2. Distribution functions and statistical averages

2.1. Formulating a statistical description from particle mechanics

We will regularly refer to a collection of particles occupying a specified region of some space as a *system*. A system may be open or closed depending on the behavior at its boundaries: at an open boundary particles and energy may be exchanged with a neighboring system; fully closed boundaries do not communicate with the rest of the world (picture perfectly reflecting walls confining a gas). *One-component systems* contain only a single kind of particle, examples being a monoatomic gas of helium atoms or pure liquid water. We will also consider *multi-component systems*, air being an example. Chemical reactions can alter the number of molecules of one chemical specie relative to that of others and the energy barrier separating them can be treated as another kind of boundary.

In mechanics we solve the equations of motion for a system of particles[1] to obtain their trajectories from given initial conditions. In statistical physics one traditionally utilizes the Hamiltonian, rather than the Newtonian, formulation of mechanics where the motion is expressed in terms of generalized coordinates, q_i, and momenta, p_i, of which we assume there are s of each for a total of 2s; for a three dimensional system of N particles we have $s = 3N$. The equations governing Hamiltonian mechanics are

$$\frac{\partial H}{\partial q_i} = -\dot{p}_i \tag{2.1a}$$

and

$$\frac{\partial H}{\partial p_i} = \dot{q}_i, \tag{2.1b}$$

where $\dot{q}_i$ denotes the time derivative of the ith coordinate and similarly for $\dot{p}_i$. For a system containing a single specie of particle and in Cartesian coordinates, the Hamiltonian is given by

$$H(p_1, p_2 \cdots p_s; q_1, q_2 \cdots q_s) = \sum_{i=1}^{s} \frac{p_i^2}{2m} + V(q_1, q_2 \cdots q_s), \tag{2.2}$$

where $V(q_1, q_2 \cdots q_s)$ is the potential acting between them. The solution of these equations of motion can be written

$$q_i = q_i(t; q_{01} \cdots q_{0s}; p_{01} \cdots p_{0s}) \tag{2.3a}$$

---

[1] For simplicity we will initially restrict to a one-component system.

and

$$p_i = p_i(t; q_{01} \cdots q_{0s}; p_{01} \cdots p_{0s}),$$ (2.3b)

where $q_{01} \cdots q_{0s}$ and $p_{01} \cdots p_{0s}$ are the initial values of the coordinates and momenta at some chosen time t_0, of which there are 6N in three dimensions.

Condensed terrestrial matter at ordinary densities contains of order 10^{22} particles / cm^3. If the system were closed (uncoupled to the outside world) and the equations of motion were solved, the result would not be very useful for practical purposes. What would one do with 10^{22} different time-dependent functions containing the same number of initial conditions as parameters? Worse yet, if the system is open (as most are), the equations cannot be solved in principle. Statistical physics, as the name implies, avoids the need to know the trajectories of all the particles by employing a *probabilistic description*. We note that most of the coordinates used to describe systems in classical mechanics are *continuous* (e.g., the particle positions and momenta). Hence our description will involve *probability densities* as opposed to probabilities.

We begin with a simple example. Suppose the world is static and we have a single particle the position of which is only known probabilistically. We write the *differential probability*, dw, that the particle will have a coordinate q that lies in an interval between q and q + dq as

$$dw = \rho(q)dq$$ (2.4)

where $\rho(q)$ is then a probability density in the space of q. Since the particle must be somewhere in this space, it is natural to impose the *normalization condition*

$$\int dw = \int dq \rho(q) = 1.$$ (2.5)

Suppose we wish to form the average $\bar{f}$ of some function $f(q)$ that depends on our coordinate q ; this too is naturally defined as

$$\bar{f} = \int dq \rho(q) f(q).$$ (2.6)

We now generalize this kind of description to a system defined by our 2s Hamiltonian coordinates, $q_1 \cdots q_s$ and $p_1 \cdots p_s$, together with time. Our differential probability dw then becomes

$$dw = \rho(t; q_1, q_2 \cdots q_s; p_1, p_2 \cdots p_s) dq_1 \cdots dq_s dp_1 \cdots dp_s,$$ (2.7)

where $\rho(t;q_1,q_2\cdots q_s;p_1,p_2\cdots p_s)$ is referred to as the *distribution function*, a probability density in this now much larger space. If our system does not interact with the outside world, which we now assume, then ρ will not be *explicitly* time dependent, i.e., $\rho \neq \rho(t)$; all time dependence is then built into the Hamiltonian coordinates themselves.

