

STATISTICAL MECHANICS OF GRAVITATING SYSTEMS

T. PADMANABHAN

Theoretical Astrophysics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Received August 1989

Contents:

1. Introduction and summary	287	Appendix 4.1. The variational problem for the entropy	330
2. Statistical equilibrium of gravitating systems: overview	288	Appendix 4.2. Stability of small isothermal spheres	332
2.1. Introduction	288	5. The approach to equilibrium: collisional relaxation	333
2.2. Fundamentals of the statistical description	289	5.1. Introduction	333
2.3. Difficulties due to the long range nature of gravity	293	5.2. The description of dynamical evolution	334
2.4. Effects due to the short distance behaviour of $V(r)$	294	5.3. Relaxation time for "hard" and "soft" collisions	335
3. Statistical equilibrium of gravitating systems: toy models	297	5.4. The nature of the fluctuating forces	337
3.1. Introduction	297	5.5. Fokker–Planck description of the collisional evolution	342
3.2. Statistical description of a binary star	298	5.6. A critical look at the derivation of the Fokker–Planck equation	347
3.3. The Lynden-Bells' and Thirring's models	304	6. The approach to equilibrium: collisionless relaxation	353
3.4. Approximate analysis of gravitating hard spheres	306	6.1. Introduction	353
3.5. Two dimensional gravity	307	6.2. The process of violent relaxation	353
4. Mean field equilibrium of gravitating systems	310	6.3. Constraints on collisionless evolution	356
4.1. Introduction	310	6.4. Dynamical description of collisionless relaxation	358
4.2. Mean field as the saddle point limit	310	7. Conclusions	360
4.3. Mean field equilibrium: the isothermal sphere	313	References	361
4.4. Stability of the mean field solution	318		
4.5. Isothermal sphere with a short distance cutoff	324		
4.6. Mean field thermodynamics and comparison with toy models	329		

Abstract:

The statistical description of a system containing a large number of particles which interact via Newtonian gravity is discussed. Such a system exhibits several peculiarities due to the long range, unshielded nature of the gravitational interaction. These features are illustrated using specific examples. Conceptual issues which are still unsettled are highlighted.

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 188, No. 5 (1990) 285–362.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 58.00, postage included.

STATISTICAL MECHANICS OF GRAVITATING SYSTEMS

T. PADMANABHAN

*Theoretical Astrophysics Group, Tata Institute of Fundamental Research, Homi Bhabha Road,
Bombay 400 005, India*



NORTH-HOLLAND

1. Introduction and summary

Consider a system of N particles interacting through Newtonian gravitational forces alone. The description and physical properties of such a system will clearly depend on how large N is. For $N = 2$ we have an exactly solvable Kepler problem. For N in the range of, say, 3–50, the gravitational few body problem can be tackled by a computer. For considerably larger N , say $N = 10^5$ – 10^{11} , it is neither feasible nor useful to try to follow the trajectories of the individual particles. We will be more interested in the average, statistical properties of such a system rather than in the individual orbits. The statistical mechanics of such a system of gravitating particles is the subject of this review.

The physical behaviour of such a system turns out to be very different from the statistical behaviour of other, more familiar, many body systems like neutral gases and plasmas. The central reason for the peculiarities exhibited by gravitating systems can be traced back to the *unshielded, long range* nature of the gravitational force. In contrast, gaseous systems have genuinely short range interactions and plasmas have an effective short range interaction due to Debye shielding. Because of this fundamental difference between these systems, the standard methods of statistical mechanics cannot be carried over in a direct manner to study gravitating systems. It is necessary to go back to the fundamentals of statistical mechanics and develop special techniques capable of handling the long range nature of gravity.

A first step in that direction will be to study specific problems in gravitational statistical mechanics and identify the common features. Hopefully, these common features will give us an indication as to how a more comprehensive and systematic theory of gravitational plasma physics can be developed. We have tried to present in this review such a discussion of several specific problems in this subject.

The review contains six more chapters. Chapters 2, 3 and 4 discuss the description and properties of the statistical *equilibrium* of a gravitating system while chapters 5 and 6 study the issues related to the *approach* towards statistical equilibrium. Chapter 7 presents a brief summary. The emphasis is on conceptual and theoretical issues and on questions which are still somewhat unsettled. We give below a brief summary of the rest of the review.

Chapter 2 begins by recollecting the standard concepts of statistical mechanics and discusses how some of these concepts get modified in the presence of long range interactions. Most important among these is the extensive nature of the energy, which does not hold good in the presence of long range interactions. This has the consequence that gravitating systems must be described by the microcanonical distribution. The canonical and microcanonical distributions will not lead to the same physical description for systems with long range forces. We also study the dependence of the ground state energy of the system on the number of particles in the system and discuss why gravitating systems do not possess the thermodynamic limit in the conventional sense of that term.

The issues discussed in chapter 2 are illustrated by some simple toy models in chapter 3. We construct the “statistical mechanics” based on a simple Hamiltonian, describing two particles of finite size confined inside a box and interacting via gravity. This system exhibits several important properties of more complicated gravitating systems in spite of the fact that it has only two degrees of freedom! In particular, this system exhibits the following two features, which seem to be generic to all gravitating systems: (1) When studied using the microcanonical ensemble – which, as we said before, is the proper distribution to use for gravitating systems – the system shows evidence for two different phases: a high temperature phase, dominated by kinetic energy, and a low temperature phase dominated by the potential energy and stabilised by some short distance cutoff which is of non-gravitational origin. Both these phases have positive specific heat. These two phases are connected at intermediate temperatures

by a region of negative specific heat; this is precisely the range in which the kinetic and potential energies of the system are comparable and the system is in virial equilibrium. (2) If the same system is studied using the canonical ensemble, the intermediate region of negative specific heat is replaced by a sharp phase transition releasing a large amount of latent heat. This suggests the following analogy: Gravitating systems in virial equilibrium are similar to normal systems (with short range forces) at the verge of a phase transition. In fact, this is the central idea which is emphasised in chapters 3, 4. We also study in chapter 3 a few other toy models to illustrate different aspects of the statistical description.

In chapter 4, we study the equilibrium of a gravitating system in the mean field limit after first deriving the conditions under which such a mean field description is valid. In the absence of any short distance cutoff to gravitational interaction, the mean field solution is given by an isothermal sphere. Several peculiar features of this equilibrium configuration (like the bounds on physical variables, Antonov instability etc.) are discussed in detail. This analysis reveals the need for a short distance cutoff if the equilibrium is to be physically meaningful. We introduce such a cutoff and study the properties of the resulting system, contrasting it with the original isothermal sphere. It turns out that this system is *remarkably similar* to the simple toy model studied in chapter 3. In fact, this mean field analysis confirms all the conjectures made earlier based on the toy models.

Chapters 5 and 6 take up the description of gravitating systems as they evolve from some initial configuration towards equilibrium. We present a careful discussion of the concept of “soft” collisions, which arise in systems with $1/r$ potential and the difficulties which they lead to. The Fokker–Planck equation describing the diffusion in velocity space is derived in the compact Landau form. Lastly we present a critique of the conventional approaches towards the derivation of equations describing the collisional evolution. It appears that a more meaningful approach could be based on the quasilinear approximations used in the plasma physics. The ingredients of such a derivation are discussed.

Chapter 6 discusses in a brief manner the concept of violent relaxation. We stress the conceptual issues which are still unsettled and emphasise the need for a dynamical description of this process. The quasilinear approximation may again be useful and its role is discussed.

No review of this kind can hope to cover all the relevant aspects of the topic and the present one is no exception. Three interesting topics which we have not discussed in this review are: (a) The solutions to the collisionless Boltzmann equation describing the “equilibrium” of gravitating systems like galaxies etc. (b) The effects of binary formation on the evolution of gravitating systems. (c) The effects due to cosmology in the form of an expanding background metric. These topics are covered in detail in three recent books by Binney and Tremaine [1987], Spitzer [1987] and Saslaw [1985].

2. Statistical equilibrium of gravitating systems: overview

2.1. Introduction

Equilibrium statistical mechanics of gaseous systems and plasmas is well understood. It is possible to develop this theory, in a reasonably straightforward manner, using certain general principles of statistical mechanics (see, e.g., Lifshitz and Pitaevskii [1982]). Unfortunately, these methods do not lead to an equally useful description when applied to a system of particles interacting via gravity. The main source of trouble happens to be the long range nature of the gravitational interaction.

The central difficulty can be appreciated through the following, somewhat intuitive, analogy. Systems normally encountered in the laboratory interact through short range molecular potentials, and can

usually exist in three different phases: as gas, liquid or solid. At high enough temperatures, the kinetic energy of the molecules will dominate over the potential energy of mutual interaction and the system will be in the gaseous phase. As the temperature is lowered, a stage will be reached when the potential and kinetic energies are comparable. The system will then make a transition to a more condensed phase, say the liquid phase. It is well known that the statistical behaviour of the system in a given phase is comparatively easier to understand than the dynamics of a system at the verge of a phase transition. The ratio between the kinetic and potential energies of laboratory systems can be easily controlled by controlling their temperature. Hence they can be forced to stay in one particular phase and studied. It is impossible to do this for a gravitating system. The virial theorem forces gravitating systems to have comparable kinetic and potential energies. Thus gravitating systems in virial equilibrium are similar to normal laboratory systems at the verge of a phase transition and hence, are difficult to analyse.

We shall make this idea more quantitative in the later sections of the review. In this chapter, we shall discuss certain general aspects of statistical mechanics which are of importance in the study of gravitating systems.

2.2. *Fundamentals of the statistical description*

Consider a system of N particles interacting via a two body potential $V(r)$ which behaves as $-Gm^2/r$ at large r . [At the moment, we are not making any assumptions regarding the short distance behaviour of $V(r)$; we shall comment on this aspect later in section 2.4.] Such a system can be described, at any instant of time, by a point in a phase space of $6N$ dimensions, parametrized by the $6N$ canonical co-ordinates and momenta (q_i, p_i) . As the system evolves the phase point traces a one-dimensional curve in this phase space.

One can view this fact in a different manner, which is more useful in statistical mechanics: Given the trajectories $q_i = q_i(t; q_i^0, p_i^0)$ and $p_i = p_i(t; q_i^0, p_i^0)$ of the particles, which depend on the initial conditions q_i^0 and p_i^0 , one can (in principle) invert these relations and obtain the functions: $q_i^0 = q_i^0(q_i, p_i; t)$ and $p_i^0 = p_i^0(q_i, p_i; t)$. These $6N$ functions are obviously conserved. One of these conserved quantities can be eliminated by a suitable choice for the origin of time, which is arbitrary for a closed system. Thus we are left with $(6N - 1)$ non-trivial constants in phase space. The motion of our system is therefore confined to a $[(6N) - (6N - 1)] = 1$ dimensional surface in phase space.

The functional form of all the $(6N - 1)$ constants of motion will be available only if the equations of motion for the system are integrated. Statistical mechanics deals with situations in which this task is impossibly difficult. In such realistic situations, we have to contend ourselves with the knowledge of some small number (k) of constants of the motion [where $k \ll (6N - 1)$], the existence of which can be ascertained from symmetry considerations. We can only conclude that the system should be confined to a $(6N - k)$ dimensional subspace of the phase space.

Seven such integrals are well known: they are energy, three components of momentum, and three components of angular momentum. In most situations, however, we will confine the system in a box and insist that the physical conclusions may depend only on the volume of the box and not on the nature and shape of the box. Such an external box is mathematically equivalent to a static potential (without any special symmetries) imposed upon the system. In this case, the total angular momentum and the total linear momentum of the contents of the box will not be conserved and we are left with energy as the only “obvious” constant of the motion. The motion will be now confined to a $(6N - 1)$ dimensional subspace.

