorem 36 *Let f be a convex or concave function, and let $z \equiv f'(x)$. Then there exists a function g so that $x = g(z)$, and the Legendre transform may be written as*

$$\mathcal{L}[f](x) = \mathcal{L}[f](g(z)) = zg(z) - f(g(z)) \equiv \tilde{\mathcal{L}}[f](z).$$

Proof: There isn't much to prove, really. As f is convex or concave, its second derivative never vanishes. Hence f' can be inverted, and the function g must exist. Furthermore, as g is the inverse of f', we get $f'(g(z)) = z$ and therefore obtain

$$\mathcal{L}[f](x) = \mathcal{L}[f](g(z)) = g(z)f'(g(z)) - f(g(z)) = g(z)z - f(g(z)),$$

which is just the desired relation. *(QED)*

Incidentally, we may also represent the Legendre transform $\tilde{\mathcal{L}}[f]$ as an extremum, a fact which will prove useful in the next section:

Theorem 37 *Let f be a concave or convex function, and let $z = f'(x)$. Then the Legendre transform $\tilde{\mathcal{L}}[f]$ can be written as*

$$\tilde{\mathcal{L}}[f](z) = \underset{y}{\mathrm{extr}}(yz - f(y)),$$

where $\underset{y}{\mathrm{extr}}$ denotes the extremum with respect to y.

Proof: Obviously, the extremum of $yz - f(y)$ with respect to y follows from the condition

$$0 \overset{!}{=} \frac{\mathrm{d}}{\mathrm{d}y}(yz - f(y) = z - f'(y),$$

which in the terminology of Theorem 36 is equivalent to $y = g(z)$. Accordingly, we obtain

$$\underset{y}{\mathrm{extr}}(yz - f(y)) = g(z)z - f(g(z)),$$

and using Theorem 36 we arrive at the proposition. *(QED)*

It is natural to ask how the Legendre transform can be inverted, and the first guess might be that one simply has to apply the transform again. Indeed, this is nearly right – but only nearly, as the following theorem illustrates.

Theorem 38 *A convex or concave function f can be related to its Legendre transform by means of the relation*

$$f(x) = \mathcal{L}[\tilde{\mathcal{L}}[f]](f'(x))$$

Note that in a practical calculation a technical problem arises: If we don't know f, we don't know f' either, and a straightforward computation of the right hand side of Eq. 38 is impossible.

Proof: Set $z = f'(x)$ and $\phi(z) \equiv \tilde{\mathcal{L}}[f](z)$. Then

$$\phi(z) = \tilde{\mathcal{L}}[f](z) = zg(z) - f(g(z)),$$

where as usual g is the inverse of f'. Thus (together with $f'(g(z)) = z$)

$$\begin{aligned}
\mathcal{L}[\tilde{\mathcal{L}}[f]](f'(x)) &= \mathcal{L}[\phi](z) = z\phi'(z) - \phi(z) \\
&= z(g(z) + zg'(z) - f'(g(z))g'(z)) - (zg(z) - f(g(z)) \\
&= zg(z) + zg'(z) - zg'(z) - zg(z) + f(g(z)) \\
&= f(g(z)) = f(x),
\end{aligned}$$

which is the desired relation. *(QED)*

As a simple example, let us consider the sine in the interval $[0, \pi/2]$. Then $f(x) = \sin x$, $z = f'(x) = \cos x$, and therefore $g(z) = \arccos z$. Hence for the Legendre transform it follows that

$$\mathcal{L}[f](x) = xf'(x) - f(x) = x \cos x - \sin x$$

or, equivalently,

$$\tilde{\mathcal{L}}[f](z) = zg(z) - f(g(z)) = z \arccos z - \sin(\arccos z).$$

Furthermore we get

$$\begin{aligned}
\mathcal{L}[\tilde{\mathcal{L}}[f]](y) &= y\tilde{\mathcal{L}}[f]'(y) - \tilde{\mathcal{L}}[f](y) \\
&= y\left(\arccos y + y\left(-\frac{1}{\sqrt{1-y^2}}\right) - \left(-\frac{1}{\sqrt{1-y^2}}\right)\cos(\arccos y)\right) \\
&\qquad\qquad - (y \arccos y - \sin(\arccos y)) \\
&= \sin(\arccos y),
\end{aligned}$$

and choosing $y = f'(x) = \cos x$ we arrive at

$$\mathcal{L}[\tilde{\mathcal{L}}[f]](f'(x)) = \sin(\arccos(\cos x)) = \sin x = f(x),$$

as should have been expected from Theorem 38. For comparison, we also note that

$$\begin{aligned}
\mathcal{L}[\mathcal{L}[f]](x) &= x\mathcal{L}[f]'(x) - \mathcal{L}[f](x) \\
&= x(\cos x - x \sin x - \cos x) - (x \cos x - \sin x) \\
&= \sin x - x^2 \sin x - x \cos x \neq f(x),
\end{aligned}$$

i.e. the Legendre transform of the Legendre transform does *not* yield the original function.

Perhaps, looking at the rather mathematical treatment so far, the reader shrugs their shoulders, wondering what one can actually use the Legendre transform for. If so, the following theorem might prove helpful:

Theorem 39 *Consider a function $f(x, y)$ which, if viewed as a function of x only, is convex or concave. Let $u \equiv \partial f/\partial x$, $v \equiv \partial f/\partial y$, and denote the*

Legendre transform of f with respect to x by $\phi(u,y) \equiv \tilde{\mathcal{L}}_x[f](u,y)$. Then x and v can be expressed as a function of u and y in the following way:

$$\begin{aligned} x(u,y) &= \frac{\partial \phi(u,y)}{\partial u} \\ v(u,y) &= -\frac{\partial \phi(u,y)}{\partial y}. \end{aligned}$$

Proof: The differential of f may be written as

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y = u\mathrm{d}x + v\mathrm{d}y, \tag{3.61}$$

whereas due to

$$\phi(u,y) = ux(u,y) - f(x(u,y),y)$$

the differential of the Legendre transform is given by

$$\mathrm{d}\phi = x\mathrm{d}u + u\mathrm{d}x - \mathrm{d}f. \tag{3.62}$$

Inserting Eq. 3.61 into Eq. 3.62 yields

$$\mathrm{d}\phi = x\mathrm{d}u + u\mathrm{d}x - u\mathrm{d}x - v\mathrm{d}y = x\mathrm{d}u - v\mathrm{d}y. \tag{3.63}$$

On the other hand we know that

$$\mathrm{d}\phi = \frac{\partial \phi}{\partial u}\mathrm{d}u + \frac{\partial \phi}{\partial y}\mathrm{d}y. \tag{3.64}$$

By comparing Eqs. 3.63 and 3.64 we finally obtain the desired result. *(QED)*

In order to convince the reader that this theorem is useful in physics, we consider a single point particle in one dimension with position q, velocity $\dot{q}$, and a time-independent Lagrangian $L = L(q,\dot{q})$. The conjugated momentum p of this particle is defined as $p \equiv \partial L/\partial \dot{q}$. Then obviously the Hamiltonian

$$H(q,p) \equiv \dot{q}p - L(q,\dot{q}(q,p))$$

represents the Legendre transform of the Lagrangian with respect to the velocity, and thus we may conclude from Theorem 39 that

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \frac{\partial L}{\partial q} &= -\frac{\partial H}{\partial q} \end{aligned}$$

The first of these equations represents one of the canonical equations of classical mechanics. And if one inserts the Lagrange equation

$$\frac{\partial L}{\partial q} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} = \dot{p}$$

into the left hand side of the second equation, one obtains

$$\dot{p} = -\frac{\partial H}{\partial q},$$

i.e. the second of the canonical equations.

3.18 General reservoirs

Throughout the last sections we have discussed various reservoirs, introducing important thermodynamic quantities along the way. However, as the reader undoubtedly noticed, the discussion did not depend particularly on the specific quantities under consideration. Indeed, we may generalize the treatment of reservoirs in a straightforward way, and may thus elucidate the underlying mathematical structure.

To this end, let us consider an arbitrary system ("system 1") placed in a heat reservoir ("system 2") which at the same time is a reservoir of some quantity A (called a *generalized thermodynamic coordinate*). As usual, we assume that the interaction potential can be neglected and that the entropy S_2 of the reservoir fulfills the relation

$$S_2(E - H_1, A - A_1) \approx S_2(E, A) - \frac{\partial S_2(E, A)}{\partial E}E_1 - \frac{\partial S_2(E, A)}{\partial A}A_1, \qquad (3.65)$$

where H_1 is the Hamiltonian of system 1 and where the E and A refer to the total system composed of system 1 and 2, which may be considered as a microcanonical ensemble. Then everything said in Sect. 3.12 applies, if we replace the volume by the quantity A and the pressure by its respective counterpart. We thus get:

Definition 36 *Consider a canonical ensemble with a variable generalized thermodynamic coordinate A in equilibrium with a heat bath and reservoir of A. The temperature T and the conjuagate coordinate α of A of this system are given by*

$$\frac{1}{T} \equiv \frac{\partial S(\bar{E}, \bar{A})}{\partial E}$$

and

$$\alpha \equiv \frac{\partial S(\bar{E}, \bar{A})}{\partial A},$$

where $\bar{E}$ and $\bar{A}$ are the most probable energy and A, respectively.

Theorem 40 *The probability density ρ of a canonical ensemble with a variable thermodynamic generalized coordinate A and a continuum of states is given by*

$$\rho(\boldsymbol{p}, \boldsymbol{q}, A) = \frac{1}{\tilde{Z}(T, \alpha)} e^{-(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)/k_{\mathrm{B}}T},$$

where T and α denote the temperare and conjugate coordinate of A, respectively, and where

$$\tilde{Z}(T, \alpha) \equiv \int \mathrm{d}\Gamma \int \mathrm{d}A \, e^{-(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)/k_{\mathrm{B}}T}$$

is the canonical partition function with variable A*.*

Theorem 41 *If two canonical ensembles with variable generalized thermodynamic quantity A are in equilibrium with each other, their temperatures and conjugate coordinates are the same.*

Of course, there is a discrete analogue to Theorem 40. However, we leave it as an exercise for the reader to write this down (cf. Theorem 24) and content ourselves with discussing the continuous case.

Evidently, the canonical partition function with variable A can be written as

$$\begin{aligned}\tilde{Z}(T, \alpha) &= \int \mathrm{d}A\mathrm{d}\Gamma \, e^{-(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)/k_{\mathrm{B}}T} \\ &= \int \mathrm{d}A \, e^{-\alpha A/k_{\mathrm{B}}T} \int \mathrm{d}\Gamma e^{-H(\boldsymbol{p},\boldsymbol{q},A)/k_{\mathrm{B}}T} = \int \mathrm{d}A \, e^{-\beta\alpha A} Z(T, A),\end{aligned}$$

where Z denotes the ordinary canonical partition function. Hence $\tilde{Z}$ can be interpreted as a Laplace transform, and from Eq. 3.47 we can infer that

$$\ln \tilde{Z}(T, \alpha) = \ln Z(T, \bar{A}(T, \alpha)) - \beta\alpha\bar{A}(T, \alpha),$$

where $\bar{A}$ is chosen to maximize $\ln Z(T, A) - \beta\alpha A$. Accordingly we obtain

$$\ln \tilde{Z}(T, \alpha) = \max_{A}(\ln Z(T, A) - \beta\alpha A). \tag{3.66}$$

In order to proceed, we now introduce the generalized free energy and enthalpy.

Definition 37 *The generalized free energy $F(T,A)$ and generalized free enthalpy $G(T,\alpha)$ of a canonical ensemble with a variable generalized thermodynamic coordinate A are defined as*

$$F(T,A) \equiv -k_{\mathrm{B}}T \ln Z(T,A)$$

and

$$G(T,\alpha) \equiv -k_{\mathrm{B}}T \ln \tilde{Z}(T,\alpha),$$

respectively.

Inserting this definition into Eq. 3.66, we obtain

$$-G(T,\alpha) = \max_{A}(-F(T,A) - \alpha A)$$

or (employing Theorem 37)

$$-G(T,-\alpha) = \max_{A}(\alpha A - F(T,A)) = \tilde{\boldsymbol{\mathcal{L}}}_A[F](T,\alpha),$$

i.e. we may cum grano salis regard the generalized free enthalpy as the Legendre transform of the generalized free energy. Thus in view of Theorem 39 it should come as no surprise that relations such as those given in Theorem 35 exist. However, we don't comment on this any further.

What we do comment on, though, is the overall behavior of the free enthalpy: It is quite straightforward to show that it must be a concave function. To see this, we start by computing its first derivative with respect to α (cf. Theorem 40),

$$\begin{aligned}
\frac{\partial G(T,\alpha)}{\partial \alpha} &= -k_{\mathrm{B}}T \frac{\partial}{\partial \alpha} \ln \tilde{Z}(T,\alpha) = -k_{\mathrm{B}}T \frac{1}{\tilde{Z}(T,\alpha)} \frac{\partial \tilde{Z}(T,\alpha)}{\partial \alpha} \\
&= -k_{\mathrm{B}}T \frac{1}{\tilde{Z}(T,\alpha)} \frac{\partial}{\partial \alpha} \int \mathrm{d}\Gamma \int \mathrm{d}A \, e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)} \\
&= -k_{\mathrm{B}}T \frac{1}{\tilde{Z}(T,\alpha)} \int \mathrm{d}\Gamma \mathrm{d}A \, (-\beta A) e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)} \\
&= \int \mathrm{d}\Gamma \int \mathrm{d}A \, A \frac{1}{\tilde{Z}(T,\alpha)} e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)} = \langle A \rangle \, .
\end{aligned}$$

From this we may easily obtain the second derivative,

$$\begin{aligned}\frac{\partial^2 G(T,\alpha)}{\partial\alpha^2} &= \frac{\partial}{\partial\alpha}\left(\frac{1}{\tilde{Z}(T,\alpha)}\int \mathrm{d}\Gamma \int \mathrm{d}A\, A e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)}\right) \\ &= -\frac{1}{\tilde{Z}(T,\alpha)^2}\frac{\partial \tilde{Z}(T,\alpha)}{\partial\alpha}\int \mathrm{d}\Gamma \int \mathrm{d}A\, A e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)} \\ &\qquad + \frac{1}{\tilde{Z}(T,\alpha)}\int \mathrm{d}\Gamma \int \mathrm{d}A\, (-\beta) A^2 e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)} \\ &= \beta\left(\int \mathrm{d}\Gamma \int \mathrm{d}A\, A \frac{1}{\tilde{Z}(T,\alpha)} e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)}\right)^2 \\ &\qquad - \beta \int \mathrm{d}\Gamma \int \mathrm{d}A\, A^2 \frac{1}{\tilde{Z}(T,\alpha)} e^{-\beta(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)} \\ &= \beta(\langle A\rangle^2 - \langle A^2\rangle).\end{aligned}$$

But because of $\langle A\rangle^2 \leqslant \langle A^2\rangle$ this implies

$$\frac{\partial^2 G(T,\alpha)}{\partial\alpha^2} \leqslant 0,$$

so that $G(T,\alpha)$ is indeed concave. From this one may show in addition that the generalized free energy consitutes a convex function.

Finally, we note that for the case of structureless point particles the only extensive variables are energy, volume, and particle number. Such systems are called *simple fluids*.

3.19 Entropy and Information

We live, as it is said, in an information era. However, what is this supposed to mean? What is information, after all?

From a naive, everyday life point of view, we might be inclined to associate information with semantics: This script contains information, as the reader (hopefully) can gain knowledge from reading it. Accordingly, we would readily embrace the idea that the sentence *I've never seen snow in my life.* provides some information, whereas we arguably would reject the idea that this is true of an arbitrary accumulation of letters such as *Xtjymnol hcvywas Tdc ashjc.*

But from the point of view of information theory, there are problems with this notion: Firstly, the semantic content of something is difficult to quantify. Secondly and more importantly, information seen in this way depends on the beholder. For example, the statement *Cogito, ergo sum.* will mean nothing to you, unless you either know about philosophy or speak some Latin.

We therefore need a more robust definition of information, which doesn't depend on semantics at all. In order to find this, let us reconsider the sentence *I've never seen snow in my life.* If it was uttered by a local person from Central Africa living far away from any mountain, we wouldn't be surprised. If, on the other hand, a Swedish person said so, we would be set thinking and would wonder whether our conversational partner might perhaps be blind. In this case, it seems, the sentence contains more information.

The difference between these two cases is of course that it is extremely less likely that the event *A Swedish person has never seen snow.* takes place. It is thus tempting to conclude that information is a function of the probability. And as we wanted to discard all semantical meaning, we should expect (or define) that this function is independent of the specific sample space under consideration.

A further property of information can be obtained, if we consider two or more independent events. Here, the total information should be just the sum of the amounts of information contained in the individual events. In other words, information constitutes an additive quantity.

Changing a probability just a little should not result in drastic changes of the respective information. We therefore also conclude that information must be a continuous function.

In addition, information must be normalized in some way. We achieve this by demanding that for some arbitrary number $a > 1$ the information corresponding to the the probability $1/a$ is just 1.

Having said all this, we may now define information in a rigorous manner.