We will again adopt the normalization condition

$$\int dw = \int dq_1 \cdots dq_s \, dp_1 \cdots dp_s \, \rho(q_1,q_2\cdots q_s;p_1,p_2\cdots p_s) = 1. \qquad (2.8)$$

The averaging procedure corresponding to Eq. (2.6) for some function f of the 2s Hamiltonian coordinates can be written as[2]

$$\overline{f} = \int dq_1 \cdots dq_s dp_1 \cdots dp_s)\rho(q_1,q_2\cdots q_s;p_1,p_2\cdots p_s)f(q_1\cdots q_s;p_1\cdots p_s). \quad (2.9)$$

The 2s dimensional space $q_1\cdots q_s; p_1\cdots p_s$ is called *phase space*. We recall that in considering the mechanics of a single particle moving in one space dimension, one often plots a trajectory in the two-dimensional space of p and q, along which the particle moves in time. By analogy our multi-particle system can be pictured as a point moving in time along a trajectory in the 2s dimensional phase space.

As a short hand notation in what follows we will often write p in place of $p_1,p_2\cdots p_s$ and q for $q_1,q_2\cdots q_s$. Similarly we will also write dpdq in place of $dp_1,dp_2\cdots dp_s dq_1 dq_2 \cdots dq_s$ for the differential volume element in this space.

2.2. The Liouville's theorem

We again emphasize that, in general, all particles in a given system are in motion; i.e., there is an implicit time dependence and the particles continually enter and leave any given volume element in phase space. We now examine this behavior in more detail.

We start by recalling that in fluid mechanics one introduces an "equation of continuity" which embodies the conservation of mass[3] for the flow of a fluid having a mass density $\rho(t,\mathbf{r})$.

---

[2] Formally, $\overline{f}$ has a time dependence here that arises from that of the Hamiltonian coordinates.

[3] The mass, M, within some volume, V , is given by $M(t) = \int_V d^3r\rho(t,\mathbf{r})$ where $\rho(t,\mathbf{r})$ is the mass density. The rate at which the mass in this volume changes is then $\dot{M}(t) = \int_V d^3r\dot{\rho}(t,\mathbf{r})$; because mass is conserved this change must arise from a flow of fluid across the surface, S, that encloses V, which is given by $= -\int_S d^2\mathbf{S} \cdot \mathbf{f}(t,\mathbf{r})$ where $\mathbf{f}(t,\mathbf{r}) = \rho(t,\mathbf{r})\mathbf{v}(t,\mathbf{r})$ is the mass crossing unit area in unit time, $\mathbf{v}$ is the fluid velocity, and $d\mathbf{S}$ points

We now generalize the equation of continuity to cover the case of our 2s dimensional phase space and view the corresponding probability density, $\rho(t,p,q)$, as a point moving in that space, which will then have the form

$$\frac{\partial \rho}{\partial t} + \sum_{i=1}^{s} \frac{\partial}{\partial q_i}\left(\rho \dot{q}_i\right) + \sum_{i=1}^{s} \frac{\partial}{\partial p_i}\left(\rho \dot{p}_i\right) = 0 .$$ (2.10)