Given the initial position of the phase point, the dynamical evolution will make it roam around this

constant energy surface. The exact location of this point at any instant will clearly depend on the starting position. In the absence of detailed knowledge regarding the initial position of the phase trajectory, we can only conclude that the system can be found somewhere on the energy surface. Because of this uncertainty in the location of the phase point we will not be able to provide unique answers to the questions regarding the variables of the system. At best, we will be able to supply a statistical description, in terms of probabilities, provided we can arrive at the probability $P(q, p, t)$ for the system to be found in an infinitesimal phase volume around the point (q, p) at time t .

It is usual, at this stage, to make the following two assumptions regarding the behaviour of the probability $P(q, p, t)$ at “sufficient” late times: (i) As $t \rightarrow \infty$, the statistical behaviour of the system is independent of time and hence $P(q, p, t) \approx P(q, p)$. (ii) In the same limit, we assume that the system does not prefer any one location of the constant energy surface over another. In other words, $P(q, p)$ will be assumed to be a constant on the energy surface and zero elsewhere. These assumptions can be valid only if the long time behaviour of the system is reasonably independent of the initial conditions. This premise, that the behaviour of the system at sufficiently late times exhibits a statistical regularity which is independent of the initial conditions, seems to be vindicated by observations.

The theoretical validity of the description outlined above rests on highly non-trivial issues related to the very foundations of statistical mechanics (see e.g. Fowler [1936], Gibbs [1902], Tolman [1936], Penrose [1970], Ter Haar [1954, 1955, 1961]). However, there is no reason to believe that gravitational interactions add any new *conceptual* complication to these questions. Hence, we shall take a simple minded attitude regarding these issues.

The above description corresponds to the “microcanonical ensemble” of statistical mechanics, valid for a closed system with fixed energy E (and volume V). This is the most basic description of statistical mechanics. We note that the average value of any phase space function $f(p, q)$ can now be computed as

$$\langle f(p, q) \rangle = \left(\frac{1}{N!g(E)} \right) \int f(p, q) \delta(E - H(p, q)) dp dq, \quad (2.1)$$

where $N!g(E)$ is the volume of phase space occupied by the constant energy surface itself. This quantity $g(E)$, usually called the density of states, is defined as

$$g(E) = \frac{1}{N!} \int \delta(E - H(q, p)) dq dp, \quad (2.2)$$

and is related logarithmically to the thermodynamic entropy S of the system,

$$S(E) = \ln g(E). \quad (2.3)$$

As is usual in classical statistical mechanics, the entropy is defined only up to an additive constant. Maximisation of entropy is thus equivalent to the maximisation of the phase volume available to the system.

The partition function $Z(\beta)$ of the system is defined by the integral

$$Z(\beta) = \int_{-\infty}^{+\infty} dE g(E) e^{-\beta E} = \int dp dq e^{-\beta H(p, q)}. \quad (2.4)$$

Note that the range of integration is from $(-\infty)$ to $(+\infty)$. If the Hamiltonian is bounded from below [i.e., if $H(p, q) > E_0$ with some finite E_0], then $g(E)$ will vanish for all $E < E_0$. One can then redefine the zero of the energy scale in such a way that the limits of integration in (2.4) change to the range $(0, +\infty)$. Then the partition function will become the Laplace transform of the density of states.

The above definitions, at this stage, are purely formal. The existence of the integrals defining $g(E)$ and $Z(\beta)$ and the usefulness of these functions will depend on the nature of the Hamiltonian, and especially on the range of the interaction.

In systems interacting via short range potentials [i.e. attractive potentials which behave as $-r^{-(3-\varepsilon)}$, with $\varepsilon > 0$, at large distances], the total energy of the system will be an extensive parameter. If we divide such a system into macroscopic subsystems then the energy of the full system will be the sum of the energies of the subsystems. When the forces are of short range, the interaction energy between the subsystems will be proportional to the surface areas of the subsystems; the energy of the subsystem, on the other hand, will increase with the volume. The interaction between the subsystems, therefore, makes a negligible contribution to the total energy.

[Another simple way of seeing this result is as follows: Consider a spherical region of radius R in which the particles are distributed uniformly. Suppose we now add a particle at the origin. The potential energy felt by this particle due to the rest of the sphere will be

$$U \propto \int_0^R 4\pi r^2 dr \rho r^{-3-\varepsilon} \propto \int_0^R dr r^{-1-\varepsilon}. \quad (2.5)$$

If $\varepsilon > 0$, then most of the contribution to this integral comes from particles near the origin. In contrast, for long range forces with $\varepsilon < 0$, most of the contribution to the potential comes from the particles at large distances. In such cases, the energy of each particle is affected by the particles at large distances; hence the energy will not be an extensive parameter.]

In gaseous systems, the forces are of short range and hence the interaction between the subsystems is ignorable. The energy will be an extensive parameter. This fact is crucial for the study of the equilibrium properties of such systems. Suppose a system with total energy E is divided into two parts with energies E_1 and $E_2 = E - E_1$. Then the following results can be easily derived (see, e.g., Landau and Lifshitz [1980]):

(a) The volume occupied by the system in the phase space is $g(E) = g_1(E_1)g_2(E - E_1)$; this expression is maximised when $\ln g(E)$ is maximised. Therefore the quantity

$$(\partial/\partial E)\ln g = \partial S/\partial E = \beta \quad (2.6)$$

must be constant in statistical equilibrium.

(b) When the full system is in equilibrium so are the large subsystems which form parts of the full system.

(c) Consider the situation in which subsystem 1 is “small” compared to subsystem 2. We can then think of subsystem 2 as a “heat reservoir” with which subsystem 1 can exchange energy. The probability for subsystem 1 to have an energy E_1 will then be proportional to the phase volume $g_1(E_1)g_2(E - E_1)$. Writing g_2 as e^{S_2} and Taylor expanding the quantity $\exp S_2(E_2) = \exp S_2(E - E_1)$ in E_1 , we easily get the result

$$P(E_1) \propto g_1(E_1) \exp(-\beta E_1). \quad (2.7)$$

This result allows one to describe systems which can exchange energy with an external heat bath. Such systems are described by a single parameter β . The probability for such a system to have an energy E is given by the “canonical distribution” of eq. (2.7). Normalising this distribution properly we get

$$P(E) = Z(\beta)^{-1} g(E) e^{-\beta E}, \quad (2.8)$$

where $Z(\beta)$ is the partition function defined earlier.

The following point should be clear from the above derivation: The microcanonical distribution can be used to study the statistical properties of any closed system. The canonical distribution derived above can be used to describe the properties of any large subsystem of a closed system, providing the energy is an extensive variable. It is, however, usual to use both canonical and microcanonical distributions in an interchangeable manner while studying the properties of gaseous systems. In fact, both distributions lead to the same predictions in such contexts. It is worthwhile to examine in some detail how the equivalence between the two distributions emerges in normal systems. (As we shall see, this equivalence between the predictions based on these two distributions breaks down in the presence of long range interactions.)

For a system described by the microcanonical distribution, the energy is fixed at some value E . When the same system is described using a canonical distribution, the energy is not a fixed quantity but can exhibit fluctuations among the members of a canonical ensemble. If we should be able to use either of the two distributions, in an interchangeable manner, then to describe the system we must satisfy the following two conditions: (a) The mean energy $U = \langle E \rangle$ of the canonical ensemble must be the same as the fixed energy of the microcanonical ensemble. (b) The root-mean-square fluctuations in the energy, in the canonical distribution, must be negligible for sufficiently large systems.

The first condition is easy to satisfy if we take the mean energy to be the same as the most probable energy of the canonical distribution. (In fact, this condition merely establishes a correspondence between the energy E describing the microcanonical distribution and the parameter β describing the canonical distribution.) The probability in (2.8) is the product of an increasing function of E , $g(E)$, and a decreasing function of E , $e^{-\beta E}$, and hence will have a maximum at some $E = U$, say. It is trivial to verify that this maximum is determined by the equation

$$(\partial S / \partial E)_{E=U} = \beta, \quad (2.9)$$

where $S(E) = \ln g(E)$ is the entropy defined from the density of states.

To verify the second condition we Taylor expand the quantity $\ln P(E)$ about the peak value. Using (2.9) we can write

$$\ln P(E) = -\beta(U - \beta^{-1}S) - \frac{1}{2}\beta^2 C_v^{-1}(E - U)^2 + \dots, \quad (2.10)$$

where, we have defined the specific heat

$$C_v = -\beta^2(\partial U / \partial \beta). \quad (2.11)$$

In other words,

$$P(E) \propto e^{-\beta(U - \beta^{-1}S)} \exp[-\frac{1}{2}\beta^2 C_v^{-1}(E - U)^2]. \quad (2.12)$$

This equation shows that the probability distribution for E , given by the canonical distribution, is a Gaussian centered at U with a dispersion proportional to $\beta^{-2}C_V$. For systems interacting through short range forces, the mean energy U and the specific heat C_V will grow as N ; therefore, the dispersion in the energy, $\Delta E \propto C_V^{1/2}$, will grow as \sqrt{N} . The relative fluctuation in E , $\Delta E/E$, will thus decrease as $N^{-1/2}$. This result, along with eq. (2.9), demonstrates the equivalence [up to $O(N^{-1/2})$] of the canonical and microcanonical descriptions for systems with short range forces.

In the above discussion we have assumed that the distribution (2.8) is very sharply peaked around the mean value of the energy so that the mean and most probable values are the same to $O(N^{-1/2})$. Since $U = \langle E \rangle = -Z'/Z$ and $\langle E^2 \rangle = Z''/Z$, where the primes denote differentiation with respect to β , it immediately follows that

$$C_V = -\beta^2 U' = \beta^2 (\langle E^2 \rangle - \langle E \rangle^2). \quad (2.13)$$

Therefore C_V defined from the canonical distribution must be positive definite.

We can also define the specific heat directly from the microcanonical distribution. From $S(E) = \ln g(E)$ we can construct the function $\beta(E) = (\partial S / \partial E)$; by inverting this function we can get $E(\beta)$, from which the specific heat may be defined as $C_V = -\beta^2 E'$. As long as the two distributions are equivalent this expression will also be positive definite. However, there is no assurance that this expression is *always* positive; in fact, it is possible to construct systems for which the microcanonical specific heat becomes negative in some range. When this happens, the equivalence between the two distributions also breaks down.

A very likely situation in which this occurs is near a phase transition. This can be seen as follows: To prove the equivalence between the two distributions, we had to assume that (2.8) is sharply peaked at the mean energy. From (2.12) it is clear that this assumption breaks down when C_V – defined from the canonical distribution – becomes very large. This is precisely what happens near a phase transition occurring at a constant temperature and involving a finite latent heat. Thus near the phase transition, there will be large fluctuations and the two distributions will lead to different predictions. As we shall see in chapter 3, the specific heat defined from the microcanonical distribution can become negative near a phase transition.

2.3. *Difficulties due to the long range nature of gravity*

There are essentially two kinds of gravitating systems for which one would like to develop a statistical theory. The first type consists of astrophysical systems like globular clusters, galaxies etc. These systems contain approximately spherical objects of finite size. Except when two of the stars are close enough to be in contact physically, they may be assumed to interact via the $-r^{-1}$ potential. The second type of gravitating systems we will be interested in are the dark matter halos around astrophysical objects which could be made of elementary fermionic particles (like neutrinos). They must be treated as point particles in classical physics; however, quantum mechanical considerations change the nature of their interactions at short distances.