Definition 38 *Consider a discrete sample space* Ω *with some probability* P*. Then the* information $I_{element}(\omega)$ *of an element* $\omega \in \Omega$ *is given by*

$$I_{element}(\omega) = f(P(\omega)),$$

with a continuous function $f :]0,1] \longrightarrow \mathbb{R}$*, which is normalized by demanding that*

$$f\left(\frac{1}{a}\right) = 1$$

for an arbitrarily chosen $a > 1$. Information is additive, i.e. the total information of two elements (possibly belonging to different sample spaces) is the sum of the individual amounts of information.

Note that f is *not* defined for vanishing probabilities. Somewhat surprisingly, this rather general definition is sufficient to entirely fix the functional form of the function f, so that indeed information is uniquely defined:

Theorem 42 *Let Ω be a discrete sample space with some probability P. Then the information $I_{element}(\omega)$ of an element $\omega \in \Omega$ has the value*

$$I_{element}(\omega) = -\log_a P(\omega).$$

Proof: Consider two independent events $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$. Then we know from Def. 38 that the total information corresponding to these events is given by

$$I_{\text{element,total}}(\omega_1, \omega_2) = I_{\text{element}}(P_1(\omega_1)) + I_{\text{element}}(P_2(\omega_2)).$$

On the other hand, we may regard the Cartesian product $\omega_1 \times \omega_2$ as an event of the sample space $\Omega_1 \times \Omega_2$. As the two events are assumed to be independent, the probability of this event must have the value

$$P(\omega_1 \times \omega_2) = P_1(\omega_1) \cdot P_2(\omega_2),$$

so that we get

$$\begin{aligned} f(P_1(\omega_1) \cdot P_2(\omega_2)) &= f(P(\omega_1 \times \omega_2)) = I_{\text{element,total}} \\ &= I_{\text{element}}(P_1(\omega_1)) + I_{\text{element}}(P_2(\omega_2)) \\ &= f(P_1(\omega_1)) + f(P_2(\omega_2)). \end{aligned}$$

As the sample spaces and probabilities are completely arbitrary, this implies that

$$f(x \cdot y) = f(x) + f(y) \tag{3.67}$$

for all $x, y \in]0, 1]$. Furthermore we have from Def. 38 that

$$f\left(\frac{1}{a}\right) = 1 \tag{3.68}$$

But the only continuous function fulfilling Eqs. 3.67 and 3.68 is the negative of the logarithm to base a. (The reader should consult a textbook on calculus for a proof of this statement.) Hence we have obtained the desired result. *(QED)*

Normally one chooses $a = 2$, so that for some $\omega \in \Omega$ $I_{\text{element}}(\omega) = -\operatorname{lb} P(\omega)$. In this case the information is measured in *bits*, which is achieved by merely appending the word *bit* to the numerical value obtained from Theorem 42 for $a = 2$.

If we are interested in the information of a sample space rather than that of an individual element, we just have to take the average of the amounts of information corresponding to the elements of the sample space.

Definition 39 *The information I of a discrete sample space with probability P is the average of the amounts of information contained in its elements,*

$$I(\Omega) \equiv \langle I_{element}(\omega) \rangle_{\omega \in \Omega} = \sum_{\omega \in \Omega} P(\omega) I_{element}(\omega).$$

Of course, we may use Theorem 42 to state an explicit expression for this information:

Theorem 43 *The information I of some sample space Ω has the value*

$$I(\Omega) = -\langle \log_a P(\omega) \rangle_{\omega \in \Omega} = -\sum_{\omega \in \Omega} P(\omega) \log_a P(\omega).$$

Apart from the somewhat formal derivation of information outlined so far, there is a more intuitive approach. As an illustration consider a single pawn on a checkerboard. Here, the probability that it is located on some arbitrary field (C5, say) obviously is 1/64. Accordingly, the information of the sample space consisting of all the events "The pawn is placed on a specific field." is given by

$$I = -\langle \operatorname{lb} P(\text{The pawn is placed on a specific field.}) \rangle = -\left\langle \operatorname{lb} \frac{1}{64} \right\rangle = 6 \text{ bit}.$$

Now imagine that we had to find out the actual position by asking questions which can be answered by "yes" or "no" only. Then it would be a good idea to succesively bisect the field where the pawn might be. Table 3.2 gives an example of this procedure.

One can see readily that independent of the actual field occupied by the pawn, six questions are required for figuring out the position. Hence the number of questions is equivalent to the information as measured in bits.

This result may be generalized. To this end consider a sample space Ω the elements ω_i of which have probabilities of the form $P_i = 2^{-n_i}$ with $n_i \in \mathbb{N}$.

Question	Answer
Is the pawn located in the rectangle with corners A1 and D8?	yes
Is it located in the rectangle with corners A1 and D4?	no
Is it located in the rectangle with corners A5 and B8?	no
Is it located in the rectangle with corners C5 and D6?	yes
Is it located in the rectangle with corners C5 and C6?	yes
Is it located on the field C5?	yes
conclusion: The pawn is on field C5.	

Table 3.2: Figuring out the position of a pwan located on the field C5 of a checkerboard. With each question the remaining space is cut in half.

Figure 3.21: Generalized method for singling out an element ω of some sample space Ω. Omega is divided up into Ω_1 and Ω_2 so that $P(\Omega) = P(\Omega_2) = 1/2$. The Ω_i in which ω lies then is taken as the new sample space, and the process is repeated as long as necessary.

Assume that one of these elements has occured. Then the best strategy for figuring out this event by means of yes-and-no questions will be to bisect Ω into Ω_1 and Ω_2 so that $P(\Omega_1) = P(\Omega_2) = 1/2$, and to ask in which half the desired element lies. This half is taken as the new sample space (i.e. its probability is renormalized to 1), and the process can be repeated as long as necessary (cf. Fig. 3.21). As the reader may show for themselves, following this recipe, one needs $n_i = \ln P_i$ questions to single out ω_i. Thus the average number of questions to be asked is just $\sum_i P_i \ln P_i$, and again we obtain the result that the information is equal to the average number of questions necessary for singling out an element of the sample space under consideration.

Alas, this isn't true in general. As a simple counterexample we may take the sample space $\Omega = \{\omega_1, \omega_2\}$ consisting of two elements with probabilities $P_1 = 1/3$ and $P_2 = 2/3$. The information of Ω is given by

$$I(\Omega) = -P_1 \operatorname{lb} P_1 - P_2 \operatorname{lb} P_2 = -\frac{1}{3} \operatorname{lb} \frac{1}{3} - \frac{2}{3} \operatorname{lb} \frac{2}{3}.$$

On the other hand, the question *Has ω_1 occured?* is sufficient to single out any of the two elements of Ω. Hence in this case the average number of necessary questions (namely one) clearly is different from the amount of information contained in the sample space.

As our discussion so far hasn't involved any physics, it should be clear that it applies to both classical and quantum mechanical systems. Here the latter may be treated fairly easily:

Theorem 44 *Consider a macrostate $\hat{\rho}$ with discrete microstates $|\psi_i\rangle$ forming an orthonormal basis. Then the information contained in the macrostate has the value*

$$I(\hat{\rho}) = -\operatorname{tr}(\hat{\rho}\log_a \hat{\rho}).$$

Proof: Let P_i denote the probability of the microstate $|\psi_i\rangle$. Then we may express ρ and $\log_a \hat{\rho}$ as

$$\begin{aligned}\hat{\rho} &= \sum_i P_i \, |\psi_i\rangle \, \langle\psi_i| \\ \log_a \hat{\rho} &= \sum_i \log_a P_i \, |\psi_i\rangle \, \langle\psi_i| \, ,\end{aligned}$$

and due to the orthonormality of the $|\psi_i\rangle$ we get

$$\begin{aligned}\operatorname{tr}(\hat{\rho}\log_a \hat{\rho}) &= -\sum_i \langle\psi_i| \left(\sum_r P_r \, |\psi_r\rangle \, \langle\psi_r| \sum_s \log_a P_s \, |\psi_s\rangle \, \langle\psi_s| \right) |\psi_i\rangle \\ &= -\sum_i \sum_r \sum_s P_r \log_a P_s \, \langle\psi_i| \left(|\psi_r\rangle \underbrace{\langle\psi_r| \, \psi_s\rangle}_{\delta_{rs}} \langle\psi_s| \right) |\psi_i\rangle \\ &= -\sum_i \sum_r P_r \log_a P_r \underbrace{\langle\psi_i| \, \psi_r\rangle}_{\delta_{ir}} \underbrace{\langle\psi_r| \, \psi_i\rangle}_{\delta_{ri}} \\ &= -\sum_i P_i \log_a P_i.\end{aligned}$$

But the last term is just the definition of information, so that the proposition has been proved. *(QED)*

Let us now turn to the case of continuous sample spaces with some probability density ρ. At first sight, one might guess that here the concept of information may be introduced by merely forming the limit of the discrete case for infinitely many elements. To be more precise, this would mean that we have to consider the limit $-\lim_{\Delta x_i \to 0} \sum_i \rho_i \Delta x_i \log_a(\rho_i \Delta\Omega_i)$, where ρ_i is the probability density for some $x_i \in \Delta x_i$. This may look nice, but actually isn't. Indeed, without loss of generality we assume that all the Δx_i are of the same size Δx and rewrite the limit as

$$\begin{aligned} -\lim_{\Delta x_i \to 0} \sum_i \rho_i \Delta x_i \log_a(\rho_i \Delta x_i) &= -\lim_{\Delta x \to 0} \sum \rho_i \Delta x \log_a(\rho_i \Delta x) \\ &= -\lim_{\Delta x \to 0} \underbrace{\sum_i \rho_i \Delta x}_{1} \log_a \Delta x - \lim_{\Delta x \to 0} \sum_i \rho_i \Delta x \log_a \rho_i \\ &= -\lim_{\Delta x \to 0} \log_a \Delta x - \langle \log_a \rho \rangle \end{aligned}$$

Evidently, the first term goes towards infinity for $\Delta x \longrightarrow 0$, so that we get the result that a continuous sample space contains an infinite amount of information. This shouldn't come as a surprise really, as we have seen above that information is related to the the number of questions necessary for singling out an element of the sample space considered. However, as a continuous sample space has infinitely many elements, it requires infinitely many questions.

Of course, while consistent with everything said so far, a definition leading to an infinite amount of information obviously is of no avail. So what can be done?

The simplest solution is to just drop the diverging term and live happily without it. And fortunately enough, that is precisely what is done:

Definition 40 *Consider a continuous sample space Ω with some probability density ρ. Then the information $\langle I \rangle (\Omega)$ of Ω is given by*

$$\langle I \rangle (\Omega) \equiv \langle \log_a \rho \rangle ,$$

where a is the constant known from the definition of information for discrete sample spaces.

It should be kept in mind that in the case of systems in statistical mechanics the usual differential phase space volume (i.e. $\mathrm{d}\Gamma = \mathrm{d}^{3N}p\mathrm{d}^{3N}q/N!h^{3N}$) is used.

Remarkably, there is a strong connection between information theory and statistical mechanics, which is hinted at by the following definition.

Definition 41 *The* information entropy *of some system (not necessarily in a state of equilibrium) is defined as* k_B *times the information contained in this system for* $a = e$*. More explicitly, the information entropy of a classical system is given by*

$$\hat{S}[\rho] \equiv k_\mathrm{B} \left\langle I \right\rangle_{a=e} = -k_\mathrm{B} \left\langle \ln \rho \right\rangle ,$$

whereas that of a quantum mechanical system has the value

$$\hat{S}[\hat{\rho}] \equiv k_\mathrm{B} \left\langle I \right\rangle_{a=e} = -k_\mathrm{B} \left\langle \ln \hat{\rho} \right\rangle$$

Here ρ *and* $\hat{\rho}$ *denote the probability density and density matrix, respectively.*

Note that this is just a definition. So far, we don't know whether information entropy has anything to do with the thermostatic entropy known from Def. 3.23. However, the terminology used implies that there might be some relation, and this is precisely what the following theorem asserts.

Theorem 45 *For equilibrium systems the information entropy is equal to the thermostatic entropy.*

Before turning to the proof of this assertion, let us discuss what the term "thermostatic entropy of an equilibrium system" is supposed to mean. Of course, if the system happens to be a microcanonical ensemble, in principle the entropy is just k_B times the logarithm of the microcanonical partition function. Still, there is a small caveat: As $\langle \rho \rangle$ doesn't make sense for delta distributions, in the following we have to employ the probability density given by Eq. 3.9.

In the case of a canonical ensemble (possibly with some variable thermodynamic coordinate such as volume) the energy isn't fixed and thus, being a function of the energy, the entropy should vary. However, as the (relative) fluctuations are small, we may define the entropy as the respective value for the average energy (and thermodynamic coordinate).

Proof: We consider microcanonical ensembles and canonical ensembles with and without a variable continuous thermodynamic coordinate, confining ourselves to the classical case. Other equilibrium systems (such as systems including a variable discrete thermodynamic coordinate or quantum mechanical ones) can be treated analogously.

microcanonical ensemble
As stated above, we assume that the probability density is given by Eq. 3.9, i.e. by

$$\rho_{\mathrm{mc}}^{\Delta}(\boldsymbol{p},\boldsymbol{q})=\frac{1}{\omega_{\Delta}(E)}(\Theta(E-H(\boldsymbol{p},\boldsymbol{q}))-\Theta(E-\Delta-H(\boldsymbol{p},\boldsymbol{q}))).$$

Hence we obtain for the information (with $a=e$)

$$\begin{aligned}\langle I\rangle &= -\left\langle \rho_{\mathrm{mc}}^{\Delta}\right\rangle\\ &= -\int \mathrm{d}\Gamma\,\frac{1}{\omega_{\Delta}(E)}(\Theta(E-H(\boldsymbol{p},\boldsymbol{q}))-\Theta(E-\Delta-H(\boldsymbol{p},\boldsymbol{q})))\cdot\\ &\qquad\cdot\ln\left(\frac{1}{\omega_{\Delta}(E)}(\Theta(E-H(\boldsymbol{p},\boldsymbol{q}))-\Theta(E-\Delta-H(\boldsymbol{p},\boldsymbol{q})))\right)\\ &= \int_{E-\Delta\leqslant H(\boldsymbol{p},\boldsymbol{q})\leqslant E}\mathrm{d}\Gamma\,\frac{1}{\omega_{\Delta}(E)}\ln\frac{1}{\omega_{\Delta}(E)}=-\frac{\ln\omega_{\Delta}(E)}{\omega_{\Delta}(E)}\underbrace{\int_{E-\Delta\leqslant H(\boldsymbol{p},\boldsymbol{q})\leqslant E}\mathrm{d}\Gamma}_{\omega_{\Delta}(E)}\\ &= -\ln\omega_{\Delta}(E),\end{aligned}$$

so that the information entropy is equal to

$$\hat{S}[\rho_{\mathrm{mc}}^{\Delta}]=k_{\mathrm{B}}\langle I\rangle=-k_{\mathrm{B}}\ln\omega_{\Delta}(E),$$

which according to Eq. 3.23 is just the thermostatic entropy.
canonical ensemble
The canonical probability density is given by

$$\rho_{\mathrm{c}}(\boldsymbol{p},\boldsymbol{q})=\frac{1}{Z(T)}e^{-H(\boldsymbol{p},\boldsymbol{q})/k_{\mathrm{B}}T},$$

and thus we get for the information (with $a=e$)

$$\begin{aligned}\langle I\rangle &= -\langle\ln\rho_{\mathrm{c}}\rangle=-\left\langle\frac{1}{Z(T)}e^{-H(\boldsymbol{p},\boldsymbol{q})/k_{\mathrm{B}}T}\right\rangle\\ &= -\langle-\ln Z(T)-H(\boldsymbol{p},\boldsymbol{q})/k_{\mathrm{B}}T\rangle=\ln Z(T)+\langle H\rangle/k_{\mathrm{B}}T.\end{aligned}$$

Multiplying both sides by $k_{\mathrm{B}}T$ and using the free energy yields

$$T\hat{S}[\rho_{\mathrm{c}}]=k_{\mathrm{B}}T\langle I\rangle=-F(T)+\langle H\rangle.$$

Comparing this result with the first equation given in Theorem 34, one sees immediately that the information entropy must be identical to the thermostatic one.

canonical ensemble with variable thermodynamic coordinate A

Let α denote the conjugate coordinate of A (cf. Def. 36). Then the probability density is of the form

$$\rho_{\text{c,variable }A}(\boldsymbol{p},\boldsymbol{q},A) = \frac{1}{\tilde{Z}(T,\alpha)} e^{-(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)/k_{\mathrm{B}}T},$$

so that we obtain for the information (again assuming $a = e$)

$$\begin{aligned}\langle I\rangle &= -\left\langle \ln \rho_{\text{c,variable A}}\right\rangle = -\left\langle \ln \frac{1}{\tilde{Z}(T,\alpha)} e^{-(H(\boldsymbol{p},\boldsymbol{q},A)+\alpha A)/k_{\mathrm{B}}T}\right\rangle \\ &= -\left\langle -\ln \tilde{Z}(T,\alpha) - H(\boldsymbol{p},\boldsymbol{q},A)/k_{\mathrm{B}}T - \alpha A/k_{\mathrm{B}}T\right\rangle \\ &= \ln \tilde{Z}(T,\alpha) + \langle H\rangle/k_{\mathrm{B}}T + \alpha\langle A\rangle/k_{\mathrm{B}}T.\end{aligned}$$

Multiplying both sides by $k_{\mathrm{B}}T$ and employing the generalized free energy and enthalpy, we get

$$T\hat{S}[\rho_{\text{c,variable }A}] = k_{\mathrm{B}}T\langle I\rangle = -G(T,\alpha) + \langle H\rangle + \alpha\langle A\rangle.$$

Comparing with the first two equations of Theorem 34 (where we may without loss of generality replace V and p by A and α, respectively), we arrive at

$$\begin{aligned}\langle H\rangle - T\hat{S}[\rho_{\text{c,variable }A}] &= G(T,\alpha) - \alpha\langle A\rangle = F(T,\langle A\rangle) \\ &= \langle H\rangle - TS(\langle H\rangle,\langle V\rangle).\end{aligned}$$

Hence information entropy and thermostatic entropy are the same. *(QED)*

We have shown that information entropy is consistent with thermostatic entropy for quilibrium systems. However, it can be applied to any system whatsoever and thus can be regarded as an extension of the original definition of entropy.