Integration of Hamilton's equations of motion for specified inter-particle forces yields $q_i = q_i(t)$ and $p_i(t)$, as written out in Eq.'s (2.3a,b). Carrying out the differentiations in Eq. (2.10) we have

$$0 = \sum_{i=1}^{s}\left[\frac{\partial}{\partial q_i}\left(\rho \dot{q}_i\right) + \frac{\partial}{\partial p_i}\left(\rho \dot{p}_i\right)\right] = \sum_{i=1}^{s}\left[\frac{\partial \rho}{\partial q_i}\dot{q}_i + \rho\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \rho}{\partial p_i}\dot{p}_i + \rho\frac{\partial \dot{p}_i}{\partial p_i}\right].$$ (2.11)

From Hamilton's equations we have

$$\frac{\partial \dot{q}}{\partial q_i} = \frac{\partial^2 H}{\partial q_i \partial p_i} = \frac{\partial^2 H}{\partial p_i \partial q_i} = -\frac{\partial \dot{p}_i}{\partial p_i}$$ (2.12)

where we assumed the equality of the cross derivatives of the Hamiltonian in the second step; hence the second and the fourth terms on the right hand side of Eq. (2.11) *cancel*, thereby leaving

$$\frac{\partial \rho}{\partial t} + \sum_{i=1}^{s}\left[\frac{\partial \rho}{\partial q_i}\dot{q}_i + \frac{\partial \rho}{\partial p_i}\dot{p}_i\right] = 0 ,$$ (2.13)

which we identify as the *total* time derivative of the distribution function[4]. Therefore we can write

$$d\rho / dt = 0 .$$ (2.14)

Although the particles actually move along trajectories in phase space, the distribution function for their motion as a whole, viewed as a point moving our 2s dimensional phase space, remains

---

outward from V. Applying Green's theorem we have $\int_V d^3 r \dot{\rho} = -\int_S d^3 r\, \nabla \cdot (\rho \mathbf{v})$. Since the volume element is arbitrary this relation must be satisfied by the integrands; i.e., we must have $\partial \rho / \partial t + \nabla \cdot (\rho \mathbf{v}) = 0$. This latter relation expresses the conservation of mass mathematically.

[4] When taking the total derivative of our function, in addition to the explicit time dependence, here assumed zero, we must also include the implicit time dependence, that contained in q_i and p_i, using the chain rule. This kind of derivative is often called "substantial" or "hydrodynamic" derivative.

constant[5]. If we limit ourselves to a *closed* system (without external time dependent forces), ρ will have no *explicit* time dependence and (2.13) takes the form

$$\sum_{i=1}^{s} \frac{\partial \rho}{\partial q_i} \cdot \dot{q}_i + \sum_{i=1}^{s} \frac{\partial \rho}{\partial p_i} \cdot \dot{p}_i = 0 . \qquad (2.15)$$

Eq.'s (2.13) and Eq. (2.15) constitute *Liouville's theorem*; in the form (2.15) it holds only so long as the interaction of our system with the rest of the world can be ignored.

2.3. Calculating an average distribution function from a microscopic perspective

The distribution function introduced above still contains all the complexity associated with the implicit time dependence of the 2s coordinates $p(t)$ and $q(t)$. In this and the next sub section we will discuss two strategies to *suppress* this complexity by introducing two different averaging procedures to convert $\rho(p(t),q(t))$ into a function $\rho(p,q)$ where the Hamiltonian coordinates have *no time dependence*. The latter can be viewed as a *statistical distribution function* as opposed to the actual (implicitly time-dependent) one. Unfortunately it is common to denote both distributions functions as $\rho(p,q)$.

Suppose we were able to solve the equations of motion. Figuratively, we can then use a "stop watch" to record the total amount time, Δt, that the system spends within a small but finite phase space volume $\prod_{i=1}^{s} \Delta p_i \Delta q_i$ centered on some point $q_1 \cdots q_s; p_1 \cdots p_s$ (see Fig. 2.1); dividing Δt by the total elapsed time, t, and taking the limit $t \to \infty$ we obtain Δw as

$$\lim_{t \to \infty} \frac{\Delta t}{t} = \Delta w . \qquad (2.16)$$

The differential probability density then follows from Eq. (2.7) by taking the following limit

$$\rho(p,q) = \lim \frac{\Delta w}{\Delta p \Delta q} . \qquad (2.17)$$

---

[5] As a result, even though the particles in an actual system are in constant motion, the introduction of a probability density to explain their averaged behavior (that does not explicitly contain time) is *consistent* with the laws of mechanics, as contained in Hamilton's equations.

