

Introduction to Statistical Mechanics

Hamiltonian systems and Liouville's Theorem

Rohan Hitchcock

2 August 2023

Why are we here?

1 The basic set-up

Statistical mechanics is concerned with dynamical systems where we don't know the precise state of the system, but we none-the-less want to make statements about the behaviour of the system overall. This is usually framed as a distinction between *microstates*, which exactly specify our system but we are not able to measure, and *macrostates*, which are measurable properties that encapsulate the properties of the system we care about but do not uniquely determine the state of our system. A typical example is a gas: the exact state of the system at a given time — the microstate — consists of the positions and momenta of every particle, while we are only able to observe quantities like pressure, volume and temperature.

We set things up as follows. For system we consider the set \mathcal{R} of all possible states of that system, which we call *phase space*. For the purpose of this talk we can take $\mathcal{R} = \mathbb{R}^n$, and in the example of the gas above we have $\mathcal{R} = \mathbb{R}^{6N}$ where N is the number of particles (each particle has three position and three momentum coordinates). Our system is dynamical, meaning it evolves over time, and so any given state of our system $\mathbf{x} \in \mathcal{R}$ is considered to be a function of time: $\mathbf{x} := \mathbf{x}(t)$, although time dependence is usually suppressed from the notation. We denote time derivatives as $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt}$.

To reflect our uncertainty about the true state of the system we represent it as a random variable \mathbf{X}_t (more precisely a sequence of random variables in time). The probability density of the state of our system at time t is denoted by $\rho(\mathbf{x}, t)$.

Definition 1 (Tentative). A *statistical mechanical system* consists of a phase space \mathcal{R} , dynamics on states $\mathbf{x} \in \mathcal{R}$, and a sequence of probability densities $\rho(\mathbf{x}, t)$ (one for each $t > 0$).

Definition 1 is not really a sufficient description of a statistical mechanical system. Of course there are a number of technical conditions which need to be satisfied (ρ should at least be a differentiable with respect to t , \mathcal{R} should be a manifold (probably smooth) with a compatible measure...), but more to the point the dynamics of the system need to be compatible with the sequence of probability densities. Given any trajectory in phase space (generated by the dynamics), $\rho(\mathbf{x}, t)$ must 'consistently assign' probability to a system following this trajectory. Rather than trying to be more precise now, for the rest of the talk we will investigate common conditions which are sufficient for ensuring a statistical mechanical system has this coherence property. Before continuing to this we note one other common condition imposed on statistical mechanical systems.

Definition 2. A statistical mechanical system is *in equilibrium*, or equivalently its density ρ is *stationary*, if

$$\frac{\partial \rho}{\partial t} = 0 \tag{1}$$

Note that even if $\rho(\mathbf{x}) = \rho(\mathbf{x}, t)$ is stationary it still depends on time via the time-dependant state \mathbf{x} .

2 The continuity equation

One way of ensuring that the probability distributions of the \mathbf{X}_t are consistent with respect to the dynamics of the physical system is to assume that there are no sources or sinks or probability mass in phase space. Formally, this amounts to assuming that the density $\rho(\mathbf{x}, t)$ satisfies a *continuity equation*. This is the same equation that, for example, the mass density of a fluid must satisfy to ensure (local) conservation of mass.

We will now derive this continuity equation, following the argument in [PB11, Chapter 2.2]. Let V be a compact region of phase space with non-zero volume with smooth boundary $S = \partial V$. The probability that the system is in the region V at time t is given by the integral

$$\mathbf{P}(\mathbf{X}_t \in V) = \int_V \rho(\mathbf{x}, t) d\mathbf{x} .$$

Note that even if ρ is a stationary distribution $\mathbf{P}(\mathbf{X}_t \in V)$ depends on time. We consider the rate $\frac{\partial}{\partial t} \mathbf{P}(\mathbf{X}_t \in V)$ changes over time. The net flow of probability mass *out* of V is given by integrating the probability flux $\rho \dot{\mathbf{x}}$ over the boundary S . If $\hat{\mathbf{n}}$ is the outward unit normal of S then this is