3.20 Principle of Maximum Entropy

The concept of information entropy introduced in the previous section lends itself to establishing an upper bound for the entropy of an arbitrary (physical) system. However, in order to see why this should be the case, we first need to prove a fairly general relation known as the Gibbs inequality. We first state its version for ordinary function:

Theorem 46 (Gibbs inequality) *Let $\rho(x)$ and $\rho_1(x)$ be probability densities of a (possibly more-dimensional) random variable x. Then the relation*

$$\langle \ln \rho_1 \rangle_\rho \leqslant \langle \ln \rho \rangle_\rho$$

holds valid.

Proof: We confine ourselves to the case of a one-dimensional random variable; the more-dimensional case can be treated completely analogously.

Let $B \equiv \langle \ln \rho_1 \rangle_\rho - \langle \ln \rho \rangle_\rho$.Then

$$\begin{aligned} B = \langle \ln \rho_1 \rangle_\rho - \langle \ln \rho \rangle_\rho = \langle \ln \rho_1 - \ln \rho \rangle_\rho &= \int \mathrm{d}x \, (\ln \rho_1(x) - \ln \rho(x)) \\ &= \int \mathrm{d}x \, \rho(x) \ln \frac{\rho_1(x)}{\rho(x)}. \end{aligned}$$

There is a small technical subtlety: The integral runs over all real numbers. (We will see below why we have to ensure this.) However, $\rho(x)$ may vanish for some $x \in \mathbb{R}$, so that in form of $\rho_1(x)/\rho(x)$ we end up with a division by zero, which is of course not well-defined. Fortunately, we are saved from grievous trouble by the fact that $\rho(x)$ appears as a prefactor and that we take the logarithm of the dubious ratio. Hence the problematic term only appears as $\rho(x) \ln(\rho_1(x)/\rho(x))$, which vanishes for $\rho(x) \longrightarrow 0$.

The careful reader might object that there is even worse trouble lurking in the depths of vanishing probability densities: For a $x \in \mathbb{R}$ with $\rho_1(x) = 0 \neq \rho(x)$ we may get a divergence we cannot discuss away any longer. True. But this divergence doesn't matter, as the upper bound to be derived in a moment remains valid.

Thus we don't really have to bother about vanishing probability densities and may proceed with our proof: For any $y \in \mathbb{R}^+$ the relation

$$\ln y \leqslant y - 1$$

is fulfilled (cf. Fig. 3.22), and thus we get

Figure 3.22: $\ln x$ and $x-1$. One can see that the relation $\ln x \leqslant x-1$ holds valid for all $x \in \mathbb{R}^{+}$.

$$\begin{aligned} B &\leqslant \int \mathrm{d}\, x \rho(x) \left(\frac{\rho_1(x)}{\rho(x)} - 1\right) = \int \mathrm{d}x \left(\rho_1(x) - \rho(x)\right) \\ &= \int \mathrm{d}x\, \rho_1(x) - \int \mathrm{d}x\, \rho(x). \end{aligned}$$

As ρ and ρ_1 are normalized and as the integral runs over all real numbers (here is where we need this), we finally obtain

$$B \leqslant \underbrace{\int \mathrm{d}x\, \rho_1(x)}_{1} - \underbrace{\int \mathrm{d}x\, \rho(x)}_{1} = 1 - 1 = 0.$$

But $B \leqslant 0$ is equivalent to $\left\langle \ln \rho_1 \right\rangle_{\rho} \leqslant \left\langle \ln \rho \right\rangle_{\rho}$. *(QED)*

We also give an operator version of the Gibbs inequalty:

Theorem 47 (Gibbs inequalty, operator version) *Let $\hat{\rho}$ and $\hat{\rho}_1$ denote density matrices. Then the logarithms of $\hat{\rho}$ and $\hat{\rho}_1$ fulfill the inequalty*

$$\left\langle \ln \hat{\rho}_1 \right\rangle_{\rho} \leqslant \left\langle \ln \hat{\rho} \right\rangle_{\rho}.$$

Proof: The proof basically runs along the same lines as that of the classical version. We may write the density matrices as

$$\begin{aligned} \hat{\rho} &= \sum_n w_n \left|\psi_n\right\rangle \left\langle\psi_n\right| \\ \hat{\rho}_1 &= \sum_n \tilde{w}_n \left|\phi_n\right\rangle \left\langle\phi_n\right| \end{aligned}$$

and their logarithms as

$$\begin{aligned} \ln \hat{\rho} &= \sum_n \ln w_n \left|\psi_n\right\rangle \left\langle\psi_n\right| \\ \ln \hat{\rho}_1 &= \sum_n \ln \tilde{w}_n \left|\phi_n\right\rangle \left\langle\phi_n\right|. \end{aligned}$$

Here the w_n, $\tilde{w}_n$ are non-negative numbers, and both the $|\psi_n\rangle$ and the $|\phi_n\rangle$ may without loss of generality be assumed to form complete orthonormal sets. The trouble arising from vanishing w_n and $\tilde{w}_n$ can be discussed away as in the proof of the classical version.

Let us denote the term $\langle \ln \hat{\rho}_1 \rangle_\rho - \langle \ln \hat{\rho} \rangle_\rho$ by B. Then as the trace can be formed with any orthornormal set, we get

$$
\begin{aligned}
B &= \langle \ln \hat{\rho}_1 \rangle - \langle \ln \hat{\rho} \rangle_\rho = \langle \ln \hat{\rho}_1 - \ln \hat{\rho} \rangle_\rho = \operatorname{tr}(\hat{\rho}[\ln \hat{\rho}_1 - \ln \hat{\rho}]) \\
&= \sum_n \langle \psi_n | \, \hat{\rho}(\ln \hat{\rho}_1 - \ln \hat{\rho}) \, | \psi_n \rangle = \sum_n \sum_k w_k \underbrace{\langle \psi_n | \, \psi_k \rangle}_{\delta_{nk}} \langle \psi_k | \, (\ln \hat{\rho}_1 - \ln \hat{\rho}) \, | \psi_n \rangle \\
&= \sum_n w_n \langle \psi_n | \, (\ln \hat{\rho}_1 - \ln \hat{\rho}) \, | \psi_n \rangle \\
&= \sum_n w_n \langle \psi_n | \ln \hat{\rho}_1 | \psi_n \rangle - \sum_n \sum_r w_n \underbrace{\langle \psi_n | \, \psi_r \rangle}_{\delta_{nr}} \langle \psi_r | \ln w_r | \psi_n \rangle \\
&= \sum_n w_n \langle \psi_n | \, (\ln \hat{\rho}_1 - \ln w_n) \, | \psi_n \rangle = \sum_n w_n \langle \psi_n | \ln \frac{\hat{\rho}_1}{w_n} | \psi_n \rangle \\
&= \sum_n \sum_s w_n \ln \frac{\tilde{w}_s}{w_n} \langle \psi_n | \, \phi_s \rangle \langle \phi_s | \, \psi_n \rangle = \sum_n \sum_s | \langle \psi_n | \, \phi_s \rangle |^2 w_n \ln \frac{\tilde{w}_s}{w_n}.
\end{aligned}
$$

If we apply the relation $\ln x \leqslant x - 1$ known already from the proof of the classical version, we obtain

$$
\begin{aligned}
B &\leqslant \sum_n \sum_s | \langle \psi_n | \, \phi_s \rangle |^2 w_n \left(\frac{\tilde{w}_s}{w_n} - 1 \right) = \sum_n \sum_r | \langle \psi_n | \, \phi_s \rangle |^2 (\tilde{w}_s - w_n) \\
&= \sum_s \tilde{w}_s \underbrace{\sum_n | \langle \psi_n | \, \phi_s \rangle |^2}_{1} - \sum_n w_n \underbrace{\sum_s | \langle \psi_n | \, \phi_s \rangle |^2}_{1} = \sum_s \tilde{w}_s - \sum_n w_n \\
&= 1 - 1 = 0,
\end{aligned}
$$

where the last but one step follows from the fact that the w_n and $\tilde{w}_s$ are to be interpreted as probabilities, which have to add up to unity (cf. Sect. 2.2). Accordingly, we have shown that $B \leqslant 0$, which is equivalent to the desired result. *(QED)*

Now consider the density matrices $\hat{\rho}$ of some quantum mechanical system Q and $\hat{\rho}_{\text{mc}}$ of a microcanonical ensemble, and assume that both systems possess the same Hamiltonian $\hat{H}$ and that for both systems the energy is fixed to the same value E (up to a small uncertainty Δ). Then from the Gibbs inequality we may infer

$$\hat{S}[\hat{\rho}] = -k_{\mathrm{B}} \left\langle \ln \hat{\rho} \right\rangle_{\hat{\rho}} \leqslant -k_{\mathrm{B}} \left\langle \ln \hat{\rho}_{\mathrm{mc}} \right\rangle_{\hat{\rho}} . \tag{3.69}$$

As the microcanonical density matrix is given by

$$\hat{\rho}_{\mathrm{mc}} = \sum_{E-\Delta \leqslant E_n \leqslant E} \frac{1}{\omega_{\Delta}(E)} \left| E_n \right\rangle \left\langle E_n \right| ,$$

and as any orthonormal basis may be used for forming a trace, the right hand side may be expressed as

$$\begin{aligned} -k_{\mathrm{B}} \left\langle \ln \hat{\rho}_{\mathrm{mc}} \right\rangle_{\hat{\rho}} &= -k_{\mathrm{B}} \sum_{E_k} \left\langle E_k \right| \hat{\rho} \ln \hat{\rho}_{\mathrm{mc}} \left| E_k \right\rangle \\ &= -k_{\mathrm{B}} \sum_{E_k} \sum_{E-\Delta \leqslant E_n \leqslant E} \ln \frac{1}{\omega_{\Delta}(E)} \left\langle E_k \right| \hat{\rho} \left| E_n \right\rangle \underbrace{\left\langle E_n \middle| E_k \right\rangle}_{\delta_{nk}} \\ &= k_{\mathrm{B}} \ln \omega_{\Delta}(E) \sum_{E-\Delta \leqslant E_n \leqslant E} \left\langle E_n \right| \hat{\rho} \left| E_n \right\rangle . \end{aligned}$$

Due to the fact that both systems under consideration have the same Hamiltonian, not only for the microcanonical ensemble, but also for Q the vector $\left| E_n \right\rangle$ corresponds to the energy E_n. In particular, this implies that for any $E_r \notin [E - \Delta, E]$ $\hat{\rho} \left| E_r \right\rangle$ must vanish, as otherwise the energy of Q wouldn't be fixed after all. Hence we obtain

$$\begin{aligned} k_{\mathrm{B}} \left\langle \ln \hat{\rho}_{\mathrm{mc}} \right\rangle_{\hat{\rho}} &= k_{\mathrm{B}} \ln \omega_{\Delta}(E) \sum_{E-\Delta \leqslant E_n \leqslant E} \left\langle E_n \right| \hat{\rho} \left| E_n \right\rangle \\ &= k_{\mathrm{B}} \ln \omega_{\Delta}(E) \sum_{E_n} \left\langle E_n \right| \hat{\rho} \left| E_n \right\rangle = k_{\mathrm{B}} \ln \omega_{\Delta}(E) \underbrace{\operatorname{tr} \hat{\rho}}_{1} = k_{\mathrm{B}} \ln \omega_{\Delta}(E) \end{aligned}$$

But the last term is just the entropy of the microcanonical ensemble, and thus from Eq. 3.69 and Theorem 45 we obtain

$$\hat{S}[\hat{\rho}] \leqslant \hat{S}[\hat{\rho}_{\mathrm{mc}}] = S(E).$$

We see that the microcanonical ensemble has the largest possible information entropy for a given Hamiltonian and energy. As should be expected, this result holds valid for the classical case as well. The proof, which is completely analogous to the quantum mechanical one, is left as an exercise for the reader.

Let us now compare a canonical ensemble with an arbitrary macrostate having the same Hamiltonian and average energy. Focussing again on the quantum mechanical case, we can represent the arbitrary macrostate and the canonical ensemble by means of their density matrices $\hat{\rho}$ and

$$\hat{\rho}_{\mathrm{c}} = \frac{1}{Z(T)} e^{-\hat{H}/k_{\mathrm{B}}T},$$

respectively. The demand of equal average energies then translates into

$$\left\langle \hat{H} \right\rangle_{\hat{\rho}} = \left\langle \hat{H} \right\rangle_{\hat{\rho}_c}. \tag{3.70}$$

From the Gibbs inequality we get

$$\begin{aligned} \hat{S}[\hat{\rho}] &= -k_{\mathrm{B}} \left\langle \ln \hat{\rho} \right\rangle_{\hat{\rho}} \\ &\leqslant -k_{\mathrm{B}} \left\langle \ln \hat{\rho} c \right\rangle_{\hat{\rho}} = -k_{\mathrm{B}} \left\langle \ln \left(\frac{1}{Z(T)} e^{-\hat{H}/k_{\mathrm{B}}T} \right) \right\rangle_{\hat{\rho}} \\ &= -k_{\mathrm{B}} \ln \frac{1}{Z(T)} + k_{\mathrm{B}} \left\langle \hat{H} \right\rangle_{\hat{\rho}} / k_{\mathrm{B}}T. \end{aligned}$$

From this, Eq. 3.70, and Theorem 45 it follows that

$$\begin{aligned} \hat{S}[\hat{\rho}] \leqslant -k_{\mathrm{B}} \ln \frac{1}{Z(T)} + k_{\mathrm{B}} \left\langle \hat{H} \right\rangle_{\hat{\rho}_c} / k_{\mathrm{B}}T = -k_{\mathrm{B}} \left\langle \ln \left(\frac{1}{Z(T)} e^{-\hat{H}/k_{\mathrm{B}}T} \right) \right\rangle_{\hat{\rho}_c} \\ = -k_{\mathrm{B}} \left\langle \ln \hat{\rho}_{\mathrm{c}} \right\rangle_{\hat{\rho}_{\mathrm{c}}} = \hat{S}[\hat{\rho}_{\mathrm{c}}] = S(\bar{E}(T)). \end{aligned}$$

Again the ensemble entropy is the largest one compatible with the given Hamiltonian and average energy. The reader should check for themselves that replacing the operators by the corresponding ordinary functions in the previous lines yields the respective proof for the classical case.

Evidently, systems with constant average energy and some constant average thermodynamic coordinate A (with conjugate variable α) can be treated in the same manner. If we consider an arbitrary quantum mechanical macrostate with the density matrix $\hat{\rho}$ and a canonical ensemble with the same Hamiltonian $\hat{H}$, average energy, and average A, represented by its density matrix

$$\hat{\rho}_{\mathrm{c,variable}\ A} = \frac{1}{\tilde{Z}(T,\alpha)} e^{-(\hat{H}+\alpha\hat{A})/k_{\mathrm{B}}T}$$

the relations

$$\left\langle \hat{H} \right\rangle_{\hat{\rho}} = \left\langle \hat{H} \right\rangle_{\hat{\rho}_{\text{c, variable } A}}$$
$$\left\langle \hat{A} \right\rangle_{\hat{\rho}} = \left\langle \hat{A} \right\rangle_{\hat{\rho}_{\text{c, variable } A}}$$

must hold valid. Employing these and (as usual) the Gibbs inequality, and Theorem 45, we get

$$\begin{aligned}
\hat{S}[\hat{\rho}] &= -k_{\text{B}} \left\langle \ln \hat{\rho} \right\rangle_{\hat{\rho}} \\
&\leqslant -k_{\text{B}} \left\langle \ln \hat{\rho}_{\text{c, variable } A} \right\rangle_{\hat{\rho}} = -k_{\text{B}} \left\langle \ln \left(\frac{1}{\tilde{Z}(T,\alpha)} e^{-(\hat{H}+\alpha\hat{A})/k_{\text{B}}T} \right) \right\rangle_{\hat{\rho}} \\
&= -k_{\text{B}} \ln \frac{1}{\tilde{Z}(T,\alpha)} + k_{\text{B}} \left\langle \hat{H} \right\rangle_{\hat{\rho}} / k_{\text{B}}T + k_{\text{B}}\alpha \left\langle \hat{A} \right\rangle_{\hat{\rho}} / k_{\text{B}}T \\
&= -k_{\text{B}} \ln \frac{1}{\tilde{Z}(T,\alpha)} + k_{\text{B}} \left\langle \hat{H} \right\rangle_{\hat{\rho}_{\text{c, variable } A}} / k_{\text{B}}T + k_{\text{B}}\alpha \left\langle \hat{A} \right\rangle_{\hat{\rho}_{\text{c, variable } A}} / k_{\text{B}}T \\
&= -k_{\text{B}} \left\langle \ln \left(\frac{1}{\tilde{Z}(T,\alpha)} e^{-(\hat{H}+\alpha\hat{A})/k_{\text{B}}T} \right) \right\rangle_{\hat{\rho}_{\text{c, variable } A}} \\
&= -k_{\text{B}} \left\langle \hat{\rho}_{\text{c, variable } A} \right\rangle_{\hat{\rho}_{\text{c, variable } A}} = \hat{S}[\hat{\rho}_{\text{c, variable } A}] = S(\bar{E}(T,\alpha), \bar{A}(T,\alpha)).
\end{aligned}$$

Therefore (not surprisingly any longer) we arrive at the result that the entropy of the canonical ensemble with a variable thermodynamic coordinate is the largest entropy consistent with the conditions imposed. The respective proof for the classical case again is obtained by replacing the operators by ordinary functions.