Figure 1.1 shows, schematically, the trajectory for a single pair of Hamiltonian coordinates (p_i, q_i) as the system enters and leaves a volume element $\Delta p_i \Delta q_i$ as the time evolves.

With the availability of $p(t)$ and $q(t)$ averages of some function $f(p(t),q(t))$ can also be calculated directly with the following procedure

$$\bar{f} = \lim_{t \to \infty} \frac{1}{t} \int_0^t f(p(t),q(t)) dt . \tag{2.18}$$

But, as stated earlier, $p(t)$ and $q(t)$ cannot be obtained for a large system.

2.4. Replacing time averages with ensemble averages

Rather than interpreting $\rho(p,q)$ as resulting from a *time average* using the prescription given by Eq. (2.17), we can regard it as arising from the averaging of a large number of identical "copies" of our system at some arbitrarily chosen time. One way to do this would be to have different but "compatible" initial conditions (with the same total energy, see below). The set of such copies is called an *ensemble*. The distribution function then measures how the ensemble of copies distribute their motion over different regions of our phase space.
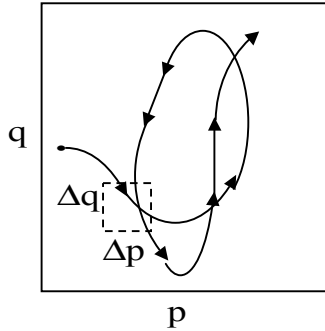

Fig. 2.1. The "trajectory" in two of the six dimensions of the phase space (p, q) of a single particle. Shown also is a two dimensional "volume" element $\Delta p \Delta q$ through which the particle can pass multiple times and within which it spends a total time Δt during some elapsed time interval t. The contribution of this particle's to the distribution function is then: $\Delta t / t = \rho(p,q) \Delta p \Delta q$.

The assumption that these two approaches, the time averages and the ensemble averages, are equivalent is called the *ergotic hypothesis*. It has long been debated, but will be accepted here as being valid.

Assuming that $\rho(p,q)$ is available, the average, $\bar{f}$, of some property $f(p,q)$ may then be calculated using Eq. (2.9) rather than (2.18). As introduced in section 1.1 for the case of discrete properties, one may also calculate the mean square fluctuation $\overline{(\Delta f)^2}$ of the quantity $f(p,q)$ as

$$\overline{(\Delta f)^2} = \int dp dq \, \rho(p,q) \big(f(p,q) - \bar{f}\big)^2$$
$$= \int dp dq \, \rho(p,q) \big(f(p,q)^2 - 2f(p,q)\bar{f} + \bar{f}^2\big)$$

$$= \int dpdq\, \rho(p,q) f(p,q)^2 - \overline{f}^2 , \qquad\qquad (2.19)$$

where we used the definition of $\overline{f}$ in the last step. According to the discussion in section 1.2 this quantity will be small relative to $\overline{f}^2$ for a system with many particles.

2.5. Additivity, statistical independence and sub systems

Experimentally, uniform macroscopic matter has the property that it can be divided into parts and, provided they are not too small, certain properties of the whole are accurately given as a sum of those parts; i.e., they are *additive*. Examples of such properties are the energy and particle number; in thermodynamics these are called *extensive quantities*. This additive property of neutral matter[6] arises from the short range character of the forces acting between its parts[7]. On the other hand the individual parts, provided they are in equilibrium with each other (and themselves), have additional properties that are the same for all the parts, such as temperature and pressure; i.e., they do not scale with the amount of material; these are called *intensive quantities.*

