III.E The H–Theorem and Irreversibility

The second question posed at the beginning of this chapter was whether a collection of particles naturally evolves towards an equilibrium state. While it is possible to obtain steady state solutions for the full phase space density ρ_N, because of time reversal symmetry these solutions are not attractors of generic non-equilibrium densities. Does the unconditional one particle PDF ρ_1, suffer the same problem? While the exact density ρ_1 must necessarily reflect this property of ρ_N, the H–theorem proves that an approximate ρ_1, governed by the Boltzmann equation, does in fact non-reversibly approach an equilibrium form. This theorem states that:

- If $f_1(\vec{p}, \vec{q}, t)$ satisfies the Boltzmann equation, then $d\mathrm{H}/dt \leq 0$, where

$$\mathrm{H}(t) = \int d^3\vec{p}\, d^3\vec{q}\ f_1(\vec{p}, \vec{q}, t) \ln f_1(\vec{p}, \vec{q}, t) \quad . \tag{III.42}$$

The function $\mathrm{H}(t)$ is related to the information content of the one particle PDF. Up to an overall constant, the information content of $\rho_1 = f_1/N$ is given by $I[\rho_1] = \langle \ln \rho_1 \rangle$, which is clearly similar to $\mathrm{H}(t)$.

Proof: The time derivative of H is

$$\frac{d\mathrm{H}}{dt} = \int d^3\vec{p}_1 d^3\vec{q}_1\ \frac{\partial f_1}{\partial t} (\ln f_1 + 1) = \int d^3\vec{p}_1 d^3\vec{q}_1\ \ln f_1 \frac{\partial f_1}{\partial t} \quad , \tag{III.43}$$

since $\int dV_1 f_1 = N \int d\Gamma \rho = N$ is time independent. Using eq.(III.41), we obtain

$$\frac{d\mathrm{H}}{dt} = \int d^3\vec{p}_1 d^3\vec{q}_1\ \ln f_1 \left(\frac{\partial U}{\partial \vec{q}_1} \cdot \frac{\partial f_1}{\partial \vec{p}_1} - \frac{\vec{p}_1}{m} \cdot \frac{\partial f_1}{\partial \vec{q}_1} \right)$$
$$- \int d^3\vec{p}_1 d^3\vec{q}_1 d^3\vec{p}_2 d^2\sigma |\vec{v}_1 - \vec{v}_2| \left[f_1(\vec{p}_1, \vec{q}_1) f_1(\vec{p}_2, \vec{q}_1) - f_1(\vec{p}_1{}', \vec{q}_1) f_1(\vec{p}_2{}', \vec{q}_1) \right] \ln f_1(\vec{p}_1, \vec{q}_1), \tag{III.44}$$

where we shall interchangeably use $d^2\sigma$, $d^2\vec{b}$, or $d^2\Omega |d\sigma/d\Omega|$ for the differential cross-section. The streaming terms in the above expression are zero, as shown through successive integrations by part,

$$\int d^3\vec{p}_1 d^3\vec{q}_1\ \ln f_1 \frac{\partial U}{\partial \vec{q}_1} \cdot \frac{\partial f_1}{\partial \vec{p}_1} = -\int d^3\vec{p}_1 d^3\vec{q}_1 f_1\ \frac{\partial U}{\partial \vec{q}_1} \cdot \frac{1}{f_1} \frac{\partial f_1}{\partial \vec{p}_1} = \int d^3\vec{p}_1 d^3\vec{q}_1 f_1 \frac{\partial}{\partial \vec{p}_1} \cdot \frac{\partial U}{\partial \vec{q}_1} = 0,$$

and

$$\int d^3\vec{p}_1 d^3\vec{q}_1\ \ln f_1 \frac{\vec{p}_1}{m} \cdot \frac{\partial f_1}{\partial \vec{q}_1} = -\int d^3\vec{p}_1 d^3\vec{q}_1\ f_1 \frac{\vec{p}_1}{m} \cdot \frac{1}{f_1} \frac{\partial f_1}{\partial \vec{q}_1} = \int d^3\vec{p}_1 d^3\vec{q}_1 f_1 \frac{\partial}{\partial \vec{q}_1} \cdot \frac{\vec{p}_1}{m} = 0.$$