We may summarize all the things said so far in a neat theorem known as *principle of maximum entropy*:

Theorem 48 (principle of maximum entropy) *The largest information entropy compatible with a given Hamiltonian and with some given values of thermodynamic quantities is that of the respective ensemble. More explicitly,*

fixed energy E:	$\hat{S}[\rho] \leqslant \hat{S}[\rho_{\text{mc}}] = S(E)$
fixed average energy $\bar{E}$:	$\hat{S}[\rho] \leqslant \hat{S}[\rho_{\text{c}}] = S(\bar{E}(T))$
fixed $\bar{E}$ and $\bar{A}$:	$\hat{S}[\rho] \leqslant \hat{S}[\rho_{c,\ variable\ A}] = S(\bar{E}(T,\alpha), \bar{A}(T,\alpha))$

In all these relations the probability densities ρ may be replaced by the corresponding density matrices $\hat{\rho}$.

In the case of the canonical ensemble, we may restate the principle of maximum entropy as a principle of minimum free energy $F(T)$, if we use

$$F[\rho] = \bar{E} - T\hat{S}[\rho].$$

We know this equation to be true for canonical ensembles, and for all other systems we may accept it as a definition. To prove the minimum principle we just have to employ the principle of maximum entropy:

$$F[\rho] = \bar{E} - T\hat{S}[\rho] \geqslant \bar{E} - T\hat{S}[\rho_{\mathrm{c}}] = F[\rho_c].$$

A completely analogous argument holds valid for quantum mechanical systems as well, as the reader might like to show for themselves. Hence the minimum free energy consistent with a given Hamiltonian and fixed average energy is the one of the canonical ensemble.

3.21 Information Entropy and Irreversibility

In the previous section we have seen that the information entropy becomes maximal for an ensemble system. This suggests the question whether perhaps there exists a time development of the entropy whichs leads to ever-increasing entropies until the maximum (i.e. the equilibrium state) is reached.

Hence, might it be that irreversibility can be explained by means of information entropy? This surely looks promising. It perhaps even looks reasonable. But it is wrong.

The sad truth is told by the following theorem:

Theorem 49 *The information entropy of a physical system with a time-independent Hamiltonian is constant in time.*

Proof: We give a quantum mechanical proof. Consider a time-dependent density matrix $\hat{\rho}(t)$ with a time-independent Hamiltonian $\hat{H}$. Then $\hat{\rho}(t)$ is given by

$$\hat{\rho}(t) = e^{\frac{i}{\hbar}\hat{H}t}\hat{\rho}(0)e^{-\frac{i}{\hbar}\hat{H}t},$$

which follows directly from

$$\hat{\rho}(t=0) = e^{\frac{i}{\hbar}\hat{H}\cdot 0}\hat{\rho}(0)e^{-\frac{i}{\hbar}\hat{H}\cdot 0} = 1 \cdot \hat{\rho} \cdot 1 = \hat{\rho}(0)$$

and from the fact that this $\hat{\rho}(t)$ fulfills the von Neumann equation (cf. Theorem 14),

$$\dot{\hat{\rho}} = \frac{i}{\hbar}\hat{H}e^{\frac{i}{\hbar}\hat{H}t}\hat{\rho}(0)e^{-\frac{i}{\hbar}\hat{H}t} - \frac{i}{\hbar}e^{\frac{i}{\hbar}\hat{H}t}\hat{\rho}(0)\hat{H}e^{-\frac{i}{\hbar}\hat{H}t} = \frac{i}{\hbar}\hat{H}\hat{\rho}(t) - \frac{i}{\hbar}\hat{\rho}(t)\hat{H} = \frac{i}{\hbar}[\hat{H},\hat{\rho}].$$

Accordingly, we may write down the information entropy as a function of time t,

$$\begin{aligned}\hat{S}[\hat{\rho}](t) &= \hat{S}[\hat{\rho}(t)] = -k_{\mathrm{B}}\left\langle \ln\hat{\rho}(t)\right\rangle = -k_{\mathrm{B}}\left\langle \ln\left(e^{\frac{i}{\hbar}\hat{H}t}\hat{\rho}(0)e^{-\frac{i}{\hbar}\hat{H}t}\right)\right\rangle \\ &= -k_{\mathrm{B}}\left\langle \frac{i}{\hbar}\hat{H}t + \ln\hat{\rho}(0) - \frac{i}{\hbar}\hat{H}t\right\rangle = -k_{\mathrm{B}}\left\langle \ln\hat{\rho}(0)\right\rangle = \hat{S}[\hat{\rho}(0)] = \hat{S}[\hat{\rho}](0),\end{aligned}$$

and we see that it doesn't depend on time. *(QED)*

Accordingly, information entropy cannot be used for studying irreversibility.

3.22 The Classical Ideal Gas in Equilibrium Ensembles

The Hamiltonian of a classical ideal gas of N identical non-interacting particles in three dimensions is given by

$$H(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m}.$$

We will now discuss this charmingly simple example for the various equilibrium ensembles we have encountered throughout this chapter. And lo behold, we will be able to recover the ideal gas law known so well from phenomenological thermodynamics.

microcanonical ensemble

We place the ideal gas in a box, so that energy E, volume V, and particle number N remain constant. Fortunately the somewhat tedious calculation yielding the specific entropy of this microcanonical ensemble has already been carried out in Sect. 3.7, where the formula

$$s(e,v) = k_{\mathrm{B}}\left\{\frac{5}{2} + \ln\left[v\left(\frac{me}{3\pi\hbar^2}\right)^{3/2}\right]\right\}$$

with $e = E/N$ and $v = V/N$ was obtained (cf. Eq. 3.24). Thus the entropy $S = Ns$ must be given by

$$S(E, V, N) = Nk_{\mathrm{B}} \ln \left[\frac{V}{N} \left(\frac{m}{3\pi\hbar^2} \frac{E}{N} \right)^{3/2} \right] + \frac{5}{2} Nk_{\mathrm{B}},$$

which is known as the *Sackur-Tetrode equation.*

canonical ensemble

Now let us put the box containing the ideal gas in thermal contact with a heat bath of temperature T. Then the canonical partition function $Z(T, V, N)$ is of the form

$$\begin{aligned} Z(T, V, N) &= \int \mathrm{d}\Gamma \, e^{-H(\boldsymbol{p}, \boldsymbol{q})} = \int \frac{\prod_{k=1}^{N} \mathrm{d}^3 p_k \, \mathrm{d}^3 q_k}{N! (2\pi\hbar)^{3N}} \exp\left(-\beta \sum_{i=1}^{N} \frac{\boldsymbol{p}_i}{2m} \right) \\ &= \frac{1}{N!} \int \frac{\mathrm{d}^{3N} q}{(2\pi\hbar)^{3N}} \int \mathrm{d}^{3N} p \, \exp\left(-\beta \frac{1}{2m} \sum_{r=1}^{3N} p_r^2 \right) \\ &= \frac{1}{N!} \frac{V^N}{(2\pi\hbar)^{3N}} \left(\int \mathrm{d}p \, e^{-\frac{p^2}{2mk_{\mathrm{B}}T}} \right)^{3N}. \end{aligned}$$

The remaining integral is a Gaussian one, the value of which is given by $\sqrt{2\pi m k_{\mathrm{B}} T}$ (cf. Footnote 1 on page 43). We therefore obtain

$$Z(T, V, N) = \frac{1}{N!} \frac{V^N}{(2\pi\hbar)^{3N}} \sqrt{2\pi m k_{\mathrm{B}} T}^{3N},$$

and inserting the *thermal de Broglie wavelength* λ defined by

$$\lambda \equiv \frac{h}{\sqrt{2\pi m k_{\mathrm{B}} T}}$$

(which can be regarded as the quantum mechanical wave length of a particle with energy $k_{\mathrm{B}} T$), it follows that

$$Z(T, V, N) = \frac{1}{N!} \left(\frac{V}{\lambda^3} \right)^N. \tag{3.71}$$

Now consider for the moment the classical Hamiltonian of N particles with an interaction independent of the momenta,

$$H(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + V(\boldsymbol{q}).$$

Here, as the reader should prove as an exercise, the canonical partition function becomes

$$\begin{aligned} Z(T,V,N) &= \frac{1}{N!(2\pi\hbar)^{3N}}\left[\prod_{i=1}^{3N}\int \mathrm{d}p_i\, e^{-\frac{p_i^2}{2mk_\mathrm{B}T}}\right]\int \mathrm{d}^{3N}q\, e^{\frac{-V(\boldsymbol{q})}{k_\mathrm{B}T}} \\ &= Z_\text{ideal gas}(T,V,N)\int \mathrm{d}^{3N}q\, e^{-V(\boldsymbol{q})/k_\mathrm{B}T}. \end{aligned}$$

The remaining spatial integral is known as a *configuration integral* and is denoted by Q_N, so that

$$Q_N \equiv \int \mathrm{d}^{3N}q\, e^{-V(\boldsymbol{q})/k_\mathrm{B}T}$$

and thus

$$Z(T,V,N) = Z_\text{ideal gas}(T,V,N)\cdot Q_N(T,V,N).$$

Hence in classical mechanics the momentum integral can be performed for all systems the interaction of which is momentum-independent.

Let us now return to the ideal gas. From Eq. 3.71 we can infer directly that

$$\begin{aligned} F(T,V,N) &= -k_\mathrm{B}T\ln Z(T,V,N) = -k_\mathrm{B}T\ln\left(\frac{1}{N!}\left(\frac{V}{\lambda^3}\right)^N\right) \\ &= k_\mathrm{B}T\ln N! - Nk_\mathrm{B}T\ln\frac{V}{\lambda^3}, \end{aligned}$$

and using the Stirling approximation $\ln N! \approx N\ln N - N$ we see that

$$\begin{aligned} F(T,V,N) &\approx Nk_\mathrm{B}T\ln N - Nk_\mathrm{B}T - Nk_\mathrm{B}T\ln\frac{V}{\lambda^3} \\ &= -Nk_\mathrm{B}T\left(1+\ln\frac{V}{N\lambda^3}\right) = Nk_\mathrm{B}T\left(1+\ln\frac{v}{\lambda^3}\right). \end{aligned}$$

Thus from Theorem 35 we may compute the pressure p,

$$p(T,\bar{V},\bar{N}) = -\frac{\partial F(T,\bar{V},\bar{N})}{\partial V} = \frac{\bar{N}k_\mathrm{B}T}{\bar{V}}.$$

Obviously we have recovered the ideal gas law.

canonical ensemble with variable volume:
The canonical partition function with variable volume $\tilde{Z}(T, p, N)$ can be obtained readily from the partition function $Z(T, V, N)$:

$$\begin{aligned}\tilde{Z}(T, p, N) &= \int \mathrm{d}V \int \mathrm{d}\Gamma\, e^{-\beta(H+pV)} = \int \mathrm{d}V\, e^{-\beta pV} \int \mathrm{d}\Gamma\, e^{-\beta H} \\ &= \int \mathrm{d}V\, e^{-\beta pV} Z(T, V, N) = \int \mathrm{d}V\, e^{-\beta pV} \frac{1}{N!} \left(\frac{V}{\lambda^3}\right)^N \\ &= \frac{1}{N! \lambda^{3N}} \int \mathrm{d}V\, V^N e^{-\beta pV}.\end{aligned}$$

By means of an N-fold partial integration it is straightforward to show that the volume integral has the value

$$\int \mathrm{d}V\, V^N e^{-\beta pV} = \frac{N!}{(\beta p)^N},$$

so that we get

$$\tilde{Z}(T, p, N) = \frac{1}{N! \lambda^{3N}} \frac{N!}{(\beta p)^N} = \frac{1}{p^N \beta^N \lambda^{3N}}.$$

Accordingly, the Gibbs free enthalpy is given by

$$\begin{aligned}G(T, p, N) &= -k_{\mathrm{B}} T \ln \tilde{Z}(T, p, N) = -k_{\mathrm{B}} T \ln \frac{1}{p^N \beta^N \lambda^{3N}} \\ &= N k_{\mathrm{B}} T (\ln p + \ln \beta + 3 \ln \lambda),\end{aligned}$$

and Theorem 35 implies

$$\bar{V}(T, p, \bar{N}) = \frac{\partial G(T, p, \bar{N})}{\partial p} = \frac{\bar{N} k_{\mathrm{B}} T}{p}.$$

Again we have arrived at the equation of state of the ideal gas.
grand canonical ensemble:
Similarly to the case of a canonical ensemble with variable volume, the grand canonical partition function $\hat{Z}(T, V, \mu)$ can be obtained from the canonical one:

$$\begin{aligned}\hat{Z}(T, V, \mu) &= \sum_{N=0}^{\infty} \int \mathrm{d}\Gamma\, e^{-\beta(H-\mu N)} = \sum_{N=0}^{\infty} e^{\beta\mu N} \int \mathrm{d}V\, e^{-\beta H} \\ &= \sum_{N=0}^{\infty} e^{\beta\mu N} Z(T, V, N) = \sum_{N=0}^{\infty} e^{\beta\mu N} \frac{1}{N!} \left(\frac{V}{\lambda^3}\right)^N.\end{aligned}$$

Employing the fugacity $z = e^{\beta\mu}$, we may rewrite $\hat{Z}(T, V, \mu)$ as

$$\hat{Z}(T, V, \mu) = \sum_{N=0}^{\infty} \frac{1}{N!} \left(\frac{zV}{\lambda^3} \right)^N .$$

The series on the right hand side constitutes the Taylor expansion of the exponential function, and hence we find that

$$\hat{Z}(T, V, \mu) = \exp\left(\frac{zV}{\lambda^3} \right) .$$

Thus the grand canonical potential is given by

$$\Omega = -k_{\mathrm{B}} T \ln \hat{Z}(T, V, \mu) = -k_{\mathrm{B}} T \frac{zV}{\lambda^3} = -k_{\mathrm{B}} T \frac{e^{\beta\mu} V}{\lambda^3},$$

so that we may conclude from Theorem 35 that

$$p(T, \bar{V}, \bar{N}) = -\frac{\partial \Omega(T, \bar{V}, \bar{N})}{\partial V} = k_{\mathrm{B}} T \frac{e^{\beta\mu}}{\lambda^3}$$

and

$$\bar{N}(T, \bar{V}, \mu) = -\frac{\partial \Omega(T, \bar{V}, N)}{\partial \mu} = k_{\mathrm{B}} T \frac{\beta e^{\beta\mu} \bar{V}}{\lambda^3} = \bar{V} \frac{e^{\beta\mu}}{\lambda^3}.$$

Comparing these two equations, we note that

$$\frac{p(T, \bar{V}, \bar{N})}{\bar{N}} = \frac{k_{\mathrm{B}} T}{\bar{V}},$$

so that – what a surprise – we regain the ideal gas law.

Having proved the equation of state of the ideal gas from several points of view, we might aswell state it as a theorem. After all, it is of utmost importance.

Theorem 50 *The equation of state of a gas consisting of classical ideal non-interacting particles is given by the* ideal gas law*,*

$$p(T, \bar{V}, \bar{N}) \bar{V} = \bar{N} k_{\mathrm{B}} T.$$

Note that both an average volume and average particle number are used in the ideal gas law.

Chapter 4

The Monte Carlo Method

In the previous chapter we have developed the statistical description of equilibrium systems, and we have seen that partition functions may be considered as the “Holy Grail” of statistical mechanics.

But while it is quite simple to state the recipe for obtaining the partition function (just compute the relevant integral or sum over phase space), it turns out to be quite difficult to implement it for arbitrary systems. As usual, analytic solutions are rare, and we normally have to resort to numerical methods.