Since the large distance behaviour of the potential is the same for both these systems, we will begin by studying the special difficulties caused by the asymptotic $-r^{-1}$ nature of the potential. Issues related to the short distance behaviour of the potential will be taken up in the next section.

Realistic astrophysical systems are “open” in the sense that particles in these systems can escape to infinity. It is easy to see that the integral defining $g(E)$ will diverge if the range of spatial integrations is

extended to infinity. (This divergence, of course, is in addition to any divergence we may encounter due to the short distance behaviour of the potential.) A similar divergence of $g(E)$ will occur even for an ideal gas if it were not confined in a box. We are, therefore, forced to introduce the first artificiality: We confine the system inside a spherical box of radius R . This assumption can be justified if we can demonstrate that the fractional rate of evaporation of particles from the system is small. We shall assume that this can be done.

Given such a confining volume and a suitable short distance behaviour of the potential, one can, in principle, compute the phase volume $g(E, V)$, the entropy $S(E, V)$ and the partition function $Z(\beta, V)$. However, these quantities behave in an unfamiliar manner for gravitating systems.

The central difficulty arises from the non-extensive nature of the energy. We can no longer divide the system into non-interacting subsystems. The conclusions (a) to (c), derived in the previous section, depended crucially on the extensive nature of energy, and are therefore invalid for a system with long range interactions like gravity. Notice, in particular, that the canonical distribution of energy – which is usually used for a subsystem in contact with a larger system – can no longer be derived from the microcanonical distribution for a gravitating system. As we shall see later, gravitating systems behave in a very peculiar manner if put in contact with a heat bath.

The microcanonical distribution, therefore, is the proper distribution to use in the study of gravitating systems. This feature also ties in with another well known property of gravitating systems, viz., that gravitating systems in virial equilibrium have negative specific heat. As we saw before, systems described by the canonical distribution cannot have negative specific heat. No such constraint exists for the microcanonical distribution.

This situation is in marked contrast with that usually encountered in the study of neutral gases (with genuinely short range interactions) or in the study of plasmas (where the long range interaction is shielded by the Debye cloud). In such systems canonical and microcanonical distributions will produce identical results to $O(N^{-1/2})$ except near a phase transition. This suggests the possibility that the behaviour of gravitating systems may be similar to the behaviour of ordinary systems near a phase transition. We shall see in chapter 3 that this is indeed true. Even though the canonical distribution cannot be *derived* from the microcanonical distribution in the presence of long range-forces we can – purely as a formal, mathematical construct – *define* the partition function for such systems. Comparing the function $E(\beta)$ obtained from the microcanonical distribution with the corresponding function obtained from the partition function one can prove that the negative specific heat region of the microcanonical distribution is replaced by a phase transition in the canonical distribution. Several such examples will be discussed in chapters 3 and 4.

One final point may also be noted: When the interactions have a short range, we can divide the system into non-interacting subsystems, thereby ensuring the overall spatial homogeneity of the equilibrium configuration. The equilibrium configurations for gravitating systems, on the other hand, are necessarily inhomogeneous.

Since the energy is no longer an extensive variable, gravitating systems do not possess a thermodynamic limit. Though this result arises because of the long range nature of the gravitational force, it is more convenient to discuss this feature in the next section along with the other effects arising due to the short distance behaviour of the potential.

2.4. *Effects due to the short distance behaviour of $V(r)$*

The statistical description outlined above is based on the phase volume $g(E)$ or, equivalently, on the entropy $S(E)$. One would like to compute these quantities in order to understand the nature of the

equilibrium configuration described by the microcanonical ensemble. The short distance behaviour of $V(r)$ needs to be specified before such a computation can be undertaken.

The behaviour of the potential $V(r)$ near $r \approx 0$ depends very much on the system we are considering. The simplest, and more unrealistic, assumption would be to treat all the constituents as *classical* point particles. Then the potential will exhibit the $-r^{-1}$ behaviour for all r . It is easy to see that no statistical equilibrium exists for such a system; the phase volume $g(E)$ will diverge for all E (if $N \geq 3$). Expression (2.2) becomes, on integrating out the momenta,

$$g(E) = \text{constant} \times \int d^3x_1 d^3x_2 \cdots d^3x_N \left(E + \frac{1}{2} \sum_{i \neq j} \frac{Gm^2}{|\mathbf{x}_i - \mathbf{x}_j|} \right)^{3N/2-1}. \quad (2.14)$$

We change the integration variables from $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ to $(s, \mathbf{x}_2, \dots, \mathbf{x}_N)$, where $s = \mathbf{x}_1 - \mathbf{x}_2$. Then (2.14) becomes

$$g(E) = \int d^3x_2 d^3x_3 \cdots d^3x_N I(\mathbf{x}_2, \dots, \mathbf{x}_N), \quad (2.15)$$

where

$$I(\mathbf{x}_2, \dots, \mathbf{x}_N) = \int ds \left(E + \frac{Gm^2}{|s|} + \sum_{j=3}^N \frac{Gm^2}{|s + \mathbf{x}_2 - \mathbf{x}_j|} + \frac{1}{2} \sum_{i,j=2}^N \frac{Gm^2}{|\mathbf{x}_i - \mathbf{x}_j|} \right)^{3N/2-1}. \quad (2.16)$$

This integral is divergent. Near $s = 0$ it behaves as

$$I = \int_0^\varepsilon ds s^2 (Gm^2/s)^{3N/2-1} = \lim_{\varepsilon \rightarrow 0} \varepsilon^{(4-3N/2)}, \quad (2.17)$$

which diverges for all $N \geq 3$ making (2.15) also diverge.

It must be emphasised that this result is *not* a peculiarity related to gravitational interactions. The phase space volume occupied by a plasma consisting of equal numbers of oppositely charged particles, will diverge in the same manner. This is, of course, to be expected because the Debye shielding affects only the large- r behaviour of the potential; the divergences we are discussing now arise from the small- r behaviour of the potential. This issue has been investigated thoroughly in the case of Coulomb systems. It has been found that ordinary matter owes its stability to quantum mechanics, which introduces an effective repulsion between particles through uncertainty and exclusion principles [Lieb 1976; Dyson 1967; Dyson and Lenard 1967; Lenard and Dyson 1968; Lieb and Thirring 1975; Lieb and Lebowitz 1972].

It seems, therefore, natural to invoke quantum theory if we have a system of point particles (say, a cloud of massive neutrinos). The phase space volume will then be finite, though impossibly complicated to compute. It is, however, possible to prove certain general results which are of interest. It can be shown that, even though a stable equilibrium state is ensured by the quantum corrections, no thermodynamic limit exists for gravitating systems [Levy-Leblond 1969; Hertel et al. 1972]. This is to be expected; the thermodynamic limit depends on the large distance behaviour of the potential while the quantum corrections only change the short distance behaviour. In Coulomb systems the existence of Debye shielding allows one to prove the existence of the thermodynamic limit once the question of stability is taken care of; such a derivation is clearly not possible for the unshielded gravitational interactions.

These results can be stated in a more quantitative form [Levy-Leblond 1969]: Consider an N particle system described by the Hamiltonian

$$H = \sum_{i=1}^N \frac{p_i^2}{2m} - \frac{1}{2} \sum_{i \neq j} \frac{Gm^2}{|r_i - r_j|}. \quad (2.18)$$

It can be shown that the ground state energy $E_0(N)$ of such a system is bounded between two limits,

$$-aN(N-1)^2(G^2m^5/\hbar^2) \leq E_0(N) \leq -bN(N-1)^2(G^2m^5/\hbar^2), \quad (2.19)$$

where a and b are *positive* constants independent of N . This result shows that, though a ground state with finite energy exists, the ground state energy scales as N^3 for large N . Therefore the binding energy per particle, $E_0(N)/N$, grows without bound as the thermodynamic limit, $N \rightarrow \infty$, is taken. It is clearly impossible to define the usual thermodynamic variables; as $N \rightarrow \infty$ the ground state becomes more and more condensed and massive.

[A qualitative understanding of the above result may be obtained by the following considerations: We estimate the total energy of the system as

$$E(N) \approx N(p^2/2m) - \frac{1}{2}N^2(Gm^2/l), \quad (2.20)$$

where l is some typical mean separation. Because of the uncertainty principle, we may assume that $l > \hbar/p$ and hence

$$E(N) > N(p^2/2m) - \frac{1}{2}N^2(Gm^2p/\hbar). \quad (2.21)$$

The ground state energy, obtained by minimising this expression with respect to p , is

$$E_0(N) \approx -N^3(G^2m^5/8\hbar^2). \quad (2.22)$$

The N^3 scaling is obvious. If the potential was changed from $-r^{-1}$ to $-r^{-n-1}$, then the ground state energy scales as $N^{(n-3)/(n-1)}$; for $n \gg 1$, corresponding to a short range interaction, we recover the $E_0 \propto N$ scaling and the thermodynamic limit. This derivation however, is not rigorous; see the comments in Lieb [1976, p. 555].

The above result was derived without assuming anything about the statistics of the particles. If the particles are fermions, a different bound can be obtained for $E_0(N)$. We can then show that

$$-AN(N-1)^{4/3}(G^2m^5/\hbar^2) \leq E_0(N) \leq -BN^{1/3}(N-1)^2(G^2m^5/\hbar^2), \quad (2.23)$$

where A and B are positive constants. The ground state energy now grows as $N^{7/3}$ for large N . Though this is a mild improvement on the previous result it is still not good enough to provide a thermodynamic limit.

The role of gravity in these results can be best appreciated by comparing them with the corresponding results for systems dominated by electromagnetic interactions. In fact, it is possible to obtain similar bounds when the particles experience both electromagnetic and gravitational forces. Consider a system consisting of N identical fermions with mass m and charge $-e$, and N particles of mass M , charge e , and

unspecified statistics. It can then be shown that

$$-CN(me^4/2\hbar^2)[1 + cN^{2/3}GM^2/e^2]^2 \leq E_0(N) \leq -DN(me^4/2\hbar^2)[1 + dN^{2/3}GM^2/e^2]^2, \quad (2.24)$$

where C , D , c and d are positive constants.

This result has interesting limiting forms. If we ignore the gravitational interaction by setting $G = 0$, then we see that $E_0(N) \propto N$, ensuring the existence of the thermodynamic limit for a neutral plasma. (This result also shows that the existence or otherwise of the thermodynamic limit is really decided by the large distance behaviour of the potential.) Setting $e = 0$, of course, will reproduce the previous results. We can also estimate the value of N for which the non-extensive nature of the gravitational contribution becomes important. This will happen when the second term inside the square brackets in eq. (2.24) becomes comparable to unity; that is for

$$N \approx N_c = (e^2/GM^2)^{3/2} \approx 10^{54}. \quad (2.25)$$

The numerical estimate is made taking M to be the proton mass. For larger systems, gravitational forces prevent the existence of the thermodynamic limit.

The short distance behaviour of $V(r)$ will be much more complicated for stellar systems. Non-gravitational forces will play a significant role when two realistic astrophysical systems – say, two stars – are close to each other. These forces can lead to diverse and complex phenomena depending on the details of the interaction. Usually, these forces can be relied upon to provide a stable ground state of the system if the masses of the stars are not too high. If the stars are too massive, then it is necessary to consider general relativistic effects, which can eventually lead to formation of black holes. These aspects are beyond the scope of the current review. In what follows, we shall bypass these difficulties by assuming that the “stars” which we consider have an effective hard core radius. This assumption allows us to concentrate on the statistical aspects of the gravitating system, without worrying about the complicating details (e.g. the gas dynamical, tidal and general relativistic effects) of close encounters. In this idealised case, where we have approximated the constituents of the system as hard spheres with some effective radius, the ground state will correspond to a closely packed configuration with all the constituents in close contact. The thermodynamic limit, of course, will not exist since that result arises from the large r behaviour of the potential. We shall study several specific models with such a short distance cutoff in later sections.