The collision term in eq.(III.44) involves integrations over dummy variables $\vec{p}_1$ and $\vec{p}_2$. The labels (1) and (2) can thus be exchanged without any change in the value of the integral. Averaging the resulting two expressions gives

$$\frac{d\mathrm{H}}{dt} = -\frac{1}{2} \int d^3\vec{q}\, d^3\vec{p}_1\, d^3\vec{p}_2\, d^2\vec{b}\, |\vec{v}_1 - \vec{v}_2|\, [f_1(\vec{p}_1)f_1(\vec{p}_2) - f_1(\vec{p}_1{}')f_1(\vec{p}_2{}')]\ln\left(f_1(\vec{p}_1)f_1(\vec{p}_2)\right).$$
(III.45)

(The arguments, $\vec{q}$ and t, of f_1 are suppressed for ease of notation.) We would now like to change the variables of integrations from the coordinates describing the initiators of the collision, $(\vec{p}_1, \vec{p}_2, \vec{b})$, to those of their products, $(\vec{p}_1{}', \vec{p}_2{}', \vec{b}')$. The explicit functional forms describing this transformation are complicated because of the dependence of the solid angle $\hat{\Omega}$ in eq.(III.39) on $\vec{b}$ and $|\vec{p}_2 - \vec{p}_1|$. However, we are assured that the Jacobian of the transformation is unity because of time reversal symmetry; since for every collision there is an inverse one obtained by reversing the momenta of the products. In terms of the new coordinates

$$\frac{d\mathrm{H}}{dt} = -\frac{1}{2} \int d^3\vec{q}\, d^3\vec{p}_1{}'\, d^3\vec{p}_2{}'\, d^2\vec{b}'\, |\vec{v}_1 - \vec{v}_2|\, [f_1(\vec{p}_1)f_1(\vec{p}_2) - f_1(\vec{p}_1{}')f_1(\vec{p}_2{}')]\ln\left(f_1(\vec{p}_1)f_1(\vec{p}_2)\right),$$
(III.46)

where we should now regard $(\vec{p}_1, \vec{p}_2)$ in the above equation, as functions of the integration variables $(\vec{p}_1{}', \vec{p}_2{}', \vec{b}')$ as in eq.(III.39). As noted earlier, $|\vec{v}_1 - \vec{v}_2| = |\vec{v}_1{}' - \vec{v}_2{}'|$ for any elastic collision, and we can use these quantities interchangeably. Finally, we relabel the dummy integration variables such that the primes are removed. Noting that the functional dependence of $(\vec{p}_1, \vec{p}_2, \vec{b})$ on $(\vec{p}_1{}', \vec{p}_2{}', \vec{b}')$ is exactly the same as its inverse, we obtain

$$\frac{d\mathrm{H}}{dt} = -\frac{1}{2} \int d^3\vec{q}\, d^3\vec{p}_1\, d^3\vec{p}_2\, d^2\vec{b}\, |\vec{v}_1 - \vec{v}_2|\, [f_1(\vec{p}_1{}')f_1(\vec{p}_2{}') - f_1(\vec{p}_1)f_1(\vec{p}_2)]\ln\left(f_1(\vec{p}_1{}')f_1(\vec{p}_2{}')\right).$$
(III.47)

Averaging eqs.(III.45) and (III.47) results in

$$\frac{d\mathrm{H}}{dt} = -\frac{1}{4} \int d^3\vec{q}\, d^3\vec{p}_1\, d^3\vec{p}_2\, d^2\vec{b}\, |\vec{v}_1 - \vec{v}_2|\, [f_1(\vec{p}_1)f_1(\vec{p}_2) - f_1(\vec{p}_1{}')f_1(\vec{p}_2{}')]$$
$$[\ln\left(f_1(\vec{p}_1)f_1(\vec{p}_2)\right) - \ln\left(f_1(\vec{p}_1{}')f_1(\vec{p}_2{}')\right)].$$
(III.48)