In mechanics certain constants of motion in a many body system arise from the underlying symmetries associated with time, translation, and rotation; when these are present the total energy, E, momentum, $\mathbf{P}$, and angular momentum $\mathbf{M}$, are constants of motion for a total of $1 + 3 + 3 = 7$ constants of motion. In what follows we will largely restrict to non-rotating systems at rest, i.e., systems with $\mathbf{M} = 0$ and $\mathbf{P} = 0$, so there is only a single additive (extensive) constant of motion, E. This is to be compared with the 6N constants of motion for a single-specie, N-particle system[8] in three dimensions. Furthermore, in the absence of chemical reactions the individual particle numbers in a multi-specie system are also constant.

From the uniformity displayed by thermodynamic systems in equilibrium it is intuitively suggestive that the vast majority of the initial conditions must be *totally irrelevant* on a macroscopic scale: two equal volumes, V, of water at the same temperature, T, and pressure, p (intuitively familiar terms that will be precisely defined in our subsequent discussion), can for all practical purposes be taken as identical. We may then anticipate that the distribution function of

---

[6] Here we restrict to over all electrically neutral matter; ionized gases, called plasmas, have longer range forces and must be treated with special techniques.

[7] The van der Waals interaction between molecules is a common example. On the other hand gravitational interactions are long range and the behavior of large bodies is decidedly not a sum of its parts.

[8] If our particles are molecules, additional degrees of freedom are present associated with rotations as a whole and vibrations of the constituent atoms.

a uniform system in equilibrium will depend explicitly on certain intensive parameters, here denoted as π_i; i.e., $\rho = \rho(p,q;\pi_i)$.[9]

We now ask how we can construct a suitable form for an equilibrium distribution function for some system. To begin, suppose we can break a system with distribution function $\rho^{(total)}(p^{total},q^{total};\pi_i)$ into two parts[10] in a minimally intrusive way such that the intensive parameters, π_i, are left undisturbed. Such parts will be referred to as *sub systems* of the total system. If we further imagine that these parts do not subsequently interact, then their distribution functions can be written as $\rho^{(1)}(p^{(1)},q^{(1)};\pi_i)$ and $\rho^{(2)}(p^{(2)},q^{(2)};\pi_i)$ with all coordinates of part 1 being independent of those in part 2. Such systems are said to be *statistically independent.* Then for each configuration of the first we have all configurations of the second: i. e., the probability density for the total system can, to a high degree of accuracy, be represented as a *product* of its parts

$$\rho^{(total)}(p^{total},q^{total};\pi_i) = \rho^{(1)}(p^{(1)},q^{(1)};\pi_i)\rho^{(2)}(p^{(2)},q^{(2)};\pi_i);$$

we then see that probabilities combine in a *multiplicative* fashion[11].
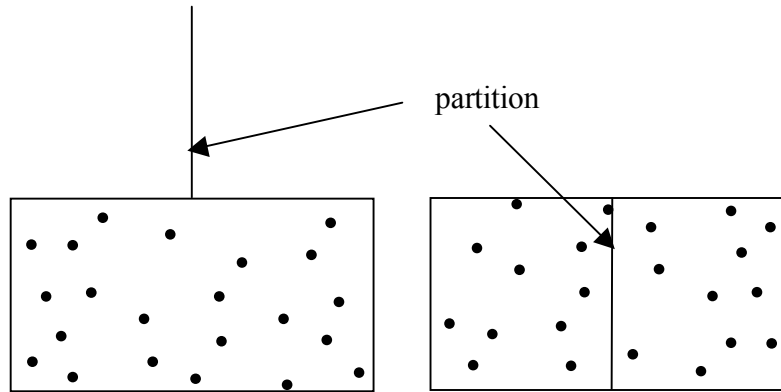

**Fig. 2.2. Partitioning a larger system into a two
smaller ones by inserting a partition.**

We now imagine a large system in statistical equilibrium which is only *figuratively* partitioned into a collection of sub systems. In that case "mixing" (e.g., an exchange of particles and energy) will subsequently occur at the "boundaries" so imposed. But the fraction of the

---

[9] For the most part we will designate intensive variables by lower case letters and extensive variables by upper case; an exception will be temperature which would otherwise be confused with time.