3. Statistical equilibrium of gravitating systems: toy models

3.1. Introduction

In the last section, we discussed several peculiar features which arise when the standard methods of equilibrium statistical mechanics are applied to a gravitating system. To understand these features quantitatively, we need to compute the microcanonical distribution function for the system. Since this is an impossibly difficult task, it is worthwhile to study some model systems which share the essential features of gravitating systems. These model systems can be analysed exactly and hence offer valuable insights into the nature of statistical equilibrium of gravitating systems.

The two essential features of gravity which are relevant to the nature of statistical equilibrium are the

following: (i) The gravitational potential well is infinitely deep; therefore, we need a short distance cutoff to ensure the stability of the ground state. (ii) The gravitational interactions have an $-r^{-1}$ behaviour at large distances; this makes the canonical and microcanonical descriptions inequivalent and forces one to use the latter. The models which we study will all have these two features incorporated in some form. We shall see that these two properties are indeed responsible for most of the interesting effects discussed in the last section. What is more, all the systems which possess the properties (i) and (ii) show striking similarity in their behaviour.

These models will also illustrate the analogy between virialised gravitating systems and normal laboratory systems undergoing a phase transition. The main difficulty in appreciating this analogy is the following: We usually study the phase transition in laboratory systems using the canonical description in which the system can exchange energy with the heat bath. For example, a liquid may undergo a phase transition at its freezing point and become a solid, releasing energy to the surroundings. This process will appear very strange if we describe it using a microcanonical distribution at constant energy. (Qualitatively, we will expect part of the system to freeze and release the energy to the rest of the system; that is, we expect both phases to co-exist.) But gravitating systems – as we saw in the last section – *can only be* described by the microcanonical distribution. Therefore, certain peculiarities in the behaviour are inevitable. To circumvent this difficulty and keep the discussion transparent we will adopt the following procedure: We will study the toy models in *both* the canonical and microcanonical descriptions (though only the latter is really relevant to the astrophysical systems we are interested in); the analogy with systems undergoing a phase transition will then be apparent in the canonical description.

3.2. *Statistical description of a binary star*

We will begin by studying the statistical mechanics of a two body system described by the Hamiltonian

$$H(\mathbf{Q}, \mathbf{P}; \mathbf{p}, r) = \frac{\mathbf{P}^2}{2M} + \frac{\mathbf{p}^2}{2\mu} - \frac{Gm^2}{r}, \quad (3.1)$$

where (\mathbf{Q}, \mathbf{P}) are the co-ordinates and momenta of the centre of mass, (\mathbf{r}, \mathbf{p}) are the relative co-ordinates and momenta, $M = 2m$ is the total mass, $\mu = m/2$ is the reduced mass and m is the mass of the individual particles. We shall further restrict the range of the r co-ordinates to the interval (a, R) . The short distance cutoff at a is equivalent to assuming that the particles are hard spheres of radius $a/2$, while the large distance cutoff can be achieved by confining the system inside a spherical box of radius R .

In the absence of the confining box and the short range cutoff, the Hamiltonian in (3.1) describes the standard Kepler problem. For finite R and a , one has to consider the effects of collisions. Angular momentum conservation does not hold in the presence of collisions. The problem becomes more complicated and, in fact, the phase trajectory becomes extremely irregular.

We are interested in the statistical mechanics of the system described by H . It may appear that a system with just two particles is a very bad example for a statistical discussion. However, this is not quite true. As we shall see, this system exhibits several peculiarities of gravitating systems [Padmanabhan 1989a,b]. We will be careful to use this model only in a limited sense. Then the effects due to the small number of degrees of freedom can be easily identified and dealt with.

We shall first study the microcanonical distribution corresponding to our system. We are interested in the phase volume $g(E)$ of the constant energy surface $H = E$, which is most easily computed as

$$g(E) = d\Gamma(E)/dE, \quad (3.2)$$

$$\Gamma(E) = \int d^3P d^3p d^3Q d^3r \Theta(E - H(P, Q, p, r)). \quad (3.3)$$

The momentum integrations and the integration over Q can be easily performed giving

$$\Gamma(E) = \frac{1}{3} AR^3 \int_a^{r_{\max}} r^2 dr (E + Gm^2/r)^3, \quad (3.4)$$

where $A = 64\pi^5 m^3/3$; therefore

$$g(E) = AR^3 \int_a^{r_{\max}} r^2 dr (E + Gm^2/r)^2. \quad (3.5)$$

It is understood that the range of integration in (3.5) should be limited to the region in which the expression in parentheses is positive. Therefore, we should use $r_{\max} = Gm^2/|E|$ if $-Gm^2/a < E < -Gm^2/R$ and use $r_{\max} = R$ if $-Gm^2/R < E < +\infty$. Since $H \geq -Gm^2/a$, we trivially have $g(E) = 0$ for $E < -Gm^2/a$. The constant A is unimportant for our discussion and hence will be omitted from the formulas hereafter.

The integration in (3.5) is straightforward and we get the following result:

$$\frac{g(E)}{(Gm^2)^3} = \begin{cases} \frac{1}{3} R^3 (-E)^{-1} (1 + aE/Gm^2)^3, & -Gm^2/a < E < -Gm^2/R, \\ \frac{1}{3} R^3 (-E)^{-1} [(1 + RE/Gm^2)^3 - (1 + aE/Gm^2)^3], & -Gm^2/R < E < \infty. \end{cases} \quad (3.6)$$

This function $g(E)$ is continuous and smooth at $E = -Gm^2/R$. The thermodynamics of the system can be now studied using $g(E)$. We define the entropy $S(E)$ and the temperature $T(E)$ of the system by the relations

$$S(E) = \ln g(E), \quad T^{-1}(E) = \beta(E) = \partial S(E)/\partial E. \quad (3.7)$$

All the interesting thermodynamic properties of the system can be understood from the $T(E)$ curve.

Consider first the case of very low energies, i.e., $-Gm^2/a < E < -Gm^2/R$. Using (3.6) and (3.7) one can easily obtain $T(E)$ and write it in dimensionless form as

$$t(\varepsilon) = \left(\frac{3}{1 + \varepsilon} - \frac{1}{\varepsilon} \right)^{-1}, \quad (3.8)$$

where we have defined $t = aT/Gm^2$ and $\varepsilon = aE/Gm^2$. This function exhibits the peculiarities characteristic of gravitating systems. At the lowest energy admissible for our system, which corresponds to $\varepsilon = -1$, the temperature t vanishes. This describes a tightly bound low temperature phase of the system

with negligible random motion. $t(\varepsilon)$ is clearly dominated by the first term of (3.8) for $\varepsilon \approx -1$. So as we increase the energy of the system, the temperature *increases*, which is the normal behaviour for a system. This trend continues up to

$$\varepsilon = \varepsilon_1 = -\frac{1}{2}(\sqrt{3} - 1) \approx -0.36, \quad (3.9)$$

at which point the $t(\varepsilon)$ curve reaches a maximum and turns around. As we increase the energy further the temperature *decreases*. The system exhibits negative specific heat in this range.

Equation (3.8) is valid from the minimum energy $-Gm^2/a$ all the way up to the energy $-Gm^2/R$. For realistic systems $R \gg a$, and hence this range is quite wide. For a small region in this range (from $-Gm^2/a$ to $-0.36Gm^2/a$) we have positive specific heat; for the rest of the region the specific heat is negative. The positive specific heat region owes its existence to the non-zero short distance cutoff. If we set $a=0$, the first term in (3.8) will vanish; we will have $t \propto -\varepsilon^{-1}$ and a negative specific heat in this entire domain.

For $E \geq -Gm^2/R$, we have to use the second expression in (3.6) for $g(E)$. In this case, we get

$$t(\varepsilon) = \left(\frac{3\{(1+\varepsilon)^2 - (R/a)[1+(R/a)\varepsilon]^2\}}{(1+\varepsilon)^3 - [1+(R/a)\varepsilon]^3} - \frac{1}{\varepsilon} \right)^{-1}. \quad (3.10)$$

This function, of course, matches smoothly with (3.8) at $\varepsilon = -a/R$. As we increase the energy, the temperature continues to decrease for a little while, exhibiting negative specific heat. However, this behaviour is soon halted at some $\varepsilon = \varepsilon_2$, say. The $t(\varepsilon)$ curve reaches a minimum at this point, turns around, and starts increasing with increasing ε . We thus enter another (high temperature) phase with positive specific heat. From (3.10), it is clear that $t \approx \frac{1}{2}\varepsilon$ for large ε . (Since $E = \frac{3}{2}NkT$ for an ideal gas, we might have expected to find $t \approx \frac{1}{3}\varepsilon$ for our system with $N=2$ at high temperatures. This is indeed what we would have found if we had defined our entropy as $\ln I$. With our definition, the energy of the ideal gas is actually $E = (\frac{3}{2}N - 1)kT$; hence we get $t = \frac{1}{2}\varepsilon$ when $N=2$. Note that this is clearly an effect arising from the small number of degrees of freedom and is not of any fundamental significance.) The form of $t(\varepsilon)$ is shown in fig. 3.1. The specific heat is positive along the portions AB and CD and is negative along BC.

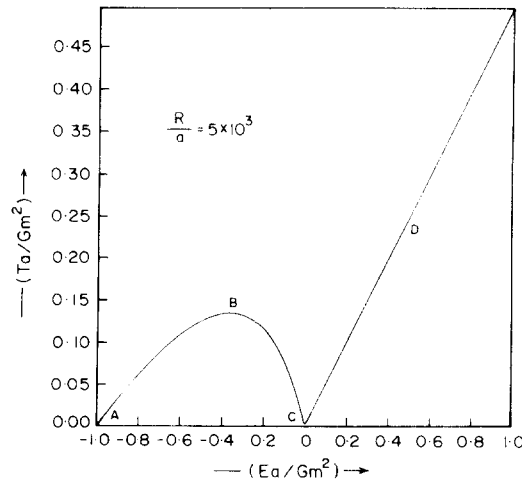


Fig. 3.1. The function $T(E)$ for a binary system. [Point C corresponds to $(0, a/2R)$ and is slightly raised for clarity.] Note that the specific heat is positive along AB and CD and is negative along BC.

The overall picture is now clear. Our system has two natural energy scales: $E_1 = -Gm^2/a$ and $E_2 = -Gm^2/R$. For $E \gg E_2$, gravity is not strong enough to keep $r < R$ and the system behaves like a gas confined by the container; we have a high temperature phase with positive specific heat. As we lower the energy to $E \lesssim E_2$, the effects of gravity begin to be felt. For $E_1 < E < E_2$, the system is unaffected by either the box or the short distance cutoff; this is the domain dominated entirely by gravity and we have negative specific heat. As we go to $E \approx E_1$, the hard core nature of the particles begins to be felt and gravity is again resisted. This gives rise to a low temperature phase with positive specific heat.