$$\int_S (\rho \dot{\mathbf{x}}) \cdot \hat{\mathbf{n}} dS = \int_V \nabla \cdot (\rho \dot{\mathbf{x}}) d\mathbf{x}$$

by Gauss's divergence theorem. Since we assume that there are no sources or sinks of probability mass, any change in $\mathbf{P}(\mathbf{X}_t \in V)$ with time must be as a result of flow into or out of V . That is,

$$\frac{\partial}{\partial t} \mathbf{P}(\mathbf{X}_t \in V) = - \int_V \nabla \cdot (\rho \dot{\mathbf{x}}) d\mathbf{x} .$$

By moving differentiation with respect to t under the integral expression for $\mathbf{P}(\mathbf{X}_t \in V)$ we find

$$\int_V \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \dot{\mathbf{x}}) d\mathbf{x} = 0$$

Since this is true of every such region V we obtain

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \dot{\mathbf{x}}) = 0 \tag{2}$$

everywhere in phase space, which is the continuity equation for probability mass. Note that if the system is in equilibrium then satisfying the continuity equation is equivalent to the probability flux $\rho \dot{\mathbf{x}}$ being divergence-free.

The continuity equation (2) can be thought of as expressing a strong ‘local’ conservation property of some quantity whose density is given by ρ . Not only must this quantity be globally conserved, but it cannot be generated in one part of phase space and destroyed in equal measure in another. Note that not all physical systems satisfy this property, a typical example being a system which dissipates energy. In this setting, the probability

mass of the system accumulates in regions which have lower energy and abandons regions of higher energy. Hence regions of low energy are sinks of probability mass and regions of high energy are sources.

Example 1. Consider a ball rolling on the surface $z = x^2 + y^2$, whose position at time $t = 0$ is drawn from some prior distribution φ on \mathbb{R}^2 . The state of this system is specified by a coordinate $(x, y) \in \mathbb{R}^2$. If the ball is subject to friction (i.e. it has a mechanism to dissipate energy) then as t increases its probability distribution concentrates on the point $(0, 0)$ in phase space. Hence we expect that $(0, 0)$ is a sink of probability mass.

We will cover such systems in a future talk, but for now we will focus on systems which satisfy the continuity equation (2).

3 Hamiltonian systems

We will now study a class of physical system, which we call a *Hamiltonian system*, whose dynamics are specified by the Hamiltonian equations. Let the phase space be \mathbb{R}^{2n} for some n . A state of the system $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ consists of two types of coordinates: position $\mathbf{q} = (q_1, \dots, q_n) \in \mathbb{R}^n$ and momentum $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}^n$. The coordinates q_i and p_i are related by a real-valued function $H(\mathbf{q}, \mathbf{p})$ of the system state, which in the case of physical systems is the total energy of the system. The positions and momenta satisfy the system of partial differential equations

$$\dot{q}_i = \frac{\partial}{\partial p_i} H(\mathbf{q}, \mathbf{p}) \quad \dot{p}_i = -\frac{\partial}{\partial q_i} H(\mathbf{q}, \mathbf{p}) \quad \text{for } i = 1, \dots, n \quad (3)$$

solutions to which give the dynamics of the system.

One example of such a system is the gas from earlier: the coordinates (q_1, q_2, q_3) and (p_1, p_2, p_3) are respectively the position and momentum of the first particle, and so on. The Hamiltonian $H(\mathbf{q}, \mathbf{p})$ is the total energy of the system of particles. If we assume the particles of the gas do not interact (as is typically done) then this is the sum of the kinetic and potential energy of each particle.