The integrand of the above expression is always positive. If $f_1(\vec{p}_1)f_1(\vec{p}_2) > f_1(\vec{p}_1{}')f_1(\vec{p}_2{}')$, both terms in square brackets are positive, while both are negative if $f_1(\vec{p}_1)f_1(\vec{p}_2) < f_1(\vec{p}_1{}')f_1(\vec{p}_2{}')$. In either case, their product is positive. The positivity of the integrand establishes the validity of the H–theorem,

$$\frac{d\mathrm{H}}{dt} \leq 0 \quad .$$
(III.49)

• *Irreversibility:* The second law is an *empirical* formulation of the vast number of everyday observations which support the existence of an *arrow of time.* Reconciling the reversibility of laws of physics governing the microscopic domain with the observed irreversibility of macroscopic phenomena is a fundamental problem. Of course, not all microscopic laws of physics are reversible: weak nuclear interactions violate time reversal symmetry, and the collapse of the quantum wave-function in the act of observation is irreversible. The former interactions in fact do not play any significant role in everyday observations that lead to the second law. The irreversible collapse of the wave-function may itself be an artifact of treating macroscopic observers and microscopic observables distinctly.[†] There are proponents of the view that the reversibility of the currently accepted microscopic equations of motion (classical or quantum) is indicative of their inadequacy. However, the advent of powerful computers has made it possible to simulate the evolution of collections of large numbers of particles, governed by *classical, reversible* equations of motion. Although simulations are currently limited to relatively small numbers of particles (10^6), they do exhibit the irreversible macroscopic behaviors similar to those observed in nature (typically involving 10^{23} particles). For example, particles initially occupying one half of a box proceed to irreversibly, and uniformly, occupy the whole box. (This has nothing to do with limitations of computational accuracy; the same macroscopic irreversibility is observed in exactly reversible integer based simulations, such as with cellular automata.) Thus the origin of the observed irreversibilities should be sought in the classical evolution of large collections of particles.

The Boltzmann equation is the first formula we have encountered that is clearly not time reversible, as indicated by eq.(III.49). We can thus ask the question of how we obtained this result from the Hamiltonian equations of motion. The key to this, of course, resides in the physically motivated approximations used to obtain eq.(III.41). The first steps of the approximation were dropping the three body collision term on the right hand side of eq.(III.30), and the implicit coarse–graining of the resolution in the spatial and temporal scales. Neither of these steps explicitly violates time reversal symmetry, and the collision term in eq.(III.37) retains this property. The next step in getting to eq.(III.41) is to replace the two–body density $f_2(-)$, *evaluated before the collision*, with the product of two one body densities according to eq.(III.32). This treats the two body densities before

---

[†] The time dependent Schrödinger equation is fully time reversible. If it is possible to write a complicated wave-function that includes the observing apparatus (possibly the whole universe), it is hard to see how any irreversibility may occur.

and after the collision differently. We could have alternatively expressed eq.(III.37) in terms of the two body densities $f_2(+)$ *evaluated after the collision.* Replacing $f_2(+)$ with the product of two one particle densities would then lead to the opposite conclusion, with $d\mathrm{H}/dt \geq 0$! For a system in equilibrium, it is hard to justify one choice over the other. However, once the system is out of equilibrium, the coordinates after the collision are more quite likely to be correlated, and hence the substitution of eq.(III.32) for $f_2(+)$ does not make sense. Time reversal symmetry implies that there should also be subtle correlations in $f_2(-)$ which are ignored in the so-called *assumption of molecular chaos.*

While the assumption of molecular chaos before (but not after) collisions is the key to the irreversibility of the Boltzmann equation, the resulting loss of information is best justified in terms of the coarse graining of space and time: The Liouville equation and its descendants contain precise information about the evolution of a pure state. This information, however, is inevitably transported to shorter scales. A useful image is that of mixing two immiscible fluids. While the two fluids remain distinct at each point, the transitions in space from one to the next occur at finer resolution on subsequent mixing. At some point, a finite resolution in any measuring apparatus will prevent keeping track of the two components. In the Boltzmann equation the precise information of the pure state is lost at the scale of collisions. The resulting one body density only describes space and time resolutions longer than those of a two-body collision, becoming more and more probabilistic as further information is lost.