In this chapter we shall therefore discuss at length a numerical method that might be called the workhorse for calculations in statistical mechanics – the Monte Carlo method. Alas, our treatment can hardly be called exhaustive, and the interested reader is kindly referred to books on Monte Carlo methods (such as [7]) for additional information.

4.1 Calculating partition functions is difficult

4.1.1 Non-interacting particles

As we will see in a moment, calculating partition functions can be a huge burden. However, let us begin our discussion with the one case where this is *not* true, namely systems of non-interacting particles.

To this end, consider such a system (taken to be classical and continuous) with N non-distinguishable particles, and denote its time-independent Hamiltonian by $H = H(\boldsymbol{p}, \boldsymbol{q})$. Then as the particles don’t interact with

Figure 4.1: Cubic lattice of spins pointing along a crystallographic axis, which is taken to lie in z direction.

each other, H must be the sum of the N single-particle Hamiltonians $h_i = h_i(\boldsymbol{p}_i, \boldsymbol{q}_i)$, i.e.

$$H(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{N} h_i(\boldsymbol{p}_i, \boldsymbol{q}_i).$$

Accordingly, the canonical partition function must be of the form

$$\begin{aligned} Z(T,V,N) &= \int \mathrm{d}\Gamma\, e^{-\beta H(\boldsymbol{p},\boldsymbol{q})} = \int \frac{\mathrm{d}^{3N}p\, \mathrm{d}^{3N}q}{N!(2\pi\hbar)^{3N}} \exp\left(-\beta \sum_{i=1}^{N} h_i(\boldsymbol{p}_i, \boldsymbol{q}_i)\right) \\ &= \frac{1}{N!} \prod_{i=1}^{N} \int \frac{\mathrm{d}^3 p_i\, \mathrm{d}^3 q_i}{(2\pi\hbar)^3} e^{-\beta h_i(\boldsymbol{p}_i, \boldsymbol{q}_i)} = \frac{1}{N!} \prod_{i=1}^{N} Z_i(T,V,N), \end{aligned}$$

where Z_i is the the one-particle partition function for the i-th particle. The result obtained looks even nicer for distinguishable particles, as here the prefactor $1/N!$ doesn't appear. At any rate, we have shown that the N-particle problem can be reduced to a set of one-particle problems, which of course is easier to solve.

If, however, we have to take some interaction into account, the exponential function doesn't factorize any longer, and that is precisely where the trouble starts. We now turn to a simple model where this actually is the case.

4.1.2 Ising Model

As an example for a system of interacting particles consider a cubic lattice of N atoms with spin 1/2 (such as, say, EuO). For simplicity, we assume that there exists a crystal anisotropy forcing all the spins to point along some crystallographic axis, which we may without loss of generality assume to be the z axis of the coordinate system used (cf. Fig. 4.1). If we can neglect any motion of the atoms (which we take for granted), the Hamiltonian $\hat{H}$ of the spin lattice just depends on the N spins. Thus, as all the spins point in z direction, $\hat{H}$ may be written as

$$\hat{H}(\hat{\boldsymbol{s}}_1, \ldots, \hat{\boldsymbol{s}}_N) = \hat{H}(\hat{s}_{1z}, \ldots, \hat{s}_{N,z}).$$

Figure 4.2: Nearest neighbors. The grey atoms are nearest neighbors of the black atom, the wite ones are not.

But we may treat all the atoms as distinguishable particles, and all the $\hat{s}_i$ commute. Hence the system is equivalent to a classical one with the Hamiltonian obtained from the quantum mechanical one by replacing the spin operators $\hat{s}_{i,z}$ by ordinary coordinates $s_{i,z}$. The allowed states are given by the set

$$M = \{s_{1z}, \ldots\ s_{N,z} \mid s_{i,z} \text{ eigenvalues of } \hat{s}_i\},$$

and as the eigenvalues of the spin operators are given by $s_{i,z} = \hbar\sigma_i/2$ with $\sigma_i \in \{-1, 1\}$ we get

$$M = \left\{ \frac{\hbar}{2}\sigma_1, \ldots, \frac{\hbar}{2}\sigma_N \;\middle|\; \sigma_i \in \{-1, 1\} \right\}. \tag{4.1}$$

Now assume we apply a homogeneous magnetic field $\boldsymbol{B}$ in z direction. Then the corresponding Hamiltonian H_{external} is of the form (cf. Sect. 3.9)

$$H_{\text{external}}(s_{1z}, \ldots, s_{N,z}) = -\sum_{i=1}^{N} \boldsymbol{\mu}_i \boldsymbol{B} = -\sum_{i=1}^{N} \frac{e\hbar}{2m_e}\sigma_i B \equiv -h\sum_{i=1}^{N}\sigma_i,$$

where $h \equiv e\hbar B/2m_e$ must not be confused with the Planck constant. Apart from this contribution to the Hamiltonian there must in addition be a contribution H_{internal} due to spin-spin interactions. This will in general be of rather short range, so that it is reasonable to *assume* that it is effective between nearest neighbors only (cf. Fig. 4.2). Hence we obtain

$$H_{\text{internal}}(s_{1z}, \ldots, s_{N,z}) = \sum_{\substack{\text{nearest} \\ \text{neighbors } i,j}} E(\sigma_i, \sigma_j).$$

Here $E(\sigma_i, \sigma_j)$ constitutes the interaction energy between the i-th and j-th spin. In order to maintain a concise notation, we define $< i, j >$ to denote all pairs of nearest neighbors, so that we can write

$$H_{\text{internal}}(s_{1z}, \ldots, s_{N,z}) = \sum_{<i,j>} E(\sigma_i, \sigma_j).$$

We take the interaction energies $E(\sigma_i, \sigma_j)$ to depend linearly on both σ_i and σ_j,

$$E(\sigma_i, \sigma_j) = -J\sigma_i\sigma_j = \begin{cases} -J & \text{for } \sigma_i = \sigma_j \\ +J & \text{for } \sigma_i = -\sigma_j \end{cases}$$

If the constant of proportionality J is positive, the system is ferromagnetic; if it is negative, we have antiferromagnetism. The total Hamiltonian H of the system now is just the sum of H_{internal} and H_{external},

$$H = H_{\text{internal}} + H_{\text{external}} = -J \sum_{<i,j>} \sigma_i\sigma_j - h\sum_{i=1}^{N} \sigma_i.$$

The model of spin lattices outlined so far is known as the *Ising model.* Its canonical partition function clearly is obtained by summing over all states,

$$Z(T, V, N) = \sum_{\sigma_1=\pm1} \dots \sum_{\sigma_N=\pm1} \exp\left(-\beta J \sum_{<i,j>} \sigma_i\sigma_j - \beta h \sum_{i=1}^{N} \sigma_i\right), \tag{4.2}$$

and similarly, the canonical average $\langle A \rangle_{\text{c}}$ of some thermodynamic coordinate $A = A(\sigma_1, \dots, \sigma_N)$ can be computed by means of

$$\langle A \rangle_c = \frac{1}{Z(T, V, N)} \sum_{\sigma_1=\pm1} \dots \sum_{\sigma_N=\pm1} A(\sigma_1, \dots, \sigma_N) \cdot \\ \cdot \exp\left(-\beta J \sum_{<i,j>} \sigma_i\sigma_j - \beta h \sum_{i=1}^{N} \sigma_i\right). \tag{4.3}$$

In general there exist no closed expressions for Z and $\langle A \rangle$. This doesn't seem to be a real problem though, as after all, both Eq. 4.2 and Eq. 4.3 involve a finite number of terms only. Hence, while surely not wanting to carry out the sums by hand, we can surely write a C program doing that job, start the calculation, go and have some lunch, come back, and enjoy the result, can't we?

Can we really? Let us have a somewhat closer look! Being modest physicists, we content ourselves with a three-dimensional lattice of $10 \times 10 \times 10 = 1000$ spins. (Not too much, isn't it?) Hence the sums $\sum_{\sigma_1=\pm1}, \dots, \sum_{\sigma_{1000}=\pm1}$

Figure 4.3: Simple numerical integration. The integrand $f(x)$ is replaced by a piecewise constant function.

lead to $2^{1000} \approx 10^{301}$ terms. If we assume that each of these terms can be computed in the Planck time, i.e. in 5.31×10^{-43} s,[1], it takes about

$$10^{301}\ \text{terms} \cdot 5.31 \times 10^{-43}\ \text{s/term} \approx 5 \times 10^{258}\ s$$

to complete the job. On the other hand, the age of the universe is roughly

$$1.5 \times 10^{10}\ \text{years} \cdot 3.15 \times 10^{7}\ \text{s/year} \approx 5 \times 10^{17}\ \text{s}.$$

Hence it takes about 10^{241} ages of the universe to carry out the calculation! *Computing partition functions by brute force is impossible.*

Accordingly, we have to think of a more appropriate (numerical) approach, and the Monte Carlo method will turn out to be a good choice. Before showing this, however, we will embark on a useful detour and discuss the numerical evaluation of integrals.

4.2 Estimating integrals

Integrals in statistical mechanics for obvious reasons tend to involve a large number of dimensions. Still, let us first focus on the somewhat simpler case of one-dimensional integration of a non-pathological function $f(x)$ from $x = a$ to $x = b$, i.e. on

$$I \equiv \int_a^b \mathrm{d}x\, f(x). \tag{4.4}$$

The most straightforward approach to solving such an integral is illustrated by Fig. 4.3: You divide the interval $[a, b]$ into N pieces of equal length $\Delta = (b - a)/N$, and replace $f(x)$ by a function which is piecewise constant on these subintervals.

To be more precise, let

$$x_i \equiv a + (i - 1)\Delta = a + (i - 1)\frac{b - a}{N}.$$

---

[1] As it seems, we aren't modest concerning the efficiency of our computer...

Then the integral is approximated by means of

$$\int_a^b \mathrm{d}x\, f(x) \approx \sum_{i=0}^{N-1} f(x_i)(x_{i+1}-x_i) = \sum_{i=0}^{N-1} f(x_i)\frac{b-a}{N}.$$

In order to estimate the error made in this approximation, we note that $f(x)$ can be Taylor expanded about the x_i,

$$f(x) = f(x_i) + f'(x_i)(x-x_i) + O((x-x_i)^2),$$

so that the integral can be written as

$$\begin{aligned}\int_a^b \mathrm{d}x\, f(x) &= \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \mathrm{d}x\, f(x)\\ &= \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \mathrm{d}x\, \left[f(x_i) + f'(x_i)(x-x_i) + O((x-x_i)^2)\right],\end{aligned}$$

and we obtain for the error ΔI that

$$\begin{aligned}\Delta I &= \int_a^b \mathrm{d}x\, f(x) - \sum_{i=0}^{N-1} f(x_i)(x_{i+1}-x_i)\\ &= \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \mathrm{d}x\, \left[f'(x_i)(x-x_i) + O((x-x_i)^2)\right].\end{aligned}$$

Assuming that $f(x)$ is sufficiently well-behaved, so that we may neglect the second-order term, we get

$$\begin{aligned}\Delta I &= \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \mathrm{d}x\, f'(x_i)(x-x_i) = \sum_{i=0}^{N-1} f'(x_i) \int_{x_i}^{x_{i+1}} \mathrm{d}x\, (x-x_i)\\ &= \sum_{i=0}^{N-1} f'(x_i)\cdot\frac{1}{2}(x_{i+1}-x_i)^2 = \sum_{i=0}^{N-1} f'(x_i)\cdot\frac{1}{2}\Delta^2\\ &= \sum_{i=0}^{N-1} f'(x_i)\cdot\frac{1}{2}\frac{(b-a)^2}{N^2} \sim O\left(N\cdot\frac{1}{N^2}\right).\end{aligned}$$

Hence the error made is of the order $O(N^{-1})$. Of course, there is a huge amount of variants of this naive version for estimating one-dimensional integrals. However, as the spirit of all these is basically the one outlined so far, the conclusions concerning the order of magnitude of error basically remains the same independent of the specific variant under consideration. Readers interested in learning more about the subtleties of this kind of numerical integration are kindly referred to [8] for more information on this subject.

There exists an alternative, more unorthodox approach for estimating the integral I given by Eq. 4.4. In order to see this, we rewrite I in the form

$$I = \int_a^b \mathrm{d}x\, f(x) = \int_{-\infty}^{\infty} \mathrm{d}x\, f(x)\Theta(x-a)\Theta(b-x)$$

and note that the probability distribution $p_{\mathrm{u}}(x)$ for a random variable X distributed uniformly in $[a,b]$ is given by

$$p_{\mathrm{u}}(x) = \frac{1}{b-a}\Theta(x-a)\Theta(b-x),$$

so that using Theorem 3 we obtain

$$I = (b-a)\int \mathrm{d}x\, p_{\mathrm{u}}(x) f(x) = (b-a)\left\langle f(X)\right\rangle .$$

Hence apart from a prefactor $b-a$, I can be regarded as the average of the integrand (provided its argument is distributed uniformly). Now consider N random variables X_i distributed uniformly in $[a,b]$. Then the $f(X_i)$ constitute N identically distributed random variables, and Theorem 11 implies that

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) = \left\langle f(X)\right\rangle + O\left(\frac{\sigma_{f(X)}}{\sqrt{N}}\right).$$

But $\sigma_{f(X)}$ doesn't depend on N, so that we finally get

$$I = (b-a)\left\langle f(X)\right\rangle = \frac{b-a}{N}\sum_{i=1}^{N} f(X_i) + O\left(\frac{1}{\sqrt{N}}\right).$$

In practical terms this means that we can estimate the value of I by generating N random numbers x_i distributed uniformly in $[a,b]$, compute the corresponding $f(x_i)$, and form their average multiplied by $b-a$, i.e.

$$I = \frac{b-a}{N}\sum_{i=1}^{N} f(x_i) + O\left(\frac{1}{\sqrt{N}}\right).$$

N	estimate for π
10	4.10049
100	3.09209
1000	3.17717
10000	3.17515
100000	3.14371
1000000	3.13951

Table 4.1: Estimates of π obtained by the C++ program shown in Fig. 4.4 for various numbers N of random numbers. Note that the results depend on the implementation of the random number generator and thus on the compiler and operating system used.

This method of employing random numbers in order to estimate integrals is known as *Monte Carlo integration*. Evidently, its convergence ($O(N^{-1/2})$) is worse than that of the "conventional method" outlined above ($O(N^{-1})$).

As a simple example of a Monte Calo integration we calculate the value of π by estimating the integral

$$I_\pi = \int_0^2 \mathrm{d}x\,\sqrt{4-x^2}.$$

A straightforward C++ program carrying out this computation is shown in Fig. 4.4; some typical results are stated in Table 4.1.

Let us now be slightly bolder and turn to the discussion of M-dimensional integrals

$$I \equiv \int_V \mathrm{d}^M x\, f(\boldsymbol{x}),$$

where for simplicity we assume that the volume V of integration is a cube of side length Δ, i.e. that the M variables $x^{(i)}$ are constrained by the relations

$$\begin{gathered}
a_1 \leqslant x^{(1)} \leqslant b_1 = a_1 + \Delta \\
a_2 \leqslant x^{(2)} \leqslant b_2 = a_2 + \Delta \\
\cdots \\
a_M \leqslant x^{(M)} \leqslant b_M = a_N + \Delta
\end{gathered}$$

```
#include <iostream.h>
#include <math.h>
#include <stdlib.h>

int main(int argc, char **argv)
  // Simple program for estimating pi by means of a
  // Monte Carlo integration of sqrt(4-x*x) from
  // a=0 to b=4.
  // This program must be called with the number of
  // random numbers to be used.
{
  long i;
  long N;                // amount of random numbers used
  double random_number;
  const double a=0;      // lower limit of integration
  const double b=4;      // upper limit of integration
  double average,pi;

  if (argc!=2) {
    cerr << "You have to supply exactly one argument.\n";
    exit(0);
  }
  N=atol(argv[1]);
  average=0;
  for (i=1;i<=N;i++) {   // computing N times the average
    random_number=a+(b-a)*drand48();
        // random number (uniformly distributed in [a,b])
    average+=sqrt(4-random_number*random_number);
  }
  average/=(double)N;
  pi=(b-a)*average;
  cout <<  N << " random numbers ---> pi=" << pi << endl;
}
```

Figure 4.4: C++ program for estimating π by means of a Monte Carlo integration of $\sqrt{4-x^2}$ from 0 to 4.