We can also consider the effect of increasing R , keeping a and E fixed. (This is more in consonance with the spirit of the microcanonical distribution which is for fixed E .) Since we imagine the particles to be hard spheres of radius $a/2$, we should only consider $R > 2a$. It is amusing to note that, if $2 < R/a < (\sqrt{3} + 1)$, there is no region of negative specific heat. As we increase R , this negative specific heat region appears. It is easy to see that increasing R effectively increases the range over which the specific heat is negative. Suppose a system is originally prepared with some E and R values such that the specific heat is positive. If we now increase R , it can happen that the system finds itself in a region of negative specific heat. This suggests the possibility that an instability may be triggered in a constant energy system if its radius increases beyond a critical value. We shall discuss such an instability in chapter 4.

In the above discussion we have treated a as a fixed quantity and rescaled all variables using Gm^2/a . To study the effect of a short distance cutoff we would like to vary a keeping R fixed. To do this it is better to rescale variables using Gm^2/R . This can be easily done and fig. 3.2 shows the behaviour of the $T(E)$ curve as the lower cutoff a is changed. As we decrease the value of a the negative specific heat region becomes more and more pronounced. In fact, if a is zero, we have negative specific heat for all

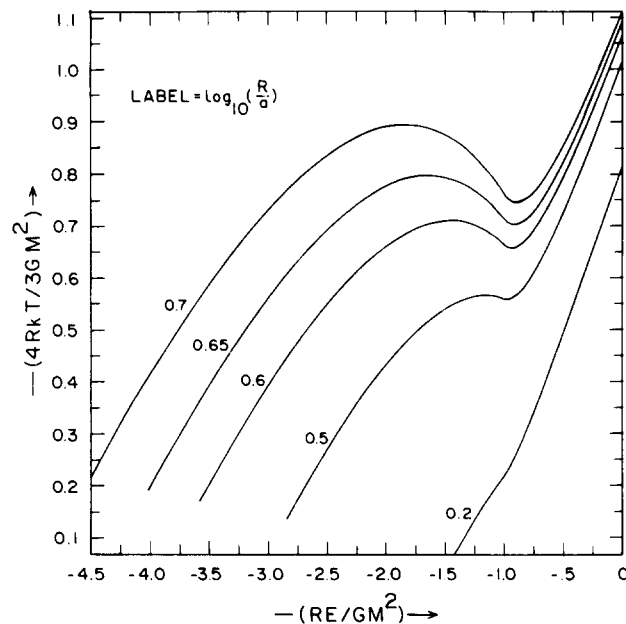


Fig. 3.2. The dependence of the $T(E)$ curve on the short distance cutoff a . Note that for $a > (\sqrt{3} + 1)^{-1}R$ there is no region of negative specific heat. As a is lowered, the negative specific heat region becomes more and more pronounced.

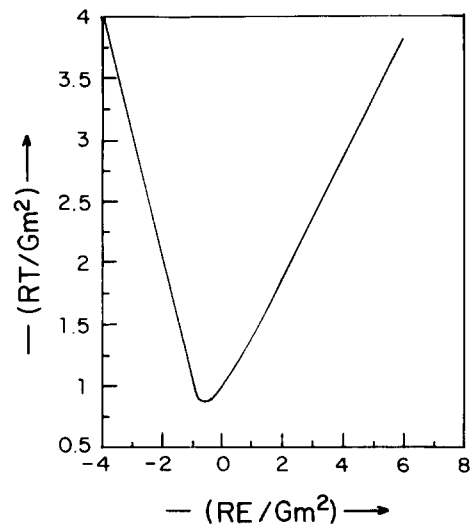


Fig. 3.3. The $T(E)$ curve for $a = 0$. The low temperature region with positive specific heat does not exist in the absence of the short distance cutoff. The system also exhibits a lower bound on the temperature.

$E < -Gm^2/R$ (see fig. 3.3). We also see from fig. 3.2 that the behaviour of the curve for large E is reasonably independent of the value of the cutoff a . We will see in chapter 4 that several realistic systems exhibit curves similar to those in fig. 3.2.

It must be noted that most real life systems are indeed found in the range $E_1 < E < E_2$. For astrophysical systems this range is quite wide since $|E_2| \ll |E_1|$. This is precisely the region with negative specific heat. Since systems described by the canonical distribution cannot exhibit negative specific heat, it follows that the canonical distribution will lead to a very different physical picture for this range of (mean) energies. In other words, the canonical and microcanonical descriptions are very different for gravitating systems in the most important range of energies. It is, therefore, of interest to look at our system from the point of view of the canonical distribution.

To do this we have to compute the partition function

$$Z(\beta) = \int d^3P d^3p d^3Q d^3r \exp(-\beta H), \quad (3.11)$$

in which the integrations over P , p and Q can be performed trivially. Omitting an overall constant which is unimportant, we can write the answer as

$$Z(\beta) = R^3 \beta^{-3} \int_a^R dr r^2 \exp(\beta Gm^2/r), \quad (3.12)$$

which, in dimensionless form, becomes

$$Z(t) = t^3 (R/a)^3 \int_1^{R/a} dx x^2 \exp(1/xt), \quad (3.13)$$

where t is the dimensionless temperature defined in (3.8). Though this integral cannot be evaluated in closed form, all the limiting properties of $Z(\beta)$ can be easily obtained from (3.13). Note that the short distance cutoff is vital here for convergence of the integral while we could have dispensed with it in the microcanonical description.

The integrand in (3.13) is large for both large and small x and reaches a minimum for $x = x_m = 1/(2t)$. The behaviour of the integral depends crucially on whether this minimum falls within the limits of integration or not. At high temperatures, $x_m < 1$ and hence the minimum falls outside the domain of integration. The exponential contributes very little to the integral and we can approximate Z adequately by

$$Z \approx t^3 \int_1^{R/a} dx x^2 (1 + 2x_m/x) = \frac{1}{3} t^3 (R/a)^6 (1 + 3a/2Rt). \quad (3.14)$$

On the other hand, if $x_m > 1$ the minimum lies between the limits of the integration. Then the exponential part of the curve dominates the integral. We can easily evaluate this contribution by a saddle point approach and get

$$Z \approx (R/a)^3 t^4 (1 - 2t)^{-1} \exp(1/t) [1 - \exp(-R/ta)] \approx (R/a)^3 t^4 (1 - 2t)^{-1} \exp(1/t). \quad (3.15)$$

As we lower the temperature, making x_m cross 1 from below, the contribution switches over from (3.14) to (3.15). The transition is exponentially sharp. The critical temperature at which the transition occurs can be estimated by finding the temperature at which the two contributions are equal. We find that

$$t_c = \frac{1}{3 \ln(R/a)}. \quad (3.16)$$

Thus for $t < t_c$, we should use (3.15) and for $t > t_c$ we should use (3.14).

Given $Z(\beta)$, all thermodynamic functions can be computed. In particular, the mean energy of the system is given by

$$E(\beta) = -(\partial \ln Z / \partial \beta). \quad (3.17)$$

This relation can be inverted to give $T(E)$, which can be compared with the $T(E)$ obtained earlier using the microcanonical distribution. (In making such a comparison we are identifying the constant energy at which the microcanonical distribution is defined with the mean energy of the canonical distribution. This is the usual dictum of statistical mechanics.) From (3.14) and (3.15) we get

$$\varepsilon(t) = aE/Gm^2 = 4t - 1, \quad \text{for } t < t_c, \quad (3.18)$$

$$\varepsilon(t) = 3t - 3a/2R, \quad \text{for } t > t_c. \quad (3.19)$$

Near $t \approx t_c$, there is a rapid variation of the energy and we cannot use either asymptotic form. The system undergoes a phase transition at $t = t_c$ absorbing a large amount of energy,

$$\Delta\varepsilon \approx 1 - \frac{1}{3 \ln(R/a)}. \quad (3.20)$$

The specific heat is, of course, positive throughout the range. This is to be expected because the canonical description cannot lead to negative specific heats.

The exact $T(E)$ curves obtained from the canonical and microcanonical distributions are shown in fig. 3.4. (For convenience, we have rescaled the $T(E)$ curve of the microcanonical distribution so that $\varepsilon \approx 3t$ asymptotically.) Notice that the descriptions match very well in the regions of positive specific heat. The negative specific heat region of the microcanonical distribution is replaced by a phase transition (rapid change in E at almost constant T) in the canonical description.

The physics of canonical description is best understood by studying E as a function of T . As we increase the temperature from zero, the energy increases from the ground state value $-Gm^2/a$ in accordance with (3.18). As the temperature approaches T_c and crosses it, a phase transition occurs in the system and the energy increases rapidly. The latent heat in the system is large enough to push the system into the high temperature phase. At still higher temperatures, the energy increases steadily with the temperature in accordance with (3.19).

We can now compare the canonical and microcanonical descriptions of our system. At both very low and very high temperatures, the descriptions match. The crucial difference occurs at the intermediate energies and temperatures. The microcanonical description predicts a negative specific heat and a reasonably slow variation of energy with temperature. The canonical description, on the other hand,

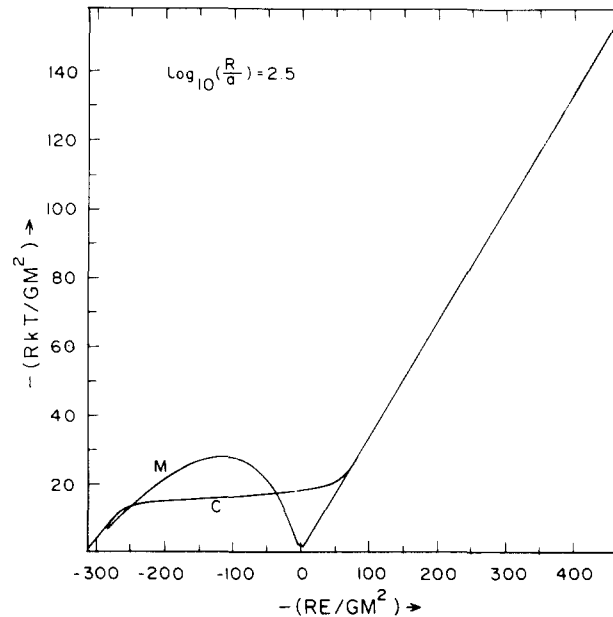


Fig. 3.4. Comparison of the $T(E)$ relations for the canonical and microcanonical distributions. The negative temperature region of the microcanonical distribution is replaced by a phase transition in the canonical distribution. The microcanonical temperature is rescaled by a factor $2/3$ for convenience of comparison.

predicts a phase transition with a rapid variation of energy with temperature. Such phase transitions are accompanied by large fluctuations in the energy, which is the main reason for the disagreement between the two descriptions.

The two descriptions, therefore, disagree for a wide range of energies. This range of energies is of great practical importance since most astrophysical systems fall in this band.

All the above conclusions can be easily generalised for a model system consisting of N *non-interacting* binaries; the partition function Z_N is just $(Z_1)^N$. All quantities derived from $\ln Z$ scale as N .

3.3. The Lynden-Bells' and Thirring's Models

Lynden-Bell and Lynden-Bell [1977] have analysed a model which has the same features as the binary star model described above. Their model describes a system with $(2N + 1)$ co-ordinates evolving through the Hamiltonian

$$H = \frac{p^2}{2m} + \sum_{i=1}^N \frac{1}{2mr^2} \left(p_{\theta_i}^2 + \frac{p_{\phi_i}^2}{\sin^2 \theta} \right) - \frac{Gm^2}{2r}. \quad (3.21)$$

The radial co-ordinate r is limited to a range of (a, R) . They study this system using both the canonical and microcanonical distributions. The properties of this system turn out to be essentially the same as those of the binary star. The microcanonical distribution exhibits high and low temperature phases connected by a $T-E$ relation which leads to negative specific heat. In the canonical description this negative specific heat branch is replaced by a phase transition at an intermediate temperature. Since this

system has a large number of degrees of freedom ($N \gg 1$), the phase transition is considerably sharper than what we have found in the case of the binary star model.