III.F Equilibrium Properties

What is the nature of the equilibrium state described by f_1, for a homogeneous gas? **(1)** *The equilibrium distribution:* After the gas has reached equilibrium, the function H should no longer decrease with time. Since the integrand in eq.(III.48) is always positive, a *necessary* condition for $d\mathrm{H}/dt = 0$ is that

$$f_1(\vec{p}_1, \vec{q}_1)f_1(\vec{p}_2, \vec{q}_1) - f_1(\vec{p}_1\,', \vec{q}_1)f_1(\vec{p}_2\,', \vec{q}_1) = 0, \qquad (\text{III.50})$$

i.e. *at each point $\vec{q}$,* we must have

$$\ln f_1(\vec{p}_1, \vec{q}) + \ln f_1(\vec{p}_2, \vec{q}) = \ln f_1(\vec{p}_1\,', \vec{q}) + \ln f_1(\vec{p}_2\,', \vec{q}). \qquad (\text{III.51})$$

The left hand side of the above equation refers to the momenta before a two-body collision, and the right hand side to the those after the collision. The equality is thus satisfied by

any *additive* quantity that is conserved during the collision. There are 5 such conserved quantities for an elastic collision: the particle number, the three components of the net momentum, and the kinetic energy. Hence, a general solution for f_1 is

$$\ln f_1 = a(\vec{q}) - \vec{\alpha}(\vec{q}) \cdot \vec{p} - \beta(\vec{q}) \left(\frac{\vec{p}^{\,2}}{2m} \right). \tag{III.52}$$

We can easily accomodate the potential energy $U(\vec{q})$ in the above form, and set

$$f_1(\vec{p}, \vec{q}) = \mathcal{N}(\vec{q}) \exp \left[-\vec{\alpha}(\vec{q}) \cdot \vec{p} - \beta(\vec{q}) \left(\frac{p^2}{2m} + U(\vec{q}) \right) \right]. \tag{III.53}$$

We shall refer to the above distribution as describing *local equilibrium*. While this form is preserved *during collisions*, it will evolve in time *away from collisions*, due to the streaming terms, unless $\{\mathcal{H}_1, f_1\} = 0$. The latter condition is satisfied for any function f_1 that depends only on $\mathcal{H}_1$, or any other quantity that is conserved by it. Clearly, the above density satisfies this requirement as long as $\mathcal{N}$, and β are independent of $\vec{q}$, and $\vec{\alpha} = 0$.

According to eq.(III.16), the appropriate normalization for f_1 is

$$\int d^3\vec{p}\, d^3\vec{q}\, f_1(\vec{p}, \vec{q}) = N. \tag{III.54}$$

For particles in a box of volume V, the potential $U(\vec{q})$ is zero inside the box, and infinite on the outside. The normalization factor in eq.(III.53) can be obtained from eq.(III.54) as

$$N = \mathcal{N}V \left[\int_{-\infty}^{\infty} dp_i \exp \left(-\alpha_i p_i - \frac{\beta p_i^2}{2m} \right) \right]^3 = \mathcal{N}V \left(\frac{2\pi m}{\beta} \right)^{3/2} \exp \left(\frac{m\alpha^2}{2\beta} \right). \tag{III.55}$$