Suppose that the probability density $\rho(\mathbf{q}, \mathbf{p}, t)$ of our Hamiltonian system satisfies the continuity equation (2), which we recall amounts to assuming there are no sources or sinks of probability mass in phase space. Following [PB11, Chapter 2.2], since $\dot{\mathbf{x}} = (\dot{\mathbf{q}}, \dot{\mathbf{p}})$ we have that the divergence of $\rho \dot{\mathbf{x}}$ is

$$\begin{aligned} \nabla \cdot (\rho \dot{\mathbf{x}}) &= \sum_{i=1}^n \left(\frac{\partial}{\partial q_i} (\rho \dot{q}_i) + \frac{\partial}{\partial p_i} (\rho \dot{p}_i) \right) \\ &= \sum_{i=1}^n \left(\frac{\partial \rho}{\partial q_i} \dot{q}_i + \rho \frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \rho}{\partial p_i} \dot{p}_i + \rho \frac{\partial \dot{p}_i}{\partial p_i} \right) \\ &= \sum_{i=1}^n \left(\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) + \rho \sum_{i=1}^n \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) \end{aligned}$$

Using the Hamilton's equations (3) we have

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

and so we have

$$\nabla \cdot (\rho \dot{\mathbf{x}}) = \sum_{i=1}^n \left(\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) = \{\rho, H\}$$

where $\{\rho, H\}$ is the *Poisson bracket*

$$\{\rho, H\} = \sum_{i=1}^n \left(\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} \right) .$$

Hence from the continuity equation (2) we arrive at the following named result.

Theorem 3 (Liouville's Theorem¹). *For a Hamiltonian system, ρ satisfies the continuity equation (2) if and only if*

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \{\rho, H\} = 0 . \quad (4)$$

Note that two statements are being made here. The fact that $\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0$ follows from the continuity equation (2) and the working above. The fact that $\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \{\rho, H\}$ is true of any Hamiltonian system and follows from the chain rule

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \dot{\mathbf{x}} \cdot (\nabla \rho)$$

and by noting that $\dot{\mathbf{x}} \cdot (\nabla \rho) = \{\rho, H\}$ by Hamilton's equations (3).

Equation (4) is (apparently) the equation defining the flow of an incompressible fluid governed by Hamiltonian H , so Liouville's Theorem can be interpreted as stating that the density ρ evolves like a fluid with the same Hamiltonian as the underlying physical system.

Lemma 4. *If the density $\rho(\mathbf{q}, \mathbf{p}, t)$ depends only on the Hamiltonian H , that is for some function $f : \mathbb{R} \rightarrow [0, 1]$ we have*

$$\rho(\mathbf{q}, \mathbf{p}, t) = f(H(\mathbf{q}, \mathbf{p}))$$

for all states (\mathbf{q}, \mathbf{p}) and times t , then ρ is both stationary (satisfies (1)) and satisfies the continuity equation (2).

Proof. It is clear that under this hypothesis $\frac{\partial \rho}{\partial t} = 0$ and so it only remains to show that $\{\rho, H\} = 0$. We have

$$\begin{aligned} \{\rho, H\} &= \sum_{i=1}^n \left(\frac{\partial (f \circ H)}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial (f \circ H)}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \\ &= (f' \circ H) \sum_{i=1}^n \left(\frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \\ &= 0 \end{aligned}$$

by the chain rule. □

Lemma 4 shows that many common choices of density ρ automatically result in the density having a nice relationship with the dynamics of the system. One common choice of density is the *Boltzmann distribution*

$$\rho_\beta(\mathbf{q}, \mathbf{p}) = \frac{1}{Z} e^{-\beta H(\mathbf{q}, \mathbf{p})} , \quad Z = \int_{\mathbb{R}^{2n}} e^{-\beta H(\mathbf{q}, \mathbf{p})} d\mathbf{q} d\mathbf{p}$$

where $\beta > 0$ is a free parameter. The choice of the Boltzmann distribution can be justified by a maximum entropy argument (which we will look at in more detail in a future talk).

¹Mathematically unrelated to Liouville's Theorem in complex analysis (although both are named after French mathematician Joseph Liouville) which says that every bounded holomorphic function $f : \mathbb{C} \rightarrow \mathbb{C}$ is necessarily constant.

References

- [PB11] R. K. Pathria and Paul D. Beale. *Statistical Mechanics*. 3rd ed. Butterworth-Heinemann, 2011.