Statistical Physics
Section 13: Classical Dynamics and The Arrow of Time

13. 1. The Reversibility Paradox

Figure 1: A gas is prepared in an initial macrostate with all the particles in the left hand region of the container. As time goes by the particle become evenly spread out through the container.

In the figure we sketch what happens if we prepare a gas into a state with all particles in the left hand region then let the system evolve. Of course, the gas molecules eventually become evenly distributed throughout the container. There is a definite direction to the time evolution. If we made a film of what was happening then ran the film backwards, so that initially evenly distributed particles evolved so that they all ended up in the left hand region, we would immediately know that the film was running backwards as what was happening would be contrary to our everyday experience.

Macroscopically we understand what's going on using Boltzmann's entropy

$$S = k \ln \Omega \tag{1}$$

where Ω is the weight of a macrostate. Here a macrostate can be characterised by the number of particles in the left hand region. The macrostate with the overwhelmingly largest weight is that where $N/2$ particles (to within corrections $O(N^{1/2})$) are on the left hand side. Thus the system begins in a macrosate with low weight and explores macrostates with larger and larger weight until it reaches the equilibrium (largest weight) macrostate and stays there. Then using Boltzmann's entropy (1) we obtain the second law of thermodynamics

$$\frac{dS}{dt} \geq 0 \tag{2}$$

Note that the second law clearly gives a direction to time i.e.there is no time-reversal symmetry.

On the other hand let us consider what's going on at the *microscopic* level. We assume our gas can be described by Newtonian dynamics, for example: for a system of N particles interacting via a potential U, Newton's laws are second order in time

$$m\ddot{\underline{r}}_i = -\underline{\nabla}_i U(\{\underline{r}\}) \quad i = 1...N \tag{3}$$

and accordingly have complete time reversal symmetry i.e. if we transform

$$t \to -t \quad \text{then} \quad \frac{d^2}{dt^2} \to \frac{d^2}{dt^2}$$

and equation (3) is invariant. Note that under this transformation the velocities would be reversed. Thus for each solution of the these equations, there is another equally good one in which the entire time dependence (and velocities) are reversed.

What this means is that in the figure if we could 'freeze' the final frame and reverse all the velocities, then this final microstate would evolve into the microstate on the left of the figure where all the particles are in the left hand region! The question is why is such a trajectory, which is allowed by the microscopic dynamics, never actually observed?

Less obviously, Schroedinger's equation, which is first order in time, but with a vital factor of $i = \sqrt{-1}$ in the right place, also has a time reversal symmetry: under the transformation

$$t \to -t \quad \text{and} \quad \psi \to \psi^*$$

one can see that Schroedinger's equation is invariant (more matheamtically speaking the time evolution is unitary). Thus on the surface it appears that microscopic dynamics, which have time-reversal symmetry are not consistent with our macroscopic experience and the second law!

Although there has been a lot of confusion and much written about this apparent paradox, many believe the resolution is rather simple: basically the point is that our everyday experience is at the macroscopic scale and in fact what we observe at the *macroscopic* scale need not exhibit the time reversal symmetry at the microscopic scale. In the above example, the particles, when we have reversed their velocities would indeed end up in the left hand region. But this occurs for that *particular* microstate we have prepared. For the overwhelming majority of microstates associated with the equilibrium macrostate, the time evolution will be as expected i.e. for there to be only small fluctuations in the number of particles in the left hand side. The probability of evolving to a microstate with all the particles in the left hand region is so overwhelmingly small that such an event is unlikely to have occurred in the age of the universe! A clearly written exposition of this view is J.L. Lebowitz *Boltzmann's Entropy and Times' Arrow* Physics Today, September 1993 p32–38.

The above discussion should underline the fact that the second law of thermodynamics is a *statistical* law which does not absolutely prohibit things happening but just implies that they are overwhelmingly unlikely.