We now divide the volume of integration into N cubes of equal volume. Then evidently the volume of such a small cube is given by Δ^M/N, so that its side length δ has the value

$$\delta = \left(\frac{\Delta^M}{N}\right)^{1/M} .$$

Similarly to the one-dimensional case, we may expand the integrand $f(\boldsymbol{x})$ about a corner $\boldsymbol{x}_{\mathrm{c},k}$ of the k-th cube V_k,

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_{\mathrm{c},k}) + (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c},k}) \cdot \nabla \boldsymbol{f}(\boldsymbol{x}_{\mathrm{c},k}) + O((\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c},k})^2)$$

and thus may approximate the integral I as

$$I = \sum_{k=1}^{N} \int_{V_k} \mathrm{d}^M x \left[f(\boldsymbol{x}_{\mathrm{c},k}) + (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c},k}) \cdot \nabla f(\boldsymbol{x}_{\mathrm{c},k}) + O((\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c},k})^2) \right] .$$

Assuming that $f(x)$ is sufficiently well-behaved, we can neglect the second-order term. Hence we see that we may estimate the integral by means of

$$\begin{aligned} I = \int_V \mathrm{d}^M x\, f(\boldsymbol{x}) &\approx \sum_{k=1}^{N} \int_{V_k} \mathrm{d}^M x\, f(\boldsymbol{x}_{\mathrm{c},k}) = \sum_{k=1}^{N} f(\boldsymbol{x}_{\mathrm{c},k}) \int_{V_k} \mathrm{d}^M x \\ &= \sum_{k=1}^{N} f(\boldsymbol{x}_{\mathrm{c},k}) V_k = \sum_{k=1}^{N} f(\boldsymbol{x}_{\mathrm{c},k}) \frac{\Delta^M}{N}. \end{aligned}$$

The imprecision ΔI of this approximation can be expressed as

$$\begin{aligned} \Delta I &= \sum_{k=1}^{N} \int_{V_k} \mathrm{d}^M x\, (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c},k}) \cdot \nabla f(\boldsymbol{x}_{\mathrm{c},k}) = \sum_{k=1}^{N} \nabla f(\boldsymbol{x}_{\mathrm{c},k}) \cdot \int_{V_k} \mathrm{d}^M x\, (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c},k}) \\ &= \sum_{k=1}^{N} \nabla f(\boldsymbol{x}_{\mathrm{c},k}) \cdot \begin{pmatrix} \delta^{M+1}/2 \\ \vdots \\ \delta^{M+1}/2 \end{pmatrix} = \frac{1}{2} \left(\frac{\Delta^M}{N}\right)^{(M+1)/M} \sum_{k=1}^{N} \nabla f(\boldsymbol{x}_{\mathrm{c},k}) \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &\sim O\left(\frac{1}{N^{1+1/M}} \cdot N\right) \sim O\left(N^{-1/M}\right) . \end{aligned}$$

Accordingly, the convergence of this method deteriorates with an increasing number of dimensions M, and in the limit of infinitely many dimensions, there is no convergence at all.

Things look utterly different for Monte Carlo integration: Here we consider N M-dimensional random vectors $\boldsymbol{X}_k$ distributed uniformly in the integration volume V (again taken to be a cube with side length Δ). Then the probability density p_{u} of the $\boldsymbol{X}_k$ can be written as

$$p_{\mathrm{u}}(\boldsymbol{x}) = \frac{1}{\Delta^M} \prod_{i=1}^{M} \Theta(x_i - a_i)\Theta(b_i - x_i),$$

and from Theorem 4 we obtain

$$\begin{aligned} I \equiv \int_V \mathrm{d}^M x\, f(x) &= \int \mathrm{d}^M x\, f(x) \prod_{i=1}^{M} \Theta(x_i - a_i)\Theta(b_i - x_i) \\ &= \Delta^M \int \mathrm{d}^M x\, p_{\mathrm{u}}(x) f(x) = \Delta^M \left\langle f(\boldsymbol{X}) \right\rangle . \end{aligned}$$

On the other hand, as the $f(\boldsymbol{X}_k)$ constitute N independent, but identically distributed random variables, we know from Theorem 11 that

$$\langle f(\boldsymbol{X}) \rangle = \sum_{k=1}^{N} f(\boldsymbol{X}_i) + O\left(\frac{1}{\sqrt{N}}\right),$$

As in the one-dimensional case, we get the result that for N random numbers distributed uniformly in the integration volume the relation

$$I = \Delta^M \left\langle f(\boldsymbol{X}) \right\rangle = \frac{\Delta^M}{N} \sum_{k=1}^{N} f(x_i) + O\left(\frac{1}{\sqrt{N}}\right),$$

which can be used for estimating the integral. Obviously, the convergence ($O(N^{-1/2})$) does not depend on M, so that for high dimensions Monte Carlo integration clearly surpasses the conventional method for estimating integrals.

Concerning the naive version of conventional numerical integration discussed above, Monte Carlo integration becomes more effective as soon as $-1/M > -1/2$, i.e. if $M > 2$. In practice, conventional algorithms may compete slightly longer; here the turnover occurs for $M \gtrsim 3 \ldots 4$. At any rate, the high-dimensional integrals occuring in statistical mechanics should be tackled by means of Monte Carlo integration.

4.3 Importance sampling

The simple Monte Carlo method outlined in the previous section involved generating random numbers according to a uniform distribution. This is generally known as *simple sampling.* While it has been declared the method of choice for high-dimensional integrals, without a serious modification it is doomed to fail for many integrals, in particular if the integrals happen to be relevant for statistical mechanics.

Perhaps the reader is surprised about this statement. After all, haven't we *proved* that Monte Carlo integration provides a reliable estimate? Yes and no. While we did indeed show that the imprecision will vanish as $CN^{-1/2}$, we said nothing whatsoever concerning the constant of proportionality C in this expression. And it is precisely this constant which may cause quite a lot of trouble.

In order to fully appreciate the problem, consider the innocent-looking function

$$f(x) \equiv \begin{cases} 10^{10} & (\frac{1}{2} \leqslant x \leqslant \frac{1}{2} + 10^{-10}) \\ 0 & (\text{otherwise}) \end{cases} .$$

Clearly the integral I of f from 0 to 2 has the value

$$I \equiv \int_0^2 \mathrm{d}x\, f(x) = 1.$$

Furthermore it is straightforward to adopt the program given in Fig. 4.4 in order to compute I; one just has to replace the line

```
average+=sqrt(4-random_number*random_number);
```

by

```
average+=random_number<0.5 ? 0 :
            (random_number<0.5+1e-10 ? 1e10 : 0);
```

(Ask your local C guru, if you don't about the meaning of the ternary operator `?`.) Running this modified program (try it!), however, turns out to be depressing: For all the numbers of random numbers from 1 to, say, 10^7 you get (in all likelihood) 0 as the estimate of I.

The reason for this can be seen fairly easily. If you generate a random number according a uniform distribution in $[0, 2]$, the probability P_{not} that it doesn't lie in the support of f (i.e. in $[1/2, 1/2 + 10^{-10}]$) is given by

$P_{\text{not}} = 1 - 10^{-10}/2$. Hence the probability P that at least one of N such random numbers lies in the support has the value

$$P = 1 - P_{\text{no}}^{N} = 1 - (1 - 10^{-10}/2)^{N},$$

so that for $N = 10^7$ we get $P \approx 0.0005$, which for all purposes is equal to zero. We therefore can assume that no random number hits the support, and of course this explains why we get zero as the estimate of I.

Obviously, if we could take many more than 10^{10} random numbers, the Monte Carlo method would yield reliable results again. And while this may seem feasible, we could have rendered any efforts of solving the problem with a computer useless by narrowing the support to, say, $[1/2, 1/2 + 10^{-30}]$.

Perhaps the reader thinks that although all this is correct, we just have discussed an artificially contrived (and thus irrelevant) example. If so, they are quite mistaken. In fact, integrals in statistical mechanics tend to be of the form

$$\int \mathrm{d}\Gamma\, A(\boldsymbol{p}, \boldsymbol{q}) e^{-\beta H(\boldsymbol{p}, \boldsymbol{q})},$$

so that by simply generating a random point in phase space we almost certainly end up with a negligible Boltzmann factor, which doesn't really contribute to the integral under consideration.

Not all is lost, though. The decisive strategy for coping with the problem just described is not to use uniformly distributed random numbers, but rather to choose a distribution which is more appropriate for the integrand under consideration. This approach is known as *importance sampling.*

In order to discuss importance sampling in more detail, let us consider an integral of the form

$$I \equiv \int_V \mathrm{d}^M x\, f(\boldsymbol{x}) g(\boldsymbol{x}),$$

where $g(x)$ is taken to be normalized,

$$\int_V \mathrm{d}^M x\, g(\boldsymbol{x}) = 1,$$

and where again we assume that the volume of integation V is a cube of side length Δ. In addition, let $g(\boldsymbol{x}) \equiv 0$ for all $\boldsymbol{x} \notin V$.Then the simple sampling approach would be to interpret I as

$$I = \Delta^M \int_V \mathrm{d}^M x\, f(\boldsymbol{x}) g(\boldsymbol{x}) p_{\text{u}}(\boldsymbol{x}).$$

However, instead of using a uniformly distributed random vector $\boldsymbol{X}$ we may also introduce a random vector $\boldsymbol{Y}$ the probability density of which is given by $p_{\boldsymbol{Y}}(y) = g(y)$, so that we may write

$$I = \int_V \mathrm{d}^M y\, f(\boldsymbol{y}) g(\boldsymbol{y}) = \int \mathrm{d}^M y\, f(\boldsymbol{y}) p_{\boldsymbol{Y}}(\boldsymbol{y}) = \langle f(\boldsymbol{Y}) \rangle \,.$$

Hence analogous to the case of simple sampling we obtain an estimate of I by generating N random number vectors $\boldsymbol{y}_k$ and using the relation

$$I = \int_V \mathrm{d}^M y\, f(\boldsymbol{y}) g(\boldsymbol{y}) = \langle f(\boldsymbol{Y}) \rangle \approx \frac{1}{N} \sum_{k=1}^{N} f(\boldsymbol{y}_k) + O\left(\frac{1}{\sqrt{N}}\right).$$

As a simple example of importance sampling let us reconsider the integral discussed at the beginning of this section. Here we may choose

$$f(\boldsymbol{x}) \equiv 1$$

and

$$g(x) \equiv \left\{ \begin{array}{ll} 10^{10} & (\frac{1}{2} \leqslant x \leqslant \frac{1}{2} + 10^{-10}) \\ 0 & (\text{otherwise}) \end{array} \right. ,$$

as this $g(x)$ surely is normalized and vanishes outside $[0, 2]$. In order to obtain random numbers distributed according to $g(y)$, in our Monte Carlo program shown in Fig. 4.4 we have to change the line

```
random_number=a+(b-a)*drand48();
```

into

```
random_number=0.5+1e-10*drand48();
```

In addition, the line

```
average+=sqrt(4-random_number*random_number);
```

needs to be replaced by

```
average+=1;
```

With these changes, independent of the number N of random numbers, the program yields 1 as its estimate of the integral. Hence by employing importance sampling we have managed to obtain the exact solution from the Monte Carlo integration.

This all sounds too good to be true, the reader might think. There must be a drawback somewhere. And of course there is. Importance sampling hinges on the fact that we can find random vectors distributed according to some given probability density. But the task of generating such random vectors is far from trivial. We will explain how to do this in the following sections.

However, let us first briefly comment on estimating sums. Here, the discussion runs along the same lines as that for the case of integrals. To see this, consider a finite set of M states $\mathcal{S} = \{s_1, \ldots, s_N\}$ and some function f of these states, and assume that we are interested in the sum

$$S \equiv \sum_{i=1}^{M} f(s_i).$$

Then we may express this sum as

$$S = M \cdot \frac{1}{M} \sum_{i=1}^{M} f(s_i) = M \left\langle f(s) \right\rangle,$$

so that if we generate N states s_k distributed uniformly among the set $\mathcal{S}$ of states, we can estimate S by means of

$$S = M \left\langle s \right\rangle \approx M \cdot \frac{1}{N} \sum_{k=1}^{N} f(s_k) + O\left(\frac{1}{\sqrt{N}}\right).$$

Of course, we may also evaluate a sum by means of importance sampling. In this case we write S in the form

$$S \equiv \sum_{i=1}^{M} f(s_i) g(s_i),$$

where g is assumed to be normalized, i.e.

$$\sum_{i=1}^{M} g(s_i) = 1.$$

Then if we generate N states s_k distributed so that $g(s_k)$ is the probability of s_k, the sum S can be approximated as

$$S \approx \frac{1}{N}\sum_{k=1}^{N} f(s_k) + O\left(\frac{1}{\sqrt{N}}\right).$$

To be more specific, note that the canonical average $\langle A\rangle$ of the Ising model with M spins is given by

$$\langle A\rangle = \sum_{\{\sigma_1,\ldots,\sigma_M} A(\{\sigma_k\})\frac{1}{Z(T)}e^{-H(\{\sigma_k\})/k_\mathrm{B}T}.$$

For these averages it is thus reasonable to carry out a Monte Carlo integration with importance sampling, where $f(s) = A(s)$ and $g(s) = e^{-H(s)/k_\mathrm{B}T}/Z(T)$. Hence we have to come up with an idea of how to construct states distributed according to the Boltzmann distribution.

4.4 Stochastic Processes

At the beginning of Sect. 2.3.3 we saw that from the point of view of probability theory, a sequence of N experiments with random variables (or, more generally, vectors) X_i can formally be regarded as a single experiment with the random vector $\boldsymbol{X} = (X_1, \ldots, X_N)$, where X_i refers to the random variable of the i-th experiment. The probability density of $\boldsymbol{X}$ then may be written in the usual way as

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{\boldsymbol{X}}(x_1, \ldots, x_N).$$

An important example of a sequence of experiments is of course the same experiment repeated at various times. In this case, it is reasonable not to simply label the various random variables related to the various times with somewhat meaningless subscripts, but to introduce a functional $X(\cdot)$, so that $X(t)$ corresponds to the random variable of the experiment carried out at the time t (and as such is *not* a real number, but a function). Similarly we might like to write $x(t_k)$ rather than x_k, so that we can express the probability density as

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{\boldsymbol{X}}(x(t_1), \ldots, x(t_N)).$$

Figure 4.5: Typical outcome of the stochastic process with $p(\boldsymbol{x}) = \prod_{i=1}^{\infty} \Theta(x(t_i))\Theta(1 - x(t_i))$, where $t_i \equiv i$ is assumed.

Obviously, we may drop the assumption that there is a finite number of times and just demand that a countable set of times is considered, so that

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{\boldsymbol{X}}(x(t_1), x(t_2), \ldots).$$

Of course, we might even allow the time to be continuous; however we then would run into trouble concerning the notation used so far. We will discard this possibility.

Temporal sequences of experiments are known as *stochastic processes*. If there is no correlation between the experiments conducted at the various times, the probability density of a stochastic processes factorizes, i.e.

$$p(x(t_1), x(t_2), \ldots) = \prod_{i=1}^{\infty} p_i(x(t_i)),$$

where here and in the following the subscript $\boldsymbol{X}$ is omitted. This property may be described by saying that the correlation time is zero for the process. As a simple example consider

$$p_i(x(t_i)) \equiv \Theta(x(t_i))\Theta(1 - x(t_i))$$

with $t_i \equiv i$. Fig. 4.5 shows a typical outcome of this process.

In general the probability of a certain outcome at the time t_k depends on all the outcomes for the times $t_i < t_k$, so that the probability densities $p(x(t_1), \ldots, x(t_k))$ of the first k and $p(x(t_1), \ldots, x(t_{k-1}))$ of the first $k - 1$ times are related as

$$p(x(t_1), \ldots, x(t_k)) = g(x(t_k)|x(t_1), \ldots, x(t_{k-1})) \cdot p(x(t_1), \ldots, x(t_{k-1})), \quad (4.5)$$

where g may be regarded as a conditional probability (cf. Def. 18). Now assume that the probability density at the time t_k depends just on that at the time t_{k-1}, i.e. that the process under consideration can only "remember" the previous step. Then g can only be a function of $x(t_k)$ and $x(t_{k-1})$, $g = g(x(t_k)|x(t_{k-1}))$, and Eq. 4.5 simplifies to

$$p(x(t_1), \ldots, x(t_k)) = g(x(t_k)|x(t_{k-1})) \cdot p(x(t_1), \ldots, x(t_{k-1})). \quad (4.6)$$

Stochastic processes for which this condition is fulfilled are known as *Markov processes* or, if the set of possible values of the random variable(s) is countable, also as *Markov chains*, and they are the only processes we shall consider in the following.

Before going on, let us introduce the notation

$$p(x_k, t_k; x_{k-1}, t_{k-1}; \ldots; x_2, t_2, x_1, t_1) \equiv p(x(t_1), x(t_2), \ldots, x(t_{k-1}), x(t_k))$$

and similarly

$$g(x_k, t_k | x_{k-1}, t_{k-1}) \equiv g(x(t_k)|x(t_{k-1})),$$

which obviously is somewhat redundant, as dropping all the t_i wouldn't really change any meaning. Note that in p the order of the arguments has been reversed. In addition, we will take t_0 rather than t_1 as the offset of our Markov chain.