[10] Here we might imagine inserting a wall in a vessel containing a liquid or gas.

[11] As an example involving discrete probabilities, the number of outcomes for tossing one dice is 6 while that for two tosses is $6 \times 6 = 36$.

subsystem that is positioned near its boundaries is small compared to that deep in its interior (and becomes progressively smaller as its size increases), so for times short compared to those required to achieve significant *mixing* we can still regard the subsystems as *statistically independent*, each dependent only on the coordinates contained within it. When such conditions are satisfied we can then write

$$\rho^{(total)}(p,q;\pi_i) = \rho^{(1)}(p^{(1)},q^{(1)};\pi_i)\rho^{(2)}(p^{(2)},q^{(2)};\pi_i)\cdots\rho^{(N)}(p^{(N)},q^{(N)};\pi_i)$$

$$= \prod_{a=1}^{N}\rho^{(a)}(p^{(a)},q^{(a)};\pi_i), \tag{2.20}$$

where N is the number of sub systems, $(p^{(a)},q^{(a)};\pi_i)$ denotes the coordinates describing the ath sub system, and the relevant intensive parameters π_i (such as temperature) are assumed the same for all the parts. We remark in passing that the "total" system could either be isolated or part of a still larger system.

2.6. The macro-canonical ensemble

From the multiplicative property of our probability densities, it immediately follows that the *logarithm* of the total probability density is equal to the *sum* of those for the sub systems:

$$\ln\rho^{total} = \ln\rho^{(1)} + \ln\rho^{(2)} + \ln\rho^{(3)} + \cdots$$

$$= \sum_{a=1}^{N}\ln\rho^{(a)}. \tag{2.21}$$

Eq. (2.21) strongly suggests that, since the logarithm of the product of the sub system probability densities is additive, $\ln\rho^{(a)}$ for each subsystem will be proportional to the associated *additive mechanical constants of motion for that subsystem*; for our subsystem at rest this only involves the energy, $E(p^{(a)},q^{(a)})$ so we write

$$\ln\left[(2\pi\hbar)^{s^{(a)}}\rho^{(a)}\right] = -\alpha^{(a)} - \beta E(p^{(a)},q^{(a)}), \tag{2.22}$$

or equivalently

$$\rho^{(a)} = \frac{1}{(2\pi\hbar)^{s^{(a)}}}e^{-\alpha^{(a)}-\beta E(p^{(a)},q^{(a)})}; \tag{2.23}$$

here β is an intensive parameter in the set π_i discussed above[12] and is *yet to be determined*, but is the same for subsystems in equilibrium,[13] whereas $(p^{(a)}, q^{(a)})$ are the Hamiltonian coordinates of the particles lying *within* subsystem a. Note that in defining this distribution we have included an additional factor $(2\pi\hbar)^{s(a)}$ which has the effect of building into the resulting probability density the semi-classical prescription that connects the number of quantum states to the differential phase space volume wherein the volume element $dq^{(a)}dp^{(a)}$ is replaced by $dq^{(a)}dp^{(a)} / (2\pi\hbar)^{s(a)}$; i.e., rather than integrating over phase space we integrate over the *density of states*.

Noting the equivalence between the energy and the Hamiltonian given in (2.2), Eq. (2.23) can also be written

$$\rho^{(a)} = \frac{1}{(2\pi\hbar)^{s^{(a)}}} e^{-\alpha^{(a)} - \beta H(p^{(a)}, q^{(a)})} . \tag{2.24}$$

A probability density of this form is called (for historical reasons) a *macro-canonical ensemble* (or a canonical ensemble for short); it is also called the *Gibbs distribution[14]*. Since the subsystems are in contact the individual $E^{(a)}$ will fluctuate, although if they are large these fluctuations will be small.

