

Unit 2-2: The Ergodic Hypothesis – Time vs Ensemble Averages

We now want to address the question of how to relate thermodynamics to mechanics more generally. Kinetic theory is one approach, but one has to be able to go beyond the simple model of non-interacting particles discussed in the previous section, and be able to incorporate interactions between the particles – the Boltzmann equation is one formalism for doing that.

Here we will adopt a more “modern” approach and use the method of statistical ensembles developed by Gibbs in the early 1900’s (so 120 years old is still “modern”!). The method of statistical ensembles has at its conceptual core the *ergodic hypothesis*.

Consider a system of N particles, each with three spatial degrees of freedom x, y, z . The system is described in classical mechanics by the Hamiltonian \mathcal{H} , which is a function of $6N$ canonical variables, $q_1, q_2, \dots, q_{3N}, p_1, p_2, \dots, p_{3N}$. The $3N$ variables q_i give the spatial coordinates of the N particles, and the $3N$ variables p_i are the corresponding canonical momenta. These $6N$ variables denote the *phase space* of the system. At any moment in time, the state of the system is specified by its position in phase space, i.e. by specifying the value of each of these $6N$ coordinates.

The time evolution of the system as it moves in phase space is given by Hamilton’s equations,

$$\dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i}, \quad \dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i}, \quad i = 1, 2, \dots, 3N \quad (2.2.1)$$

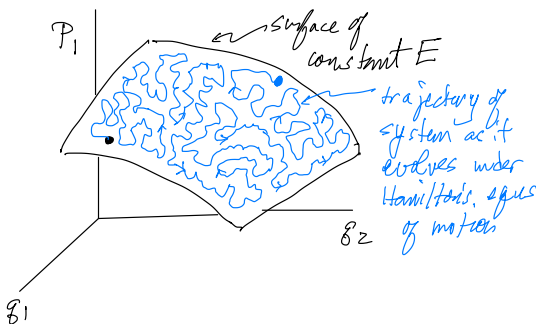
Solving Hamilton’s equations gives the phase space trajectory of the system $\{q_i(t), p_i(t)\}$.

In general, assuming the system is isolated from all external degrees of freedom, the total energy of the system will be conserved as the system moves on its phase space trajectory. If the total energy of the system is E , then the condition $\mathcal{H}[q_i, p_i] = E$ defines a $6N - 1$ dimensional surface in phase space to which the system’s trajectory is confined (when I write $[q_i, p_i]$, I will mean the set of all $3N$ of the q_i and all $3N$ of the p_i).

If one wanted to compute the measured value of some physical quantity X , averaged over an interval of time τ , it would be given by,

$$\langle X \rangle = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} dt X[q_i(t), p_i(t)] \quad (2.2.2)$$

where $X[q_i(t), p_i(t)]$ is the value that the quantity X takes when the system is at coordinates $\{q_i(t), p_i(t)\}$ in phase space.



In general, for large N , we expect that the system trajectory in phase space will be some horribly complicated curve on the surface of constant energy E , that we have no hope of computing directly (note, since Hamilton’s equations give a unique trajectory for each initial condition $\{q_{i0}, p_{i0}\}$, such a phase space trajectory may never intersect itself).

[Even if we had the best computers in the universe and beyond, we could not do the calculation because most classical mechanical systems with many interacting degrees of freedom develop “deterministic chaos” after a sufficiently long time. Although the equations of motion are deterministic, small changes in the initial conditions lead to exponentially growing changes in the final state after a long time period of evolution. So any uncertainty in the initial conditions, no matter how small (and the finite bit size of words in computer memory always results in a finite accuracy), will ultimately lead to unpredictability of the state of the system after sufficiently long time.]

To compute $\langle X \rangle$ we therefore need to make an assumption. The *ergodic hypothesis* says that, for a system in equilibrium, during any time interval τ sufficiently long, the location of the system in phase space $\{q_i(t), p_i(t)\}$ is *equally*

likely to be *anywhere* on the surface of constant energy E . If this is so, then we can write,

$$(i) \quad \langle X \rangle = \int dq_i dp_i X[q_i, p_i] \rho(q_i, p_i) \quad \text{where} \quad (ii) \quad \rho(q_i, p_i) = C \delta(\mathcal{H}[q_i, p_i] - E) \quad (2.2.3)$$

where C is a normalizing factor such that $\int dq_i dp_i \rho(q_i, p_i) = 1$.

The distribution $\rho(q_i, p_i)$ is called the *density matrix*, and it is the probability density for the system, in equilibrium, to be found at a particular point $\{q_i, p_i\}$ in phase space. According to the ergodic hypothesis, for a system with fixed energy E , that probability distribution is a uniform constant on the surface of constant energy E and zero elsewhere.

In other words, in the absence of any further information, we assume that in equilibrium all microscopic states $\{q_i, p_i\}$ consistent with a give set of macroscopic thermodynamic variables, E, V, N , are equally likely.

Computing averages as in Eq. (2.2.3)(i) is called ensemble theory. Using the density matrix ρ of the form in Eq. (2.2.3)(ii) is called the microcanonical ensemble.

In *ensemble theory* one abandons any effort to compute thermodynamic properties from the explicit time dependent trajectory of the system in phase space as in Eq. (2.2.2). Rather one describes the thermodynamic state as represented by a particular *ensemble* given by a density matrix $\rho(q_i, p_i)$.

One can interpret the ensemble average, as in Eq. (2.2.3)(i), as the value one would find, not for a single isolated system moving on its trajectory, but for the average of a collection of systems distributed in phase space according to the density ρ . The ergodic hypothesis asserts that the time average of Eq. (2.2.2) and the ensemble average of Eq. (2.2.3) are equal.

Equilibrium is described by a density matrix ρ that does not vary in time. We will soon see other ensembles besides the microcanonical ensemble of Eq. (2.2.3)(ii).

The ergodic hypothesis cannot in general be proven (it has been proven only for some very special simple systems – see the work of Sinai). But the existence of thermodynamics as an empirically consistent theory suggests why the ergodic hypothesis may be true. We can consider the following two points.

1) Liouville's Theorem of classical mechanics, which we will discuss in the next set of notes, shows that any $\rho(q_i, p_i)$ that is independent of time, and so may describe equilibrium, must be constant on constant energy surfaces.

2) By positing a thermodynamic description we assume that the macroscopic properties of a system are completely described by a set of only a few macroscopic variables, such as E, V , and N . If the ergodic hypothesis were not true, then there would be parts of phase space with the same value of E that never “saw” each other – i.e. a trajectory in one part would not enter the other part, and vice versa. One could imagine, therefore, that systems in these two disjoint regions of phase space might have different properties, i.e. have different time averages of some particular property $X[q_i, p_i]$. One therefore might expect them to represent thermodynamically *distinguishable* states. But this would contradict the assumption that E alone is the important thermodynamic variable.

Alternatively, if ergodicity fails, there might be some other important macroscopic variable (for example magnetization) which one overlooked. The disjoint regions of the constant energy surface could correspond to different values of this new macroscopic variable. This reflects back to a comment made early in Notes 1-1, that in making a thermodynamic description of a system one must first correctly identify all the relevant macroscopic variables. A globally conserved quantity is always a candidate for such a macroscopic variable. Once one has identified all the macroscopically relevant variables, the ergodic hypothesis implies that further, more microscopic, information about the state of the system will not effect the thermodynamic behavior, i.e. all microscopic states consistent with a set of macroscopic variables are all equally likely.