Using this notation and iteratively applying the Markov condition Eq. 4.6, we obtain

$$\begin{aligned}
&p(x_k, t_k; x_{k-1}, t_{k-1}; \ldots; x_0, t_0)\\
&\qquad = g(x_k, t_k|x_{k-1}, t_{k-1}) p(x_{k-1}, t_{k-1}; x_{k-2}, t_{k-2}; \ldots; x_0, t_0)\\
&\qquad = g(x_k, t_k|x_{k-1}, t_{k-1}) g(x_{k-1}, t_{k-1}|x_{k-2}, t_{k-2}) \cdot\\
&\qquad\qquad\qquad \cdot p(x_{k-2}, t_{k-2}; x_{k-3}, t_{k-3}), \ldots, x_0, t_0)\\
&\qquad = \ldots\\
&\qquad = g(x_k, t_k|x_{k-1}, t_{k-1}) g(x_{k-1}, t_{k-1}|x_{k-2}, t_{k-2}) \cdot \ldots \cdot\\
&\qquad\qquad\qquad \cdot g(x_2, t_2|x_1, t_1) g(x_1, t_1|x_0, t_0) p(x_0, t_0).
\end{aligned}$$

Here $p(x_0, t_0)$ consitutes the initial condition of the Markov process. Now (loosely speaking) we normally are interested in the probability density of the random variable $X(t_k)$ independent of what the previous outcomes have been, i.e. we are interested in the probability density $p(x_k, t_k)$ given by

$$p(x_k, t_k) = \sum_{x_{k-1} \in \Omega, \ldots, x_0 \in \Omega} p(x_k, t_k; x_{k-1}, t_{k-1}; \ldots; x_0, t_0), \tag{4.7}$$

Figure 4.6: Time evolution of a Markov process. $g(x, t_k|y, t_{k-1})$ constitutes the transition probability from the element y at t_{k-1} to x at t_k, and in order to obtain the probability $p(x, t_k)$ all elements y at t_{k-1} must be considered.

where Ω is the sample space under consideration. This probability may be rewritten by means of the Markov condition as

$$\begin{aligned} p(x_k, t_k) &= \sum_{x_{k-1}\in\Omega,\ldots,x_0\in\Omega} g(x_k, t_k|x_{k-1}, t_{k-1}) p(x_{k-1}, t_{k-1}; \ldots; x_0, t_0) \\ &= \sum_{x_{k-1}\in\Omega} g(x_k, t_k|x_{k-1}, t_{k-1}) \cdot \\ &\qquad \cdot \sum_{x_{k-2}\in\Omega,\ldots,x_0\in\Omega} p(x_{k-1}, t_{k-1}; x_{k-2}, t_{k-2}; \ldots; x_0, t_0), \end{aligned}$$

and applying Eq. 4.7 for $p(x_{k-1}, t_{k-1})$ rather than $p(x_k, t_k)$, we obtain

$$p(x_k, t_k) = \sum_{x_{k-1}\in\Omega} g(x_k, t_k|x_{k-1}, t_{k-1}) p(x_{k-1}, t_{k-1}).$$

If we replace x_{k-1} by y and drop the subscript $_k$ in x_k, we thus arrive at what is known as the *Chapman-Kolmogorov equation*,

$$p(x, t_k) = \sum_{y\in\Omega} g(x, t_k|y, t_{k-1}) p(y, t_{k-1}).$$

Fig. 4.6 illustrates this equation. In order to cast it into a somewhat more useful form, let us assume that the random variable $X(t)$ can assume N distinct values $x^{(i)}$, which don't depend on the specific time chosen. Then we can define a *probability vector* $\boldsymbol{p}(t)$ as

$$\boldsymbol{p}(t) \equiv (p(X(t) = x^{(1)}), p(X(t) = x^{(2)}), \ldots, p(X(t) = x^{(N)}))^{\mathrm{T}},$$

and in addition we may introduce the matrix $\hat{g}(t_k, t_{k-1})$ with the elements $\hat{g}_{ij}(t_k, t_{k-1}) \equiv g(x^{(i)}, t_k|x^{(j)}, t_{k-1})$. But this implies that we may write the Chapman-Kolmogorov equation in matrix form,

$$\boldsymbol{p}(t_k) = \hat{g}(t_k, t_{k-1}) \boldsymbol{p}(t_{k-1}),$$

Figure 4.7: *Left:* Example of an ergodic Markov process. The number at an arrow from $x^{(j)}$ to $x^{(j)}$ corresponds to the component $\hat{g}_{ij}$ of the stochastic matrix. If no arrow is shown, $\hat{g}_{ij} = 0$ is assumed for this transition. For simplicity, homogeneity is assumed. *Right:* Example of a non-ergodic process. No other element can be reached from $x^{(4)}$.

which has the advantage that we may use the formalism of linear algebra for discussing it in more detail. Note that in principle we might live without the condition that N is finite. However, in the case of an infinite, but countable set of $x^{(i)}$, the technical details get more troublesome. We will skip the details and just tacitly assume that, if need be, all the results obtained for the finite case remain valid in the infinite case as well.

So far, we didn't put any severe constraints on the Markov processes considered. However, in order to proceed, we have to concentrate on processes which don't change with time and which cover all the sample space elements. But first let us put these properties in a more rigid form:

Definition 42 *A Markov process is called* homogeneous*, if its transition matrix doesn't change with time, i.e. if*

$$\hat{g}_{ij}(t_k, t_{k-1}) \equiv g(x^{(i)}|x^{(j)}) \equiv \text{const.}$$

Definition 43 *A Markov process is called ergodic, if it is possible to reach all elements of the sample space from any given element, i.e. if for each $x^{(i)}$ and $x^{(j)}$ there exists an N so that*

$$[\hat{g}(t_N, t_{N-1}) \cdot \hat{g}(t_{N-1}, t_{N-2}) \cdot \ldots \cdot g(t_1, t_0)]_{ij} \neq 0.$$

An example of an ergodic and an example of a non-ergodic Markov process are given in Fig. 4.7. Especially in the case of a homogeneous Markov chain we may assume without loss of generality that the discrete times t_k are equidistant, so that $t_{k+1} - t_k \equiv \Delta$ for all k. In this case we denote the stochastic matrix and its components by

$$\hat{g}_\Delta \equiv \hat{g}(t + \Delta, t)$$

and

$$\hat{g}_\Delta(x|y) \equiv g(x, t + \Delta|y, t).$$

If a probability distribution generated by a Markov process is constant in time, it is called a *stationary state*. We may put this in a slightly more formal form as follows:

Definition 44 *A probability vector $\boldsymbol{p}(t)$ is called a stationary state of some Markov process, if it is an eigenvector of the stochastic matrix with eigenvalue 1, i.e. if (cf. the Chapman-Kolmogorov equation)*

$$\boldsymbol{p}(t_k) = \hat{g}(t_k, t_{k-1})\boldsymbol{p}(t_{k-1}) = \boldsymbol{p}(t_{k-1})$$

for all times t_k.

Do all Markov chains have a stationary state? Surely not. As a simple counterexample consider the homogeneous Markov chain given by the stochastic matrix

$$\hat{g}_\Delta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The N-th power of this matrix can easily be computed,

$$\hat{g}_\Delta^N = \begin{cases} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv \hat{G}_1 & (N \text{ odd}) \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \equiv \hat{G}_2 & (N \text{ even}) \end{cases}.$$

But $\hat{G}_1$ possesses the eigenvectors[2] $(1/2, 1/2)$ and $(1/2, -1/2)$, whereas $\hat{G}_2$ has the eigenvectors $(1, 0)$ and $(0, 1)$. Hence there are no common eigenvectors, and there exists no stationary state.

Note that the Markov process considered here is both ergodic and homogeneous, so that these two properties are not sufficient for the existence of stationary states. However, as we will see in a moment, they are nearly so. In order to see why (and what we have to add for sufficiency), we need an important theorem from linear algebra known as Perron's theorem. In order to understand it, the reader should remember that a matrix $A = (A_{ij})$ is called positive, if all the A_{ij} are positive. Analogously, a positive vector $\boldsymbol{x} = (x_i)$ has positive x_i only.

---

[2] As the eigenvectors are supposed to be probability vectors, they are normalized so that the sum of their components is 1.

Theorem 51 (Perron's theorem) *Let A be a positive matrix. Then there exists a real and positive eigenvalue $\lambda_{\max}$ which is not degenerate and the absolute value of which is larger than those of the other eigenvalues. In addition, the corresponding eigenvector $\boldsymbol{e}_{\max}$ is positive.*

Perron's theorem was proved by Perron in 1907 and generalized to irreducible matrices by Frobenius in 1912. [1, 10] Hence it or its generalized version are aloso known as *Frobenius-Perron theorem.*

Proof: The proof is somewhat lengthy, and rather than attempting to give it, we adopt the viewpoint of John 20, 29.[3] Readers feeling uneasy about this may find a proof of the generalized version of the Frobenius-Perron theorem in [1] or [10]. *(QED)*

Perron's theorem forms the key ingredient for the following remarkable theorem on the convergence of Markov processes.

Theorem 52 *Let the $n \times n$ matrix $\hat{g}_\Delta$ be the stochastic matrix of a homogeneous and ergodic Markov process with constant timesteps Δ, and assume that all diagonal elements of $\hat{g}_\Delta$ are positive. Then there exists a stationary state $\boldsymbol{p}_{\text{stat}}$ of the Markov process. In addition, the Markov process converges to $\boldsymbol{p}_{\text{stat}}$ independent of the initial state, i.e.*

$$\lim_{N\to\infty} \hat{g}_\Delta^N \boldsymbol{p}(t_0) = \boldsymbol{p}_{\text{stat}}$$

for all probability vectors $\boldsymbol{p}(t_0)$.

Proof: We start by showing that there exists a $K \in \mathbb{N}$ so that the K-th power of $\hat{g}_\Delta$ is positive. To this end consider some arbitrary, but fixed i and j. Then due to the ergodicity of the Markov process, there must be a $K(i,j) \in \mathbb{N}$ for which $(\hat{g}_\Delta^{K(i,j)})_{ij} > 0$. Furthermore, as all the components of both $\hat{g}_\Delta$ and $\hat{g}_\Delta^{K(i,j)}$ can be regarded as transition probabilities, it is clear that they need to be non-negative. Hence we may conclude that

$$(\hat{g}_\Delta^{K(i,j)+1})_{ij} = (\hat{g}_\Delta^{K(i,j)}\hat{g}_\Delta)_{ij} = \sum_{k=1}^{n} \underbrace{(\hat{g}_\Delta^{K(i,j)})_{ik}}_{\geqslant 0} \underbrace{(\hat{g}_\Delta)_{kj}}_{\geqslant 0} \geqslant \underbrace{(\hat{g}_\Delta^{K(i,j)})_{ij}}_{>0} \underbrace{(\hat{g}_\Delta)_{jj}}_{>0} > 0.$$

(Remember that $(\hat{g}_\Delta)_{jj} > 0$ was assumed in the theorem.) Thus by induction we see that $(\hat{g}_\Delta^{K'})_{ij} > 0$ for all $K' > K(i,j)$. But then if K denotes the

---

[3] "blessed are they that have not seen, and yet have believed" [6]

maximum of all the $K(i,j)$, $(\hat{g}_\Delta^{K'})_{rs}$ must be positive independent of the specific r and s chosen, i.e. $\hat{g}_\Delta^K$ is positive, which is what we set out to show.

Now consider an arbitrary positive eigenvector $\boldsymbol{e}$ of $\hat{g}_\Delta^K$ with eigenvalue λ, and assume that it is normalized to be a probability vector. Due to the fact that as a product of stochastic matrices, $\hat{g}_\Delta$ must itself be a stochastic matrix, we can infer for the choice $\boldsymbol{p}(t_0) = \boldsymbol{e}$ that

$$1 = \sum_{i=1}^n (\hat{g}_\Delta^K \boldsymbol{p}(t_0))_i = \sum_{i=1}^n (\hat{g}_\Delta^K \boldsymbol{e})_i = \sum_{i=1}^n (\lambda \boldsymbol{e})_i = \lambda \sum_{i=1}^n (\boldsymbol{e})_i = \lambda,$$

so that the only possible eigenvalue for probability vectors is 1. On the other hand, as $\hat{g}_\Delta^K$ is positive, we can deduce from Perron's theorem that it has a non-degenerate (real) eigenvalue $\lambda_{\max}$ with a positive eigenvector $\boldsymbol{e}_{\max}$. Hence we have seen that concerning probability vectors, $\hat{g}_\Delta^K$ possesses exactly one eigenvector $\boldsymbol{e}_{\max}$, and the corresponding eigenvalue $\lambda_{\max}$ is 1.

Now let us take $\boldsymbol{e}_{\max}$ as the initial state $\boldsymbol{p}(t_0)$. Then

$$\begin{aligned}\hat{g}_\Delta^K \boldsymbol{p}(t_1) &= \hat{g}_\Delta^K \hat{g}_\Delta \boldsymbol{p}(t_0) = \hat{g}_\Delta^{K+1} \boldsymbol{e}_{\max} = \hat{g}_\Delta \hat{g}_\Delta^K \boldsymbol{e}_{\max} = \hat{g}_\Delta(\lambda_{\max} \boldsymbol{e}_{\max}) \\ &= \lambda_{\max} \hat{g}_\Delta \boldsymbol{p}(t_0) = \lambda_{\max} \boldsymbol{p}(t_1) = \boldsymbol{p}(t_1),\end{aligned}$$

which implies that $\boldsymbol{p}(t_1)$ is also an eigenvector of $\hat{g}_\Delta^K$. Thus from the uniqueness of the eigenvector it follows that

$$\boldsymbol{p}(t_1) = \boldsymbol{e}_{\max} = \boldsymbol{p}(t_0).$$

In other words, $\boldsymbol{e}_{\max}$ is a eigenvector of the stochastic matrix and hence a stationary state.

It remains to be shown that the Markov process converges to the stationary state thus found. We prove this first for the special case that the stochastic matrix has n (possibly complex) linear independent eigenvectors $\boldsymbol{e}_k$ with eigenvalues λ_k. Then the eigenvectors form a basis, and choosing $\boldsymbol{e}_1 \equiv \boldsymbol{e}_{\max}$, we can write an arbitrary inititial state in the form

$$\boldsymbol{p}(t_0) \equiv q\boldsymbol{e}_{\max} + \sum_{k=2}^n r_k \boldsymbol{e}_k, \tag{4.8}$$

where $q, r_k \in \mathbb{C}$. We therefore get

$$\hat{g}_\Delta^N \boldsymbol{p}(t_0) = \hat{g}_\Delta^N \left(q\boldsymbol{e}_{\max} + \sum_{k=2}^n r_k \boldsymbol{e}_k \right) = q\boldsymbol{e}_{\max} + \sum_{k=2}^n r_k \lambda_k^N \boldsymbol{e}_k$$

But from Perron's theorem we know that the absolute values of the λ_k for $k \geqslant 2$ must be less than that of $\lambda_{\max}$, i.e. $|\lambda_k| < 1$ for $k \geqslant 2$. Accordingly, $|\lambda_k^N|$ and thus also λ_k^N vanish in the limit of infinite N, and we are left with

$$\lim_{N\to\infty} \hat{g}_\Delta^N \boldsymbol{p}(t_0) = q\boldsymbol{e}_{\max}.$$

But as $\hat{g}_\Delta^N$ is a stochastic matrix, so will be $\lim_{N\to\infty} \hat{g}_\Delta^N$. Hence the right hand side in this equation must be a probability vector, so that $q = 1$ and

$$\lim_{N\to\infty} \hat{g}_\Delta^N \boldsymbol{p}(t_0) = \boldsymbol{e}_{\max}.$$

Let us now turn to the general case, where we cannot assume that the eigenvectors of the stochastic matrix form a basis. This means, of course, that we cannot rely on the existence of the decomposition of Eq. 4.8 any longer. However, assuming that there are u eigenvectors $\boldsymbol{e}_k$ with eigenvalues λ_k and choosing $\boldsymbol{e}_1 \equiv \boldsymbol{e}_{\max}$, we may express the matrix $\hat{g}_\Delta^N$ as

$$\hat{g}_\Delta^N = \frac{C(1)}{\overset{1}{\chi}(1)} + \sum_{k=2}^{u} \frac{1}{(m_k-1)!} \left[\frac{C(\lambda)}{\overset{k}{\chi}(\lambda)} \lambda^N \right]^{m_k-1} \Bigg|_{\lambda=\lambda_k},$$

where m_k is the degree of degeneracy of λ_k and $\chi(\lambda)$ denotes the characteristic polynomial,

$$\chi(\lambda) \equiv \prod_{k=1}^{N} (\lambda - \lambda_k)^{m_k}.$$

$C(\lambda)$ constitutes the reduced adjoint matrix, $C(\lambda) \equiv (\lambda \cdot 1 - \hat{g}_\Delta)^{-1}\chi(\lambda)$, and $\overset{k}{\chi}(\lambda)$ is defined as

$$\overset{k}{\chi}(\lambda) \equiv \frac{\chi(\lambda)}{(\lambda - \lambda_k)^{m_k}}.$$

We refer the reader to Ch. XIII, § 7 of [1] and Ch. V, § 3 of [2] for a proof. As above, for $k \geqslant 2$, λ_k^N and thus $\lambda_k^{N\cdot m_k}$ vanish in the limit of $N \longrightarrow \infty$, so that

$$\lim_{N\to\infty} \hat{g}_\Delta^N = \frac{C(1)}{\overset{1}{\chi}(1)}.$$

Hence $\hat{g}_\Delta^N$ converges. Now assume there existed an initial state $\boldsymbol{p}(t_0)$ which doesn't converge to a stationary state. Then there would be some $\epsilon, \epsilon' > 0$ so that for every $N \in \mathbb{N}$ we could find a $N' \in \mathbb{N}$ with

$$\max_i |(\hat{g}_\Delta^{N'+1} \boldsymbol{p}(t_0) - \hat{g}_\Delta^{N'} \boldsymbol{p}(t_0))_i| > \epsilon$$

or, equivalently,

$$\max_{i,j}|(\hat{g}_\Delta^{N'+1} - \hat{g}_\Delta^{N'})_{ij}| > \epsilon'.$$

Thus $\hat{g}_\Delta^N$ wouldn't converge, contrary to what we have just shown. Accordingly, any initial state must converge into the stationary state $\boldsymbol{e}_{\max}$, and the theorem has been proved. *(QED)*

The conceptual importance of Theorem 52 should not be underestimated: If a physical system can be modeled by a homogeneous, ergodic Markov chain, it will end up in a (or rather *the*) stationary state, independent of where it started off. This goes a long way towards explaining why the simple recipe advocated at the beginning of Chapter 3 for preparing an equilibrium system ("just leave the system alone and wait long enough") actually works.