We must emphasize again that the distribution function (2.24) is not the distribution function of an actual system, which would involve the time dependent trajectories of all its particles, which in turn involves their initial conditions. As we have argued a description at that level of detail would be useless, even if it could be determined. Averages of various quantities computed with the Gibbs distribution together with their fluctuations will differ from those for a

---

[12] In simpler cases the parameter $\alpha^{(a)}$ (see below) is related to the normalization; in Chapter 5 it will be shown that $\beta = (k_B T)^{-1}$ where T is the absolute temperature, which is the same for all subsystems in thermal equilibrium, and k_B is Boltzmann's constant. If our system was translating and rotating, Eq. (2.23) would be replaced by

$$\ln\left[(2\pi\hbar)^{s^{(a)}} \rho^{(a)} \right] = -\alpha^{(a)} - \beta E^{(a)} - \boldsymbol{\gamma} \cdot \mathbf{P}^{(a)} - \boldsymbol{\delta} \cdot \mathbf{M}^{(a)}$$

where $\mathbf{P}^{(a)}$ and $\mathbf{M}^{(a)}$ are the translational and angular momentum and the vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ encompass six additional parameters that are equal among subsystems; π_i then consists of a total of *seven* parameters.

[13] The sign convention used in Eq. (2.22) differs from that of Landau and Lifshitz but is in accord with that used in other treatments.

[14] Named after Josiah Willard Gibbs,1839 - 1903, an American physicist who together with J. C. Maxwell and L. Boltzmann laid the foundations of statistical physics.

real system by small amounts characteristic of the law of large numbers. We not in passing that even with these restrictions Eq. (10.22) satisfied Liouville's equation in the form (2.15), as is easily verified.

It is important to reiterate that what allowed us to construct the Gibbs distribution (2.24) was the application of the following principles: i) for times such that mixing can be ignored, a large system can be divided into smaller (though still macroscopic) sub systems; ii) these subsystems can then be taken as statistically independent (Eq. (2.20)), and hence the logarithm of their distribution functions is additive; iii) the energies of these sub systems are also additive; and iv) these two quantities should then be linearly related (Eq. (2.22)).

Since the coordinates of the subsystems are taken as distinct, we can apply our normalization condition to each of them *independently*[15]

$$\frac{1}{(2\pi\hbar)^{s^{(a)}}} \int dp^{(a)} dq^{(a)} \rho^{(a)}(p^{(a)}, q^{(a)}) = 1.$$ (2.25)

This will also satisfy the condition that the distribution function for the system as a whole is normalized.

2.7. Taking sub-system averages: the partition function

Recalling Eq (2.9), using Eq. (2.24), and suppressing the subsystem index a, the average of a quantity $f(p,q)$ is given by

$$\overline{f} = \int \frac{dpdq}{(2\pi\hbar)^s} f(p,q) e^{-\alpha - \beta H(p,q)}.$$ (2.26)

From the normalization condition (2.25) we have

$$\int \frac{dpdq}{(2\pi\hbar)^s} e^{-\alpha - \beta H(p,q)} = 1.$$ (2.27)

Therefore

$$\frac{1}{(2\pi\hbar)^s} e^{-\alpha} = \frac{1}{\int dpdq\, e^{-\beta H(p,q)}};$$ (2.28)

---

[15] An alternate normalization is to equate (2.27) to the number of particles in the subsystem, $N^{(a)}$; the normalization of ρ^{total} would then give $N = \sum_a N^{(a)}$ consistent with the particle numbers being an additive constant of motion.