13. 2. Classical dynamics

We start our study of dynamics with a classical approach. The most convenient starting point is not Newton's laws as such, but Hamilton's equations of motion. These read

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i} \qquad ; \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} \tag{4}$$

where i is a label running from 1 to $3N$ and the p_i and q_i are the momentum and position variables for our N particles. Formally, the set of variables $\{q_i\}$ are arbitrary *generalised coordinates* and $\{p_i\}$ their *conjugate momenta*. However, for a system of point particles of mass m interacting via a potential $U(\{q\}) = U(\{\underline{r}\})$, the momenta and position coordinates can be chosen as ordinary Cartesian ones. The *Hamiltonian function* is then the total energy, written in terms of the p's and q's:

$$\mathcal{H} = \sum_i \frac{p_i^2}{2m} + U(\{q\}) = E \tag{5}$$

so that Hamilton's equations of motion read

$$\dot{q}_i = \frac{p_i}{m} \qquad \dot{p}_i = -\frac{\partial U}{\partial q_i} \tag{6}$$

The first of these is the usual definition of the momentum, and the second is Newton's second law. So Hamilton's equations hold no mystery for the systems of interest here, though the results below have greater generality.

Recall the concept of **Phase Space**: this is a $6N$ dimensional space spanned by the p_i and q_i co-ordinates of the whole assembly. A point in the phase space corresponds to the microstate of the assembly. To simplify notation we denote a point in phase space by

$$\underline{X} = (q_1, \ldots q_{3N}, p_1 \ldots p_{3N}) \tag{7}$$

We consider a *probability distribution.* This is defined so that

$$\rho(\underline{X}, t)\ d\Gamma \tag{8}$$

is the probability of finding the system in some small region

$$\mathrm{d}\Gamma = \prod_{i=1}^{3N} \mathrm{d}p_i \mathrm{d}q_i \tag{9}$$

of phase space around the point $\underline{X}$.

Now recall the concept of an **Ensemble**

- We consider some large fixed number M of similar systems (obeying the same equations of motion, and not interacting with each other).
- each member of the ensemble is represented by a moving point in phase space, and ρ is the density of these *representative points.*
- $M = \int \rho(\underline{X}, t)\ d\Gamma$ is conserved. Thus representative points each separately obey the equations of motion, and cannot be created or destroyed. Accordingly, we can think of a "fluid" of representative points moving through phase space.

13. 3. Liouville's theorem

Liouville's theorem states that the fluid of representative points is *incompressible.* To prove it, we consider some region ω of phase space and let s be its surface. (These are $6N$ and $6N - 1$ dimensional entities, respectively.) Define also the $6N$ dimensional *phase space velocity* of a point $\underline{X}$ under the dynamics

$$\underline{V} = \dot{\underline{X}} = (\dot{p}_1, \dot{p}_2, ...\dot{p}_{3N}, \dot{q}_1, \dot{q}_2, ...\dot{q}_{3N}) \tag{10}$$

Then

$$-\frac{\partial}{\partial t}\int_\omega \rho\, d\Gamma = \int_s \rho \underline{V}\cdot \underline{\mathrm{d}s} \tag{11}$$

where the area element $\underline{\mathrm{ds}}$ is directed outward normals to the surface s. The left hand side of (11) is the rate of decrease in the number of representative points contained within the region ω. The right hand side is the rate at which representative points leave the region by crossing its surface. Clearly, since representative points are not created or destroyed, these two things are equal; the above equation is the *continuity equation* of the fluid. This can be rewritten, using the divergence theorem,

$$\int_s (\rho\underline{V})\cdot \underline{\mathrm{d}s} = \int_\omega \underline{\nabla}\cdot(\rho\underline{V})\mathrm{d}\Gamma \tag{12}$$

as

$$\int_\omega \left[\frac{\partial\rho}{\partial t} + \underline{\nabla}\cdot(\underline{V}\rho)\right]\mathrm{d}\Gamma = 0 \tag{13}$$

where

$$\underline{\nabla} = \left(\frac{\partial}{\partial p_1}, \frac{\partial}{\partial p_2}, \ldots \frac{\partial}{\partial p_{3N}}, \frac{\partial}{\partial q_1}, \frac{\partial}{\partial q_2}, \ldots \frac{\partial}{\partial q_{3N}}\right) \tag{14}$$