We finish this section by providing a version of Theorem 52 which requires only somewhat weakened conditions and will prove helpful when discussing the Metropolis algorithm in the next section.

Theorem 53 *Let the $n \times n$ matrix $\hat{g}_\Delta$ be the stochastic matrix of a homogeneous and ergodic Markov process with constant timesteps Δ, and assume that for all i either the diagonal element g_{ii} is positive or there exists a k such that both $(\hat{g}_\Delta)_{kk}$ and $(\hat{g}_\Delta)_{ik}$ are positive. Then there exists a stationary state $\boldsymbol{p}_{\text{stat}}$ of the Markov process. In addition, the Markov process converges to $\boldsymbol{p}_{\text{stat}}$ independent of the initial state, i.e.*

$$\lim_{N\to\infty} \hat{g}_\Delta^N \boldsymbol{p}(t_0) = \boldsymbol{p}_{\text{stat}}$$

for all probability vectors $\boldsymbol{p}(t_0)$.

Proof: Evidently, it is sufficient to prove that for a sufficiently large K' $\hat{g}_\Delta^{K'}$ is positive, as then we can just use the proof provided for Theorem 52. We may assume without loss of generality that for $1 \leqslant i \leqslant r$ the $(\hat{g}_\Delta)_{ii}$ are positive, whereas for $r+1 \leqslant i \leqslant n$ they vanish. Then by arguments completely analogous to those given at the beginning of the proof for Theorem 52, it can be shown that there exists a K such that the $(\hat{g}_\Delta^K)_{ij}$ are positive for all $i \leqslant r$. However, choosing any $k > r$, we know that there must be a $l \leqslant r$ with $(\hat{g}_\Delta)_{kl} > 0$, and thus we get

$$(\hat{g}_\Delta^{K+1})_{kk} = \sum_{s=1}^{n} (\hat{g}_\Delta)_{ks} (\hat{g}_\Delta^K)_{sk} = \underbrace{(\hat{g}_\Delta)_{kl}}_{>0} \underbrace{(\hat{g}_\Delta^K)_{lk}}_{>0} + \sum_{\substack{s=1 \\ s\neq l}} \underbrace{(\hat{g}_\Delta)_{ks}}_{\geqslant 0} \underbrace{(\hat{g}_\Delta^K)_{sk}}_{\geqslant 0} > 0.$$

Furthermore, as a power of a stochastic matrix for an ergodic Markov process, $\hat{g}_{\Delta}^{K+1}$ is itself a stochastic matrix for an ergodic process. We therefore can use the arguments given at the beginning of the proof of Theorem 52 to show that there exists a K' so that $(\hat{g}_{\Delta}^{K})^{K'} = \hat{g}_{\Delta}^{K \cdot K'}$ is positive, which is all we had to prove. *(QED)*

4.5 The Metropolis algorithm

At the end of Sect. 4.3 we realized that in order to employ a Monte Carlo simulation with importance sampling, we need to construct states which are distributed according to a given probability density. Having discussed Markov prpcesses, we are now in the position to successfully tackle this problem.

In order to do so, we start by noting that in the case of a finite number of states (which we shall be content with), the (discrete) probability density can be interpreted as a probability vector. Hence we might paraphrase our task by stating that we need to construct some given probability vector.

But lo behold! From Theorem 52 we know that all homogeneous, ergodic Markov processes converge to their (unique) stationary state. Hence, if we have a Markov process with the desired probability density as its stationary state, we have a solution to our problem: We just have to start with an arbitrary probability vector and run the Markov process sufficiently long. In a way, we have replaced our old problem by a new one, as we still have to provide a recipe of how to find an appropriate Markov process for a given stationary state.

This, however, will turn out to be simpler than might be thought at first sight. But let us leave aside the question for the moment and reconsider the Chapman-Kolmogorov equation,

$$p(x, t_k) = \sum_{y \in \Omega} g(x, t_k | y, t_{k-1}) p(y, t_{k-1}).$$

Subtracting $p(x, t_{k-1})$ on both sides of the equation, we get

$$p(x, t_k) - p(x, t_{k-1}) = \sum_{y \in \Omega} g(x, t_k | y, t_{k-1}) p(y, t_{k-1}) - p(x, t_{k-1}),$$

and remembering that for any stochastic matrix $\sum_{y \in \Omega} g(y, t' | x, t) = 1$ holds valid, we obtain the Master equation.

Figure 4.8: Interpretation of the Master equation. The right hand side of Eq. 4.9 may be regarded as a sum of terms describing scattering in and out of an element in phase space. *Left:* There is a gain at $x \in \Omega$ due to transitions from (other) sample space elements ("scattering in"). *Right:* There is a loss at x due to transitions from this to other elements in sample space ("scattering out").

Theorem 54 *The time development of a probability distribution in a Markov process is given by the* Master equation*:*

$$p(x, t_k) - p(x, t_{k-1}) = \sum_{y \in \Omega} [g(x, t_k | y, t_{k-1}) p(y, t_{k-1}) - g(y, t_k | x, t_{k-1}) p(x, t_{k-1})] \tag{4.9}$$

From a "physical" point of view, the master equation may be viewed as a description of scattering in and out of some sample space element. This is illustrated by Fig. 4.8.

Dividing both sides of the Master equation by $t_k - t_{k-1}$ yields

$$\begin{aligned} \frac{p(x, t_k) - p(x, t_{k-1})}{t_k - t_{k-1}} = \sum_{y \in \Omega} [g^*(x, t_k | y, t_{k-1}) p(y, t_{k-1}) \\ - g^*(x, t_k | y, t_{k-1}) p(x, t_{k-1})], \end{aligned}$$

where the transition probabilities per time g^* are defined by

$$g^*(x, t_k | y, t_{k-1}) \equiv \frac{g(x, t_k | y, t_{k-1})}{t_k - t_{k-1}}.$$

Thus in the limit of infinitely small time steps we obtain the continuous version of the Master equation:

$$\frac{\mathrm{d}p(x, t)}{\mathrm{d}t} = \sum_{y \in \Omega} [g^*(x, t | y, t) p(y, t) - g^*(y, t | x, t) p(x, t).$$

If for all $x, y \in \Omega$ (and all times) the relation $g(x, t_k | y, t_{k-1}) p(y, t_{k-1}) = g(y, t_k | x, t_{k-1}) p(x, t_{k-1})$ holds valid, we say that *detailed balance* is fulfilled. In this case, obviously the left hand side of the Master equation vanishes, and we get the following theorem.

Figure 4.9: Metropolis algorithm. An arbitrary initial s_{new} has to be supplied. After sufficiently many loops s_{new} can be taken as the output of the algorithm.

Theorem 55 *A state which fulfils detailed balance is a stationary state of the Markov process under consideration.*

This theorem is good news indeed: In order to construct a homogeneous Markov process for a given stationary state we just have to come up with a stochastic matrix $\hat{g}$ so that $\hat{g}$ and the given state fulfil detailed balance.

We can now finally come back to the simulation of canonical ensembles, which can be carried out by means of the *Metropolis algorithm.* In order to discuss this algorithm, let us consider a system with n states and a Hamiltonian H.

Basically the Metropolis algorithm consists in a repeated application of a step turning a given initial state s_{old} into a new state s_{new}. Here, the step can be divided into two parts: First, an ergodic (and possibly random) construction is employed in order to obtain a state s' from s_{old}. Second, the energies $H(s')$ and $H(s_{\text{old}})$ are compared. If the energy of s' is less than that of s_{old}, it is used as the new state s_{new}. Otherwise a random number r distributed uniformly between 0 and 1 is generated. If $r < e^{-\beta(H(s')-H(s_{\text{old}}))}$, s' is chosen as s_{new}; if $r \geqslant e^{-\beta(H(s')-H(s_{\text{old}}))}$, s_{old} is kept. In other words,

$$s_{\text{new}} = \begin{cases} s' & \text{for } H(s') < H(S_{\text{old}}) \\ s' & \text{for } H(s') \geqslant H(S_{\text{old}}) \text{ and } r < e^{-\beta(H(s')-H(s_{\text{old}}))} \\ s_{\text{old}} & \text{for } H(s') \geqslant H(S_{\text{old}}) \text{ and } r \geqslant e^{-\beta(H(s')-H(s_{\text{old}}))} \end{cases} . \tag{4.10}$$

Fig. 4.9 may serve as an illustration. The idea behind the Metropolis algorithm is of course that the state obtained after sufficiently many repetitions will be distributed according to the Boltzmann distribution.

Theorem 56 *Consider a system with n states s_i and a Hamiltonian H. Assume*

1. *that for all i, j it is equally likely to get $s' = s_i$ from $s_{\text{old}} = s_j$ as it is to get $s' = s_j$ from $s_{\text{old}} = s_i$, and*
2. *that for all i either $s' = s_i$ can be obtained from $s_{\text{old}} = s_i$ or there exists a k so that $H(s_k) < H(s_i)$ and $s' = s_i$ can be obtained from $s_{\text{old}} = s_k$.*

Then after sufficiently many steps the state yielded by the Metropolis algorithm will be distributed according to the Boltzmann distribution, i.e. $p(s_i) = e^{-\beta H(s_i)}/Z$.

Admittedly, assumption (2) made in this theorem seems slightly technical. However, it isn't superfluous. For example, consider a single spin which may either be directed parallel or antiparallel to some given direction. Obviously, if there is no external field, the energy of this two-state system doesn't on the state whatsoever. Now we might realize the construction of s' required by the Metropolis algorithm by

$$s' = \begin{cases} s_2 & \text{for } s_{\text{old}} = s_1 \\ s_1 & \text{for } s_{\text{old}} = s_2 \end{cases}$$

Then the stochastic matrix of the Markov process corresponding to the Metropolis algorithm has the form

$$\hat{g} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and we know from the beginning of the previous section that there is no convergence for this matrix. Hence the Metropolis algorithm in this case doesn't yield the correct result.

It has to be stressed, though, that this is a rather artificial and atypical example. In practice, the Metropolis algorithm converges to the equilibrium distribution even if assumption (2) of Theorem 56 isn't fulfilled. After all, it was needed only to ensure that some power of the stochastic matrix is positive. But of course while the assumption is sufficient for this, it clearly isn't necessary.

Proof: The Metropolis algorithm may be viewed as a Markov process, the stochastic matrix $\hat{g}$ of which may be written in the form

$$\hat{g}_{ij} = \hat{C}_{ij} \cdot \hat{T}_{ij},$$

where $\hat{C}$ and $\hat{T}$ are the stochastic matrices corresponding to the construction of s' (from s_{old}) and s_{new} (from s'). As $\hat{C}$ is ergodic by definition and as $\hat{T}$ doesn't render any transition impossible, it is evident that $\hat{g}$ must be ergodic as well. Homogeneity is ensured by definition.

Now consider a state s_i for which $\hat{g}_{ii} = 0$. Then by assumption there exists a k with $H(s_k) < H(s_i)$ and $\hat{C}_{ik} > 0$. For this k $0 < T_{ik} < 1$ holds valid, and thus both $\hat{g}_{kk}$ and $\hat{g}_{ik}$ must be positive. Hence we see that either $g_{ii} > 0$ or there is a k with $\hat{g}_{kk} > 0$ and $\hat{g}_{ik} > 0$.

Accordingly, all the conditions of Theorem 53 are met, and we can conclude that the Markov process converges to a stationary state. It remains to be shown what this stationary state looks like.

As we would like to prove that the Metropolis algorithm yields the Boltzmann distribution $p(s_i) = e^{-\beta H(s_i)}/Z$, it seems reasonable to try and see whether this happens to be the stationary state we are looking for. In order to convince ourselves that this indeed the case, let us consider two arbitrary states s_i and s_k, which may or may not be the same.

Let us assume first that $H(s_i) < H(s_k)$. Then from Eq. 4.10 we know that $T_{ik} = 1$, so that

$$\hat{g}_{ik} p(s_k) = \hat{T}_{ik} \hat{C}_{ik} p(s_k) = \hat{C}_{ik} \frac{e^{-\beta H(s_k)}}{Z}. \tag{4.11}$$

Conversely, as the probability for a random number distributed uniformly in $[0,1]$ to be smaller than x is just x, we can gather from Eq. 4.10 that $\hat{T}_{ki} = e^{-\beta(H(s_k)-H(s_i))}$. Furthermore, from assumption (1) we have that $\hat{C}_{ik} = \hat{C}_{ki}$. We therefore get

$$\hat{g}_{ki} p(s_i) = \hat{T}_{ki} \hat{C}_{ki} p(s_i) = e^{-\beta(H(s_k)-H(s_i))} \cdot \hat{C}_{ik} \cdot \frac{e^{-\beta H(s_i)}}{Z} = \hat{C}_{ik} \frac{e^{-\beta H(s_k)}}{Z}. \tag{4.12}$$

By comparing Eqs. 4.11 and 4.12, we are led to the conclusion that detailed balance is fulfilled,

$$\hat{g}_{ik} p(s_k) = \hat{g}_{ki} p(s_i).$$

We now turn to the case of $H(s_i) \geqslant H(s_k)$. Here, using the same arguments as above, we obtain the equations

$$\hat{g}_{ik} p(s_k) = \hat{T}_{ik} \hat{C}_{ik} p(s_k) = e^{-\beta(H(s_i)-H(s_k))} \cdot \hat{C}_{ik} \cdot \frac{e^{-\beta H(s_k)}}{Z} = \hat{C}_{ik} \frac{e^{-\beta H(s_i)}}{Z}$$

and

$$\hat{g}_{ki} p(s_i) = \hat{T}_{ki} \hat{C}_{ki} p(s_i) = \hat{C}_{ik} \frac{e^{-\beta H(s_i)}}{Z},$$

so that again we obtain obtained balance. Hence from Theorem 55 we can infer that the Boltzmann distribution is indeed the stationary state of the Markov process. Hence after sufficiently many steps, the state resulting from the Metropolis algorithm will be distributed according to this distribution, which is precisely what we wanted to prove. *(QED)*

Strictly speaking, one has to wait until the Metropolis algorithm has (sufficiently) converged. The number of steps (or "time") this takes is known as *equilibration time*. Furthermore, as the new state obtained from a Metroplis step is not independent from the old one, it is clear that there should lie some steps between two subsequent states used for a Monte Carlo simulation. The number of steps one has to wait for this is called *correlation time*.

Finally, let us briefly discuss the implementation of the Ising model with N spins. Basically, this just requires turning the Metroplis algorithm into a computer program, which is quite straightforward. However, we have to clarify how to construct the state s' from a state s_{old}. One solution would be to generate a random integer k distributed uniformly in $[1, N]$ and then to flip the k-th spin. It is easy to see that this method fulfils the conditions of Theorem 56.

Another solution is to flip the first spin in the first step, the second in the second step, the third in the third step, and so forth. N steps (during which each spin is flipped once) are known as a *sweep* of the lattice. While this method surely works in practice, it should be noted that it violates assumption (2) of Theorem 56.

The reader is encouraged to try and compose a program for solving the Ising model by means of the methods just outlined. An exemplary C++ program is given in Appendix **??**. Those that would like to include a graphical interface are referred to [3] for more information.

Appendix A

Simple C program for the Ising model

In the following, a simple C program for solving the Ising model (cf. Sect. 4.1.2) by means of the Metropolis algorithm (cf. 4.5). Some comments are offered at the bottom of the program.

still to be included

Bibliography

[1] F.R. Gantmacher. *Matrix Theory*, volume 2. Chelsea, 1959.

[2] F.R. Gantmacher. *Matrix Theory*, volume 1. Chelsea, 1959.

[3] Wolfgang Kinzel and Georg Reents. *Physics by Computer*. Springer-Verlag, 1998.

[4] L.D. Landau and E.M. Lifshitz. *Mechanics*. Pergamon Press, 1960.

[5] H.J. Larson. *Introduction to Probability Theory and Statistical Inference*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1969.

[6] J. Moffatt. *The New Testament*. Harper & Brothers, 1935.

[7] M.E.J. Newman and G.T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, 1999.

[8] W.H. Press et al. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.

[9] D. Ruelle. *Statistical Mechanics: Rigorous Results*. Mathematical Physics Monograph Series. W.A. Benjamin, 1969.

[10] R.S. Varga. *Matrix Iterative Analysis*. Springer Series in Computational Mathematics. Springer, 2000.