Thirring [1970] (also see Hertel and Thirring [1971, 1972]) has studied the statistical mechanics of the following N -body system: A set of N particles are confined inside a volume V . The potential between the particles is such that any two particles interact with each other and feel a constant potential if – and only if – both of them are inside an interaction volume V_0 . This is equivalent to assuming an interaction potential $u(\mathbf{x}_i, \mathbf{x}_j)$ of the form

$$u(\mathbf{x}_i, \mathbf{x}_j) = 2k\theta_{V_0}(\mathbf{x}_i)\theta_{V_0}(\mathbf{x}_j), \quad (3.22)$$

where $\theta_{V_0}(\mathbf{x})$ is unity if \mathbf{x} is inside the region V_0 and is zero otherwise. The phase volume $g(E)$ can be computed exactly for this system. It turns out that the $T(E)$ curve for this system is quite similar to those of the binary star and the Lynden-Bells' models. The thermodynamical behaviour of these systems is identical for all practical purposes.

This result may appear surprising at first sight because the potential in (3.22) does not have an $1/r$ behaviour. A closer scrutiny, however, identifies the really crucial ingredient which is common to all the three models: *the non-extensive nature of the total energy*. It is easy to see that the potential in (3.22) contributes a term $-kN_0^2$ to the total energy, where N_0 is the number of particles contained in the volume V_0 . This term makes the energy non-extensive and leads to the, by now familiar, consequences.

We can verify this feature in the following manner: Consider a system described by the phase volume

$$g(E) = \frac{1}{N_0!} \int d^{3N_0}x d^{3N_0}p \delta\left(E_0 - \sum_i p_i^2 + kN_0^2\right), \quad (3.23)$$

in which we have put in by hand a non-extensive potential energy $-kN_0^2$. It is, of course, trivial to compute (3.23) and evaluate the entropy $S = \ln g$. We get

$$S_0(E_0) = N_0 \ln V_0 - \frac{5}{2}N_0 \ln N_0 + \frac{3}{2}N_0 \ln(E_0 + kN_0^2), \quad (3.24)$$

where we have used the Stirling approximation for the factorials. Suppose we now let this system interact with an ordinary ideal gas having the entropy

$$S_1(E_1) = N_1 \ln V_1 - \frac{5}{2}N_1 \ln N_1 + \frac{3}{2}N_1 \ln E_1. \quad (3.25)$$

Maximising the total entropy ($S_0 + S_1$) subject to the constraints $N_1 + N_0 = N$ and $E_1 + E_0 = E$, we can determine the equilibrium parameters of the system. In particular, the T - E relation can be determined from the maximum value of S_{total} . It is given, in a parametric form, by the relations [Thirring 1970]

$$t = \varepsilon - 2\alpha + \alpha^2, \quad (3.26)$$

$$\varepsilon = 2\alpha - \alpha^2 + 3(1 - \alpha) / \ln\left(\frac{V_1}{V_0} \frac{1 - \alpha}{\alpha}\right), \quad (3.27)$$

where $t = 3T/2Nk$ and $\varepsilon = 1 + E/kN^2$. One can easily verify that this $T(E)$ curve has the same qualitative form as the ones we have studied so far. It is, therefore, clear that the peculiar

thermodynamic features of the systems we have studied arise because of the non-extensive nature of the energy in the condensed phase of the system.

3.4. Approximate analysis of gravitating hard spheres

The models considered so far are based on Hamiltonians, which permitted exact analysis. For that very reason, these Hamiltonians do not represent realistic systems. The lessons learned from these toy models, however, can be useful in devising an *approximate* analysis of an *exact* Hamiltonian. Such an analysis was performed by Aronson and Hansen [1972] with interesting results. Their model also stresses the importance of the non-extensive nature of the energy.

They consider a system of N gravitationally interacting hard spheres, each of volume b , enclosed inside a spherical container of radius R . In the canonical description, we have to evaluate the partition function

$$Z(\beta) = \int \prod_{i=1}^N d^3 p_i d^3 x_i \exp[-\beta H(p_i, x_i)], \quad (3.28)$$

where H is the exact Hamiltonian of the system. Once the partition function is known, we can compute the free energy F and entropy S by the standard relations

$$F = -\beta \ln Z(\beta), \quad S = \beta(E - F). \quad (3.29)$$

Since we cannot evaluate $Z(\beta)$ exactly, we try the following approximate ansatz.

The study of the toy models suggests that the system can exist in two very different phases. At high temperatures, the spheres will be distributed uniformly over the entire region of volume $\propto R^3$. On the other hand, at low temperatures, most of the particles would have formed a small compact core of radius R_c . In both these limits, the partition function can be approximated by restricting the domain of integration and replacing the exact potential by the average potential energy of the system. We will then obtain, for the high temperature phase,

$$E_+ = 3N/2\beta - 3Gm^2N^2/5R, \quad (3.30)$$

$$F_+ = -(N/\beta) \ln(V/N - b) + (3N/2\beta) \ln \beta - 3Gm^2N^2/5R, \quad (3.31)$$

$$S_+ = N \ln(V/N - b) - \frac{3}{2}N \ln \beta, \quad (3.32)$$

and for the low temperature phase

$$E_- = 3N/2\beta - \frac{3}{5}(4\pi/3gb)^{1/3}Gm^2N^{5/3}, \quad (3.33)$$

$$F_- = (N/\beta) \ln b(g-1) + (3N/2\beta) \ln \beta - \frac{3}{5}(4\pi/3gb)^{1/3}Gm^2N^{5/3}, \quad (3.34)$$

$$S_- = N \ln b(g-1) - \frac{3}{2}N \ln \beta. \quad (3.35)$$

In arriving at these expressions, we have assumed that the radius R_c on the N body condensate is given

by

$$\frac{4}{3} \pi R_c^3 = gNb, \quad (3.36)$$

where g is a geometrical factor of the order of unity which depends on the nature of the close packing.

The transition temperature at which the description switches over from (3.30)–(3.32) to (3.33)–(3.35) can be easily estimated by equating the free energies of the two phases. This gives, for the critical temperature, the relation

$$\ln\left(\frac{V_{\text{core}} - Nb}{V - Nb}\right) = \frac{3}{5} \beta_c Gm^2 N \left(\frac{1}{R} - \frac{1}{R_{\text{core}}}\right), \quad (3.37)$$

which can be approximated further by noting that

$$V_{\text{core}} - Nb = (g - 1)Nb \approx Nb \approx V_{\text{core}}, \quad (3.38)$$

and that $V \gg V_{\text{core}}$. We then get

$$T_c \approx \frac{3}{5} (Gm^2 N / R_c) / \ln(V / V_{\text{core}}). \quad (3.39)$$

Note that this expression is quite similar to (3.16).

The expressions obtained above are really valid only at $T \gg T_c$ and at $T \ll T_c$. However, if the system really undergoes a sharp phase transition at $T \approx T_c$ then we will incur negligible error in extrapolating the formulas all the way up to T_c from both ends. We will then get a T – E relation which looks very similar to the ones we have obtained earlier.

The microcanonical distribution can be evaluated in the same approximation. We will then see that the phase transition is replaced by a large region of negative specific heat.

3.5. Two dimensional gravity

In the models studied so far we have put in a short distance cutoff to the potential. This cutoff is absolutely essential for the existence of the integrals defining the microcanonical distribution (for $N > 2$) and canonical distribution (for any N). If this cutoff is removed, then the phase volume – and consequently, the entropy – will increase without bound.

The situation is different in two dimensions. It turns out that the microcanonical distribution exists for a two dimensional gravitating system even in the absence of a cutoff. However, the canonical description exists only above a critical temperature. To see this, let us consider the phase volume $g(E)$,

$$g(E) = \int \prod_{i=1}^N d^2x_i d^2p_i \delta(E - H) = A \int \prod_{i=1}^N d^2x_i \left(E - \frac{1}{2} \sum_{i \neq j} 2Gm^2 \ln|x_i - x_j|\right)^{N-1}, \quad (3.40)$$

where we have used the Hamiltonian for a two-dimensional gravitating system of N particles with logarithmic potential,

$$H = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \frac{1}{2} \sum_{i \neq j} 2Gm^2 \ln|x_i - x_j|, \quad (3.41)$$

and performed the momentum integrations. (Note that the gravitational potential due to a point mass m located at the origin of a two dimensional space is $\phi(r) = 2Gm \ln r$.) We will assume that the particles are confined to a large square box of length $2L$ so that the range of each x_i is $(-L, +L)$. The integral is then well defined for large x_i . It is easy to see that the integral does not diverge at short distances even though there is no cutoff: the expression in the large parentheses in (3.40) does contribute a divergent factor, $(\ln s)^{N-1}$, near the origin, but this divergence is suppressed by the factor s in the area element $d^2x = s ds$. Thus $g(E)$ is finite, though not calculable in closed form.

Let us consider the partition function, which is given by

$$\begin{aligned} Z(\beta) &= \int \prod_{i=1}^N d^2p_i d^2x_i \exp(-\beta H) \\ &= \left(\frac{2\pi}{\beta}\right)^N \int \prod_{i=1}^N d^2x_i \exp\left[-\beta\left(\frac{1}{2} \sum_{i \neq j} 2Gm^2 \ln|x_i - x_j|\right)\right]. \end{aligned} \quad (3.42)$$

It does not exist for all β . Simple power counting near the origin shows that the integral exists only if

$$\beta < \beta_c = 2[Gm^2(N-1)]^{-1}. \quad (3.43)$$

In other words, the canonical description of the system exists only at sufficiently high temperatures [Salzberg 1965] (note that the gravitational potential used by Salzberg differs by a factor 2 from the one used above).

The same result can be obtained in a somewhat different manner, by computing the volume dependence of Z . Introducing the variables $y_i = x_i/L$ in (3.42) and differentiating the expression with respect to L , we can easily obtain the relation

$$\frac{\partial Z(\beta, L)}{\partial L} = \frac{2N}{L} [1 - \frac{1}{2}\beta Gm^2(N-1)]Z(\beta, L). \quad (3.44)$$

Since the two dimensional volume V is just πL^2 , the pressure P can be written as

$$P = \beta^{-1} \partial \ln Z / \partial V = (2\pi L\beta)^{-1} \partial \ln Z / \partial L. \quad (3.45)$$

Combining (3.44) and (3.45) we immediately get the equation of state

$$PV/T = N[1 - (N-1)Gm^2/2T]. \quad (3.46)$$

This equation reconfirms the trouble at low temperatures: the pressure becomes negative for all V if $T < T_{\min} = \frac{1}{2}(N-1)Gm^2 \approx \frac{1}{2}NGm^2$. The system, if it is in contact with a heat bath, will collapse under its own gravity. A short distance cutoff will make Z finite and will provide a stable ground state with positive pressure.