Since the region ω is arbitrary, the term in square brackets in Eq.13 must vanish everywhere. Hence we obtain the continuity equation in local form:

$$-\frac{\partial\rho}{\partial t} = \underline{\nabla}\cdot(\underline{V}\rho) \equiv \sum_{i=1}^{3N}\left[\frac{\partial}{\partial p_i}(\dot{p}_i\rho) + \frac{\partial}{\partial q_i}(\dot{q}_i\rho)\right] \tag{15}$$

$$\Rightarrow -\frac{\partial\rho}{\partial t} = \underline{V}\cdot(\underline{\nabla}\rho) + \rho(\underline{\nabla}\cdot\underline{V}) \equiv \sum_{i=1}^{3N}\left[\dot{p}_i\frac{\partial\rho}{\partial p_i} + \dot{q}_i\frac{\partial\rho}{\partial q_i}\right] + \rho\sum_{i=1}^{3N}\left[\frac{\partial\dot{p}_i}{\partial p_i} + \frac{\partial\dot{q}_i}{\partial q_i}\right] \tag{16}$$

But it follows from Hamilton's equations (check for yourself) that

$$\underline{\nabla}\cdot\underline{V} = \frac{\partial\dot{p}_i}{\partial p_i} + \frac{\partial\dot{q}_i}{\partial q_i} = 0 \tag{17}$$

and hence we have

$$-\frac{\partial\rho}{\partial t} = (\underline{V}\cdot\underline{\nabla})\rho = \sum_{i=1}^{3N}\left[\dot{p}_i\frac{\partial\rho}{\partial p_i} + \dot{q}_i\frac{\partial\rho}{\partial q_i}\right] \tag{18}$$

which may be written

$$\boxed{\frac{\mathrm{d}\rho}{\mathrm{d}t} \equiv \frac{\partial\rho}{\partial t} + (\underline{V}\cdot\underline{\nabla})\rho = 0} \tag{19}$$

$\frac{\mathrm{d}}{\mathrm{d}t}$ is a total time derivative, here meaning a time-derivative that moves with the fluid of representative points. This is *Liouville's equation.* It shows that the motion of representative points in phase space is not just that of a fluid, but that of an *incompressible* fluid. In other words, whatever the density of these points in the neighbourhood of a given one, they move together in such a way that this is constant in time. Thus, (19) means

> The local density ρ as seen by an observer moving with a representative point is constant in time

Above we just stated that the terms in (19) comprise the total time derivative. More generally (see tutorial) we can show that any well-behaved function u of the canonical variables $\{q_i, p_i\}$ satisfies

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + [u, \mathcal{H}]$$

where $[u, \mathcal{H}]$ is the Poisson bracket of u with $\mathcal{H}$ and we can write Liouville's equation in the form (see tutorial)

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{\partial \rho}{\partial t} + [\rho, \mathcal{H}] = 0$$

Finally we note that to have *stationary* distribution

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

which is a necessary condition for equilibrium, we require that $[\rho, H] = 0$. Choosing $\rho =$ *constant* or $\rho = \rho(\mathcal{H})$ recovers the microcanonical and canonical ensembles respectively (see tutorial).

13. 4. The entropy paradox

If we now choose ρ to be normalized to unity (rather than M), it becomes once again the probability density for a single system, rather than the density of representative points in an ensemble. Clearly it obeys the same equation as before: the probability density evolves incompressibly. The Gibbs entropy is defined in terms of probabilities as

$$S_{\text{Gibbs}} = -k \sum p_i \ln p_i \tag{20}$$

which, in the classical limit, approaches

$$S_{\text{Gibbs}} = -k \int d\Gamma \, \rho \ln(\rho) + \text{const.} \tag{21}$$

where the integral is over all phase space.

[The additive constant contains terms like $\ln(h^{-3N})$ which come from the conversion factor between a smooth density in phase space and a sum over discrete states. These do not affect the present discussion.]

If we take S_{Gibbs} as our definition of S, then it follows from Liouville's theorem (see tutorial 8.1) that

$$\frac{\partial S_{\text{Gibbs}}}{\partial t} = -k \frac{\partial}{\partial t} \int d\Gamma \, \rho \ln \rho = 0 \tag{22}$$

So, the Gibbs entropy is constant in time, rather than increasing. This seems to imply that reversible microphysics (from which Liouville's equation starts) is incompatible with irreversible macrophysics (the second law).