11

i.e., α is here *fixed by the normalization*, and we have included the factor $(2\pi\hbar)^{-s}$ for convenience. We can then write Eq. (2.26) as

$$\bar{f} = \frac{\int dpdq f(p,q) e^{-\beta H(p,q)}}{\int dpdq\, e^{-\beta H(p,q)}} \; . \tag{2.29}$$

This provides a *general prescription* for computing sub system averages. The denominator in this expression is called the *partition function* and is denoted as Z:

$$Z = \int dpdq e^{-\beta H(p,q)} \; . \tag{2.30}$$

Let us examine the following

$$-\frac{\partial}{\partial\beta}\ln Z = \frac{\int dpdq H(p,q) e^{-\beta H(p,q)}}{\int dpdq\, e^{-\beta H(p,q)}} = \bar{E} \; ; \tag{2.31}$$

i.e., it is the average value of the energy. Other thermodynamic averages may be obtained by taking related derivatives of the function Z .

2.8. The probability density for an isolated system: the microcanonical ensemble

In section 2.4 we used the multiplicative property of probability densities to construct a form that is suitable for a group of systems in equilibrium with each other (or a with larger one that is sometimes called a thermostat). If on the other hand one imagines a system that is *totally isolated* from the outside world, then energy, momentum and angular momentum would be strictly conserved, having some values E_0, $\mathbf{P}_0$, and $\mathbf{M}_0$; the differential probability for the system as a whole, dw^{total}, would then have to contain the form

$$dw^{\text{total}} \propto \delta(E - E_0)\delta(\mathbf{P} - \mathbf{P}_0)\delta(\mathbf{M} - \mathbf{M}_0) \; . \tag{2.32}$$

The form (2.34) explicitly builds in the constraints imposed by the seven additive constants of motion for an isolated system, as described in section 2.4.

We can simplify the description by imagining our system is contained in a rigid box, the position and orientation of which is fixed in space; interactions of the particles with the walls of that box will then build into the resulting motions of the particles the vanishing of $\mathbf{P}$ and $\mathbf{M}$ (without the need for the corresponding delta functions). In this case Eq. (2.32) can be written

$$dw^{\text{total}} \propto \delta(E - E_0) \; . \tag{2.33}$$

12

For historical reasons differential probabilities of this form are called *micro-canonical ensembles.* Experimentally such a system is difficult, if not impossible, to achieve in practice.[16] Furthermore they are awkward to work with and for this reason are not commonly utilized in practice. However they are important from a foundational perspective, particularly when developing a quantum mechanical formulation of statistical physics, which we take up in the next chapter.

Problems

2.1. Suppose you are given a three dimensional gas of N non-interacting atoms of mass m in a volume V. The Hamiltonian associated with the translational motion of a single atom is

$$H = (p_x^2 + p_y^2 + p_z^2)/2m.$$

    a) Use the partition function to calculate the total energy, E, of the gas.
Now from thermodynamics we know that $dE = -PdV$.

    b) Calculate the pressure P.
Later we will show that $\beta = 1/k_B T$ where k_B is Boltzmann's constant.

    c) Use this to calculate the *Equation of State* for the gas which relates the pressure, P, volume, V, and temperature, T, this is the *Perfect Gas Law*; define the constant R that occurs in the conventional form of this law as given in elementary chemistry texts.
Thermodynamics also tells us that the heat capacity at constant volume is given $C_V = (\partial E/\partial T)_V$.

    d) Calculate C_V for this gas.

2.2. Suppose you are given a one-dimensional oscillator with the Hamiltonian

$$H = \frac{p_x^2}{2m} + \frac{1}{2}m\omega_0^2 x^2$$

where ω_0 is the frequency.

    a) Use the partition function to calculate the average energy of this oscillator.
        Now assume your oscillator can vibrate in *three dimensions* and that you have N such oscillators in a volume V (assume the oscillators do not interact with each other; this is called an *Einstein solid*).
    b) Using the expression from the previous problem calculate C_V.

---

[16] Heat conduction through surfaces or container walls, and direct coupling via electromagnetic radiation, are always present.

Avogadro's number of such oscillators would have a total heat capacity of 3R where R is the constant in the perfect gas law.

c) What is the value of R? (This is the *Dulong–Petit* law discussed in elementary chemistry texts).