The equation of state derived above is exact in the sense that it is obtained from the exact partition function of the system. Katz and Lynden-Bell [1978] have performed an analysis of this system in the mean field limit. In the mean field limit, the system is described by the smooth gravitational potential

$\phi(\mathbf{x})$ and density distribution $\rho(\mathbf{x})$, which are related by the two-dimensional Poisson equation,

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{d\phi}{dr} \right) = 4\pi G\rho(r). \tag{3.47}$$

The thermodynamical equilibrium is introduced by taking

$$\rho(r) = A \exp[-\beta\phi(r)], \tag{3.48}$$

where the constants A and β are to be determined in terms of the total mass M and energy E of the system. (The domain of validity of such a mean field approximation will be discussed in chapter 4.) It turns out that these equations can be solved exactly [Ostriker 1964, Stodolkiewicz 1963] to give

$$\phi(r) = 2GM \ln r + 2\beta^{-1} \ln[1 - \frac{1}{2}GM\beta(1 - r^2/R^2)] + \text{constant}, \tag{3.49}$$

where R is the radius of the confining box. All other physical variables like density $\rho(r)$, pressure on the confining wall $P(R) = \beta^{-1}\rho(R)$, and the energy

$$E = \frac{M}{\beta} - \frac{1}{2} \int \rho\phi \, d\mathbf{x}, \tag{3.50}$$

can be computed from $\phi(r)$. We get

$$E = \frac{1}{2}GM^2 [2 \ln(r/R) + 2(T/T_c) + (T/T_c)^2 \ln(1 - T_c/T)], \tag{3.51}$$

$$PV = Nk(T - T_c), \tag{3.52}$$

where $T_c = GM^2/2Nk = \frac{1}{2}Gm^2N/k$ is the same critical temperature as that found earlier. It is remarkable that both the mean field description and the exact analysis leads to the same equation of state for the system. It is clear that the system cannot exist for $T < T_c$; the pressure becomes negative, the potential at the origin diverges.

The $T(E)$ curve for this system – in the mean field approximation – is shown in fig. 3.5. Note that

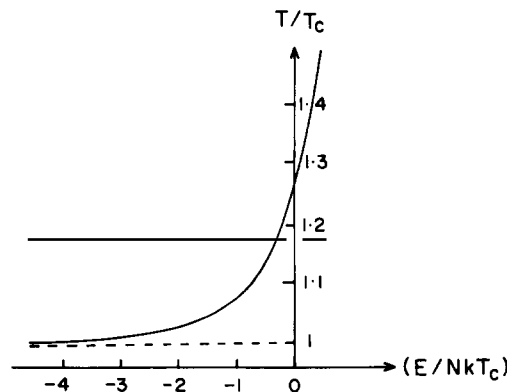


Fig. 3.5. The $T(E)$ relation for a two dimensional gravitating system, in the mean field limit. There is no region of negative specific heat but the system exhibits a lower bound on the temperature.

there is no region of negative specific heat for this two dimensional system. As the energy of the system is lowered, the temperature continuously decreases and approaches T_c asymptotically. We will see in chapter 4 that the three dimensional isothermal solution behaves very differently.

4. Mean field equilibrium of gravitating systems

4.1. Introduction

The statistical mechanical description presented so far is exact as regards the granularity of the system. We have used the correct N particle distributions and have not resorted to any continuum approximations. In this section we shall study the physics of the gravitating systems in the mean field limit, which ignores the granularity and correlations present in the N particle system. As we shall see, the system exhibits several interesting features even in this limit.

4.2. Mean field as the saddle point limit

The statistical definition of entropy used in the previous sections involves a $6N$ dimensional integration over the whole phase space. Under certain circumstances, it is possible to approximate this expression by one which involves only a six dimensional integration. Such an approximation, which we shall call the mean field approximation, reduces the problem to a tractable level.

Consider a system of N particles interacting with each other through the two body potential $U(x, y)$. The entropy S of this system, in the microcanonical description, is defined through the relation

$$e^S = g(E) = \frac{1}{N!} \int d^{3N}x d^{3N}p \delta(E - H) = \frac{A}{N!} \int d^{3N}x \left(E - \frac{1}{2} \sum_{i \neq j} U(\mathbf{x}_i, \mathbf{x}_j) \right)^{3N/2}, \quad (4.1)$$

wherein we have performed the momentum integrations and replaced $(3N/2 - 1)$ by $3N/2$. We shall approximate the expression in (4.1) in the following manner.

Let the spatial volume V be divided into M (with $M \ll N$) cells of equal size, large enough to contain many particles but small enough for the potential to be treated as a constant inside each cell. (It is assumed that such an intermediate scale exists.) Instead of integrating over the particle co-ordinates $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ we shall sum over the number of particles n_a in the cell centred at \mathbf{x}_a (where $a = 1, 2, \dots, M$). Using the standard result that the integration over $(N!)^{-1} d^{3N}x$ can be replaced by

$$\sum_{n_1=1}^{\infty} \frac{1}{n_1!} \sum_{n_2=1}^{\infty} \frac{1}{n_2!} \cdots \sum_{n_M=1}^{\infty} \frac{1}{n_M!} \delta\left(N - \sum_a n_a\right) \left(\frac{V}{M}\right)^N, \quad (4.2)$$

we can rewrite (4.1) as

$$\begin{aligned} e^S &= \sum_{n_1=1}^{\infty} \frac{1}{n_1!} \sum_{n_2=1}^{\infty} \frac{1}{n_2!} \cdots \sum_{n_M=1}^{\infty} \frac{1}{n_M!} \delta\left(N - \sum_a n_a\right) \left(\frac{V}{M}\right)^N \left(E - \frac{1}{2} \sum_{a \neq b}^M n_a U_{ab} n_b\right)^{3N/2} \\ &\approx \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \cdots \sum_{n_M=1}^{\infty} \delta\left(N - \sum_a n_a\right) \exp S\{n_a\}, \end{aligned} \quad (4.3)$$

where

$$S[\{n_a\}] = \frac{3}{2} N \ln \left(E - \frac{1}{2} \sum_{a \neq b}^M n_a U(\mathbf{x}_a, \mathbf{x}_b) n_b \right) - \sum_{a=1}^M n_a \ln(n_a M / eV). \quad (4.4)$$

In arriving at the last expression we have used the Stirling approximation for the factorials and ignored the unimportant constant A . The mean field limit is now obtained by retaining in the sum in (4.3) only the term for which the summand reaches the maximum value, subject to the constraint on the total number. That is, we claim

$$\sum_{\{n_a\}} e^{S[\{n_a\}]} \approx e^{S[\{n_{a,\max}\}]}, \quad (4.5)$$

where $n_{a,\max}$ is the solution to the variational problem

$$\left(\frac{\delta S}{\delta n_a} \right)_{n_a = n_{a,\max}} = 0 \quad \text{with} \quad \sum_{a=1}^M n_a = N. \quad (4.6)$$

Imposing this constraint by a lagrange multiplier and using expression (4.4) for S , we obtain the equation satisfied by $n_{a,\max}$,

$$\frac{1}{T} \sum_{b=1}^M U(\mathbf{x}_a, \mathbf{x}_b) n_{b,\max} + \ln(n_{a,\max} M / V) = \text{constant}, \quad (4.7)$$

where we have defined the temperature T as

$$\frac{1}{T} = \frac{3}{2} N \left(E - \frac{1}{2} \sum_{a \neq b}^M n_a U(\mathbf{x}_a, \mathbf{x}_b) n_b \right)^{-1} = \beta. \quad (4.8)$$

We see from (4.4) that this expression is also equal to $\partial S / \partial E$; therefore T is indeed the correct thermodynamic temperature. We can now return back to the continuum limit by the replacements

$$n_{a,\max} M / V = \rho(x_a), \quad \sum_{a=1}^M \rightarrow \frac{M}{V}. \quad (4.9)$$

In this limit the extremum solution (4.7) is given by

$$\rho(\mathbf{x}) = A \exp[-\beta\phi(\mathbf{x})], \quad \phi(\mathbf{x}) = \int d^3y U(\mathbf{x}, \mathbf{y}) \rho(\mathbf{y}), \quad (4.10)$$

which, in the case of gravitational interactions, becomes

$$\rho(\mathbf{x}) = A \exp[-\beta\phi(\mathbf{x})], \quad \phi(\mathbf{x}) = -G \int \frac{\rho(\mathbf{y}) d^3y}{|\mathbf{x} - \mathbf{y}|}. \quad (4.11)$$

Equation (4.11) represents the equilibrium configuration for a gravitating system in the mean field limit. The constant β is already determined through (4.8) in terms of the total energy of the system.

The constant A has to be fixed in terms of the total number (or mass) of the particles in the system. A more formal derivation of the above result can be given using the functional integral representation of the partition function. It turns out that the saddle point approximation of the functional integral leads to the mean field description (see, e.g., Horowitz and Katz [1977]).

Two important points need to be noted about the mean field result we have obtained: (i) The various manipulations in (4.1) to (4.7) tacitly assume that the expressions we are dealing with are finite. Unfortunately, for gravitational interactions without a short distance cutoff, the quantity e^S – and hence all the terms we have been handling – are divergent. We should, therefore, remember that a short distance cutoff is needed to justify the entire procedure and that (4.11) – which is based on a strict r^{-1} potential and does not incorporate any such cutoff – can only be approximately correct. We shall continue to work with (4.11) because of its mathematical convenience and will describe the effects due to the short distance cutoff in section 4.5. (ii) Condition (4.7), which we have written down, is only an *extremum* condition. To make sure that the entropy is a true maximum we need to look at the second variation of S . The procedure in (4.5) is justified only if S is at least at a local maximum. To decide whether the extremum is actually a maximum or not we need to perform a variational calculation of S up to second order. This is done in section 4.4.

We have derived the mean field limit directly from the phase volume by using the saddle point approximation. It is also possible to obtain the same result in a more intuitive approach, which is very useful for studying the dynamical evolution of the system. This is the approach which we shall follow for the rest of this section.

The central quantity in this approach is the one-particle distribution function $f(\mathbf{x}, \mathbf{p}, t)$, which may be defined through the (operational) relation

$$dm = f(\mathbf{x}, \mathbf{p}, t) d^3\mathbf{x} d^3\mathbf{p}, \quad (4.12)$$

where dm is the total mass of the particles contained in the cell $(\mathbf{x}, \mathbf{p}; \mathbf{x} + d^3\mathbf{x}, \mathbf{p} + d^3\mathbf{p})$. This function is obtained from the full N body distribution function of the system by integrating out $(N - 1)$ variables,

$$f(\mathbf{x}_1, \mathbf{p}_1, t) \equiv \int f_N(\mathbf{x}_1, \mathbf{p}_1; \mathbf{x}_2, \mathbf{p}_2; \mathbf{x}_3, \mathbf{p}_3; \dots; \mathbf{x}_N, \mathbf{p}_N) d^3\mathbf{x}_2 d^3\mathbf{p}_2 \cdots d^3\mathbf{x}_N d^3\mathbf{p}_N. \quad (4.13)$$

We shall assume that the functions which appear in (4.12) and (4.13) are smooth. This excludes, for example, the function f being given as a sum of delta functions. If we use such a smoothed-out description, then it is clear that f contains far less information than f_N . In particular the correlation between the particles is washed out in $f(\mathbf{x}, \mathbf{p}, t)$. Thus f could provide a useful description of the system only as long as the correlations are not significant.