13. 5. Coarse graining

Suppose we have a classical system (or ensemble) whose probability density (or density of representative points) is localized in some region. Then intuitively, one would expect this to spread out with time, and cause an increase in S. Liouville's theorem says this is not true. What actually happens, however, is that although the density moves incompressibly, it spreads into a complicated region of extreme threadiness:

This is very much like stirring ink into water: in principle, there is always ink in one place and water in another (so the local density of ink has never changed from its initial value). However, the regions are so intertwined that any sensible person would describe the ink as having a uniform density, lower than the original one.

Adopting this attitude, we can demand that the Gibbs entropy be redefined as

$$\bar{S} = -k \int d\Gamma\, \bar{\rho} \ln \bar{\rho} \tag{23}$$

where $\bar{\rho}$ is a *coarse grained* probability density, defined by averaging ρ over some fixed local scale Λ in phase space.

It is obvious that this coarse-graining procedure will tend to smooth out the local probability density, not sharpen it up. Moreover, such a smoothing operation can only increase $\bar{S}$. To see this, note that the function $s(\rho) = -k\rho \ln \rho$ is *concave* (tutorial). So if we consider adjacent patches of equal volume w in phase space, with local densities a and b and entropies $s_1 = ws(a)$ and $s_2 = ws(b)$, and merge them to give a single patch of volume $2w$ and density $(a+b)/2$, we have

$$\bar{s} = 2ws((a+b)/2) \geq w[s(a) + s(b)] = s_1 + s_2 \tag{24}$$

The geometrical statement of this is as follows: i.e.

Figure 2: Property of concave function

For a concave function: The function of the average $\geq$ the average of the function. For a convex function the opposite is true.

The same proof generalises easily to *any* form of local averaging operation.

It follows that if we define a fixed coarse-graining length Λ and, as time proceeds, keep averaging the density so that no details below this scale remain, the resulting entropy $\bar{S}$ can only increase with time:

$$\frac{d\bar{S}}{dt} \geq 0 \tag{25}$$

which is the second law of thermodynamics. In the present context it is sometimes called (for historical reasons) "the H theorem". There are many more complicated proofs (see textbooks) all of which reduce eventually to the above statement about the concavity of the function $s(\rho)$.

13. 6. Discussion

Actually the result of section 13.4 could have been deduced without any calculation since we know the Gibbs entropy represents the missing information about the assembly. Now since we consider deterministic dynamics which are reversible it is clear that by running the dynamics of a representative point we don't lose any information about that point. Consequently there is no increase in the missing information about the assembly and the Gibbs entropy remains constant.

So the problem of the arrow of time and the second law really boils down to a problem with the Gibbs entropy and how it differs from the Boltzmann entropy in out of equilibrium situations.

In the coarse-graining process, we do explicitly throw away information; only by doing so is the Gibbs entropy allowed to increase. The Boltzmann entropy on the other hand is inherently of a 'coarse-grained' nature since it is concerned with macrostates specified by macroscopic variables.

Therefore the increase of entropy is linked to *our knowledge about the system*, rather than anything it is doing internally, in a manner that may appear dubious. Can it be possible that macroscopic and reproducible phenomena, such as heat flow, depend on *how we handle information*? Perhaps yes, since the division between work and heat is somewhat arbitrary. Were we able to track all the particle positions, there would be no need to talk about heat energy, or heat flow.

As another example of 'coarse-graining' leading to irreversbility, consider Newton's law with a viscous term

$$m\ddot{\underline{r}}_i + \kappa\dot{\underline{r}}_i = -\underline{\nabla}_i U(\{\underline{r}\}) \quad i = 1...N \tag{26}$$

The presence of the 1st order time derivative implies the system does not have time-reversal symmetry and aslo implies that system dissipates energy. But the viscous drag on a particle is a result of complicated processes involving molecular collisions etc and in principle we could write out all these other processes using microscopic time-reversible equations of form (3) (and keeping energy conserved). The viscous term in (26) approximates microscopic reversible processes by a 'coarse-grained' description.