Hence, the properly normalized Gaussian distribution for momenta is

$$f_1(\vec{p}, \vec{q}) = n \left(\frac{\beta}{2\pi m} \right)^{3/2} \exp \left[-\frac{\beta(\vec{p} - \vec{p}_0)^2}{2m} \right], \tag{III.56}$$

where $\vec{p}_0 = \langle \vec{p} \rangle = m\vec{\alpha}/\beta$ is the mean value for the momentum of the gas, which is zero for a stationary box, and $n = N/V$ is the particle density. From the Gaussian form of the distribution it can be easily concluded that the variance of each component of the momentum is $\langle p_i^2 \rangle = m/\beta$, and

$$\langle p^2 \rangle = \langle p_x^2 + p_y^2 + p_z^2 \rangle = \frac{3m}{\beta} \quad . \tag{III.57}$$

(2) *Equilibrium between two gases:* Consider two different gases (a) and (b), moving in the same potential U, and subject to a two-body interaction $\mathcal{V}_{ab}\left(\vec{q}^{\,(a)} - \vec{q}^{\,(b)}\right)$. We can define one-particle densities, $f_1^{(a)}$, and, $f_1^{(b)}$, for the two gases respectively. In terms of a generalized collision integral

$$C_{\alpha,\beta} = -\int d^3\vec{p}_2 d^2\Omega \left|\frac{d\sigma_{\alpha,\beta}}{d\Omega}\right| |\vec{v}_1 - \vec{v}_2| \left[f_1^{(\alpha)}(\vec{p}_1, \vec{q}_1) f_1^{(\beta)}(\vec{p}_2, \vec{q}_1) - f_1^{(\alpha)}(\vec{p}_1{}', \vec{q}_1) f_1^{(\beta)}(\vec{p}_2{}', \vec{q}_1) \right],$$
$$\text{(III.58)}$$

the evolution of these densities is governed by a simple generalization of the Boltzmann equation to

$$\begin{cases} \dfrac{\partial f_1^{(a)}}{\partial t} = -\left\{ f_1^{(a)}, \mathcal{H}_1^{(a)} \right\} + C_{a,a} + C_{a,b} \\[4mm] \dfrac{\partial f_1^{(b)}}{\partial t} = -\left\{ f_1^{(b)}, \mathcal{H}_1^{(b)} \right\} + C_{b,a} + C_{b,b} \end{cases} \qquad \text{(III.59)}$$

Stationary distributions can be obtained if all six terms on the right hand side of eqs.(III.59) are zero. In the absence of inter-species collisions, i.e. for $C_{a,b} = C_{b,a}$, we can obtain independent stationary distributions $f_1^{(a)} \propto \exp\left(-\beta_a \mathcal{H}_1^{(a)}\right)$ and $f_1^{(b)} \propto \exp\left(-\beta_b \mathcal{H}_1^{(b)}\right)$. Requiring the vanishing of $C_{a,b}$ leads to the additional constraint,

$$f_1^{(a)}(\vec{p}_1) f_1^{(b)}(\vec{p}_2) - f_1^{(a)}(\vec{p}_1{}') f_1^{(b)}(\vec{p}_2{}') = 0, \implies$$
$$\beta_a \mathcal{H}_1^{(a)}(\vec{p}_1) + \beta_b \mathcal{H}_1^{(b)}(\vec{p}_2) = \beta_a \mathcal{H}_1^{(a)}(\vec{p}_1{}') + \beta_b \mathcal{H}_1^{(b)}(\vec{p}_2{}') \qquad \text{(III.60)}$$

Since the total energy $\mathcal{H}_1^{(a)} + \mathcal{H}_1^{(b)}$ is conserved in a collision, the above equation can be satisfied for $\beta_a = \beta_b = \beta$. From eq.(III.57) this condition implies the equality of the kinetic energies of the two species,

$$\left\langle \frac{p_a^2}{2m_a} \right\rangle = \left\langle \frac{p_b^2}{2m_b} \right\rangle = \frac{3}{2\beta} \qquad \text{(III.61)}$$

The parameter β thus plays the role of an *empirical temperature* describing the equilibrium of gases.

(3) *The equation of state:* To complete the identification of β with temperature T, consider a gas of N particles confined to a box of volume V. The gas pressure results from the force exerted by the particles colliding with the walls of the container. Consider a wall element of area A perpendicular to the x direction. The number of particles impacting this area, with momenta in the interval $[\vec{p}, \vec{p} + d\vec{p}]$, over a time period δt, is

$$d\mathcal{N}(\vec{p}) = \left(f_1(\vec{p}) d^3\vec{p} \right)\left(A\, v_x\, \delta t \right) \qquad \text{(III.62)}$$

The final factor in the above expression is the volume of a cylinder of height $v_x \delta t$ perpendicular to the area element A. Only particles within this cylinder are close enough to impact the wall during δt. As each collision imparts a momentum $2p_x$ to the wall, the net force exerted is

$$F = \frac{1}{\delta t} \int_{-\infty}^{0} dp_x \int_{-\infty}^{\infty} dp_y \int_{-\infty}^{\infty} dp_z f_1(\vec{p}) \left(A \frac{p_x}{m} \delta t \right) (2p_x). \qquad \text{(III.63)}$$

As only particles with velocities directed towards the wall will hit it, the first integral is over half of the range of p_x. Since the integrand is even in p_x, this restriction can be removed by dividing the full integral by 2. The pressure P is then obtained from the force per unit area as

$$P = \frac{F}{A} = \int d^3\vec{p} \, f_1(\vec{p}) \frac{p_x^2}{m} = \frac{1}{m} \int d^3\vec{p} \, p_x^2 \, n \left(\frac{\beta}{2\pi m} \right)^{3/2} \exp\left(-\frac{\beta p^2}{2m} \right) = \frac{n}{\beta} \quad , \qquad \text{(III.64)}$$

where eq.(III.56) is used for the equilibrium form of f_1. Comparing with the standard equation of state, $PV = Nk_B T$, for an ideal gas, leads to the identification, $\beta = 1/k_B T$.

(4) *Entropy:* As discussed earlier, the Boltzmann H-function is closely related to the information content of the one-particle PDF ρ_1. We can also define a corresponding Boltzmann entropy,

$$S_B(t) = -k_B \mathrm{H}(t), \qquad \text{(III.65)}$$

where the constant k_B reflects the historical origins of entropy. The H-theorem implies that S_B can only increase with time in approaching equilibrium. It has the further advantage of being defined through eq.(III.42) for situations that are clearly out of equilibrium. For a gas in equilibrium in a box of volume V, from eq.(III.56), we compute

$$\begin{aligned}
\mathrm{H} &= V \int d^3\vec{p} \, f_1(\vec{p}) \ln f_1(\vec{p}) \\
&= V \int d^3\vec{p} \, \frac{N}{V} (2\pi m k_B T)^{-3/2} \exp\left(-\frac{p^2}{2m k_B T} \right) \left[\ln\left(\frac{n}{(2\pi m k_B T)^{3/2}} \right) - \frac{p^2}{2m k_B T} \right] \\
&= N \left[\ln\left(\frac{n}{(2\pi m k_B T)^{3/2}} \right) - \frac{3}{2} \right].
\end{aligned} \qquad \text{(III.66)}$$

The entropy is now identified as

$$S_B = -k_B \mathrm{H} = Nk_B \left[\frac{3}{2} + \frac{3}{2} \ln (2\pi m k_B T) - \ln\left(\frac{N}{V} \right) \right]. \qquad \text{(III.67)}$$

The thermodynamic relation, $TdS_B = dE + PdV$, implies

$$\begin{aligned}
\left.\frac{\partial E}{\partial T}\right|_V &= T \left.\frac{\partial S_B}{\partial T}\right|_V = \frac{3}{2}Nk_B, \\
P + \left.\frac{\partial E}{\partial V}\right|_T &= T \left.\frac{\partial S_B}{\partial V}\right|_T = \frac{Nk_BT}{V}
\end{aligned} \qquad . \tag{III.68}$$

The usual properties of a monatomic ideal gas, $PV = Nk_BT$, and $E = 3Nk_BT/2$, can now be obtained from the above equations. Also note that for this classical gas, the zero temperature limit of the entropy in eq.(III.67) *is not* independent of the density n, in violation of the third law of thermodynamics.