Given the fundamental Hamiltonian of the system, it should be possible to write down an (approximate) equation satisfied by f . This task, however, is not easy and we shall postpone a detailed discussion of this question to chapter 5. For the present purpose, it is enough to note that the equation obeyed by f has the following form:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} - \nabla \phi \cdot \frac{\partial f}{\partial \mathbf{v}} = C(f), \quad (4.14)$$

where $\mathbf{v} = \mathbf{p}/m$ and $\phi(\mathbf{x}, t)$ is the mean gravitational field produced by f ,

$$\phi(\mathbf{x}, t) = -G \int \frac{f(\mathbf{y}, \mathbf{v}, t) d^3\mathbf{y} d^3\mathbf{v}}{|\mathbf{x} - \mathbf{y}|}. \quad (4.15)$$

The right hand side of (4.14), $C(f)$, describes the effect of ‘‘collisions’’ on the evolution of f . As we shall see in chapter 5, the collisional evolution of the system, driven by $C(f)$, exhibits two reasonable properties: (i) The mean field energy E ,

$$\begin{aligned} E = K + U &\equiv \frac{1}{2} \int \mathbf{v}^2 f d^3\mathbf{x} d^3\mathbf{v} - \frac{G}{2} \int \frac{f(\mathbf{x}, \mathbf{v})f(\mathbf{x}', \mathbf{v}')}{|\mathbf{x} - \mathbf{x}'|} d\mathbf{x} d\mathbf{v} d\mathbf{x}' d\mathbf{v}' \\ &= \frac{1}{2} \int \mathbf{v}^2 f d^3\mathbf{x} d^3\mathbf{v} + \frac{1}{2} \int \rho(\mathbf{x})\phi(\mathbf{x}) d^3\mathbf{x}, \end{aligned} \quad (4.16)$$

and mass M ,

$$M = \int f d^3\mathbf{x} d^3\mathbf{v}, \quad (4.17)$$

are conserved in the evolution. (ii) The mean field entropy, defined by

$$S = - \int f \ln f d^3\mathbf{x} d^3\mathbf{p}, \quad (4.18)$$

does not decrease (and, generically, increases) during the evolution. It is, therefore, reasonable to ask: In the class of all f with the same E and M , which f maximises the mean field entropy S ?

Since we are using a gravitational potential without a short distance cutoff, there will be no *global* maximum for the entropy; that is, we can construct configurations in which S is arbitrarily high. It is easy to devise such configurations by distributing part of the matter in a tightly bound core and the rest of the material as a halo at a large distance. The question, therefore, is really the following: are there configurations which are *local* maxima of the entropy? To answer this question we have to consider the variation

$$\delta S = (\delta S / \delta f) \delta f + \frac{1}{2} (\delta^2 S / \delta f^2) (\delta f)^2 + \dots \quad (4.19)$$

The vanishing of the first term defines the extremum condition and the sign of the second term decides whether the extremum is stable or merely a saddle point.

This variational problem is solved in appendix 4.1. The extremum condition is, of course, the same as (4.11) obtained above and corresponds to the saddle point evaluation of the phase volume. We will discuss this extremum solution in the next section and the second variation of the entropy in section 4.4.

4.3. The mean field equilibrium. The isothermal sphere

It can easily be shown that among all solutions to (4.11) the spherically symmetric configuration maximises the entropy (see, e.g., Antonov [1962]). This solution represents what is known as the gravitational isothermal sphere. To analyse the properties of this solution, it is convenient to combine

the two equations in (4.11) into the single equation

$$\nabla^2 \phi = 4\pi G A e^{-\beta\phi} = 4\pi G \rho_c e^{-\beta[\phi(x) - \phi(0)]}, \quad (4.20)$$

where $\rho_c = \rho(0)$. Continuity requirements at the surface of the confining volume imply the boundary condition $\phi(R) = -GM/R$ on the potential.

This extremum solution is parametrized by two positive definite constants β and ρ_c . The crucial question is the following: Can any value of E ($-\infty < E < +\infty$), R ($0 < R < \infty$) and M ($0 < M < \infty$) be accommodated by a suitable choice of ρ_c and β ? We will see that the answer is *no*. There is a lower bound on the allowed value for the combination RE/GM^2 ; if this quantity is lower than a critical value $\lambda_c \approx -0.335$ then an isothermal sphere solution cannot exist [Antonov 1962; Lynden-Bell and Wood 1968; Padmanabhan 1989c].

This can be most easily seen if we use a different set of (dimensionless) variables. Using ρ_c and β we define a length, mass and potential,

$$L_0 \equiv (4\pi G \rho_c \beta)^{1/2}, \quad M_0 \equiv 4\pi \rho_c L_0^3, \quad \phi_0 \equiv \beta^{-1} = GM_0/L_0, \quad (4.21)$$

and express the radial distance r , density ρ , mass $M(r)$ contained within a radius r , and potential ϕ in terms of these variables in dimensionless form,

$$x \equiv r/L_0, \quad n \equiv \rho/\rho_c, \quad m \equiv M(r)/M_0, \quad y \equiv \beta[\phi - \phi(0)]. \quad (4.22)$$

These variables (y, n, m) satisfy the equations

$$y' = m/x^2, \quad m' = nx^2, \quad n' = mn/x^2. \quad (4.23)$$

In terms of $y(x)$ the isothermal equation becomes

$$\frac{1}{x^2} \frac{d}{dx} \left(x^2 \frac{dy}{dx} \right) = e^{-y}, \quad (4.24)$$

with the boundary condition $y(0) = y'(0) = 0$.

This equation is invariant under the transformation $y \rightarrow y + a$, $x \rightarrow kx$ with $k^2 = e^a$. This invariance implies that, given a solution with some value of $y(0)$, we can obtain the solution with any other value of $y(0)$ by simple scaling. Therefore, only one of the two integration constants in (4.24) is really non-trivial. This fact allows us to reduce the degree of the equation from two to one by a judicious choice of variables [Chandrasekhar 1939]. One such set of variables is the following:

$$v \equiv m/x, \quad u \equiv nx^3/m = nx^2/v. \quad (4.25)$$

In terms of u and v the equation describing the isothermal sphere can be written as

$$\frac{u}{v} \frac{dv}{du} = -\frac{u-1}{u+v-3}. \quad (4.26)$$

The boundary condition $y(0) = y'(0) = 0$ translates into the following: $v = 0$ for $u = 3$, and $dv/du = -5/3$ at $(u, v) = (3, 0)$.

The major advantage of the variables u, v is the following: The *form* of the solution to (4.26) can be ascertained, without any numerical work, purely from the study of the singular points of eq. (4.26). The solution curve starts at $(3, 0)$ [which corresponds to $x = 0$ in (4.24)] and spirals indefinitely around the point $(1, 2)$ as x tends to infinity (see fig. 4.1). All isothermal spheres must necessarily lie on this curve. The nature of this curve allows us to put interesting bounds on physical quantities including energy [Padmanabhan 1989c].

To see this, we shall compute the total energy E of the isothermal sphere. The potential and kinetic energies are

$$U = - \int_0^R \frac{GM(r)}{r} \frac{dM}{dr} dr = - \frac{GM_0^2}{L_0} \int_0^{x_0} mnx dx ,$$

$$K = \frac{3}{2} \frac{M}{\beta} = \frac{3}{2} \frac{GM_0^2}{L_0} m(x_0) = \frac{GM_0^2}{L_0} \frac{3}{2} \int_0^{x_0} nx^2 dx , \tag{4.27}$$

where $x_0 = R/L_0$. The total energy is, therefore,

$$E = K + U = \frac{GM_0^2}{2L_0} \int_0^{x_0} dx (3nx^2 - 2mnx)$$

$$= \frac{GM_0^2}{2L_0} \int_0^{x_0} dx \frac{d}{dx} (2nx^3 - 3m) = \frac{GM_0^2}{L_0} (n_0x_0^3 - \frac{3}{2}m_0) , \tag{4.28}$$

where $n_0 = n(x_0)$ and $m_0 = m(x_0)$. The dimensionless quantity RE/GM^2 is given by

$$\lambda = \frac{RE}{GM^2} = \frac{L_0x_0}{GM_0^2m_0^2} \frac{GM_0^2}{L_0} (n_0x_0^3 - \frac{3}{2}m_0) = \frac{1}{v_0} (u_0 - \frac{3}{2}) . \tag{4.29}$$

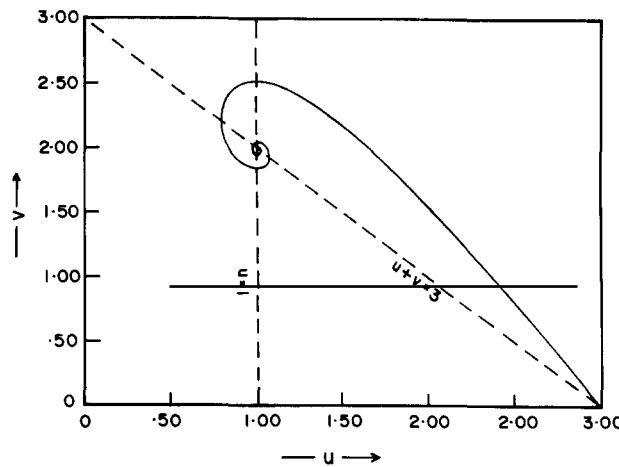


Fig. 4.1. The $u-v$ curve for the isothermal sphere without any short distance cutoff.

Note that the combination RE/GM^2 is a function of (u, v) alone. Let us now consider the constraints on λ . Suppose we specify some value for λ by specifying R, E and M . Then such an isothermal sphere *must* lie on the curve

$$v = \lambda^{-1} \left(u - \frac{3}{2} \right), \quad (4.30)$$

which is a straight line through the point $(1.5, 0)$ with slope λ^{-1} . Since, on the other hand, *all* isothermal spheres must lie on the $u-v$ curve, an isothermal sphere can exist only if the line in (4.30) intersects the $u-v$ curve.

For large positive λ (positive E) there is just one intersection (see fig. 4.2). When $\lambda = 0$ (zero energy), we still have a unique isothermal sphere. [For $\lambda = 0$, (4.30) is a vertical line through $u = 3/2$.] When λ is negative (negative E), the line can cut the $u-v$ curve at more than one point; thus more than one isothermal sphere can exist with a given value of λ . [Of course, specifying M, R, E individually will remove this non-uniqueness.] But as we decrease λ (more and more negative E) the line in (4.30) will slope more and more to the left; and when λ is smaller than a critical value λ_c , the intersection will cease to exist. Thus no isothermal sphere can exist if RE/GM^2 is below a particular critical value λ_c . This fact follows immediately from the nature of the $u-v$ curve and eq. (4.30). The numerical value of λ_c can also be found by simple analytic approximations to the curve around the singular point. It turns out to be about -0.335 .

It is instructive to look at the variation of λ as a function of x_0 ; see fig. 4.3. If we fix ρ_c and β and increase R , we will be increasing $x_0 = R/L_0$. This will take us along the curve from $(3, 0)$ – which corresponds to $x_0 = 0$ – towards the spiral. At each x_0 we have a corresponding (u_0, v_0) and λ . The behaviour of λ as a function of x_0 can be easily read off from the $u-v$ curve. We see that λ decreases

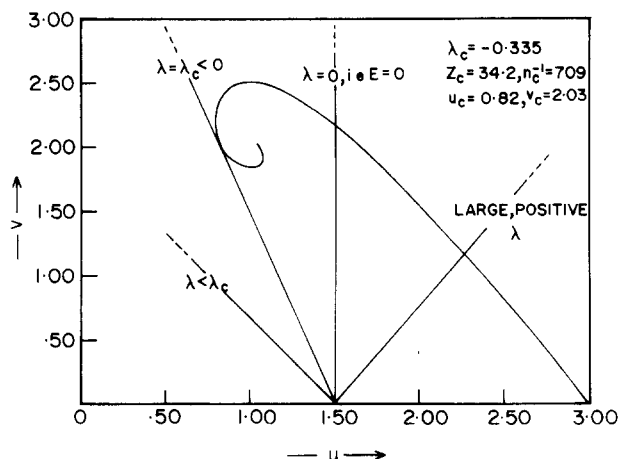


Fig. 4.2. Construction illustrating the origin of the lower bound on λ for isothermal spheres.

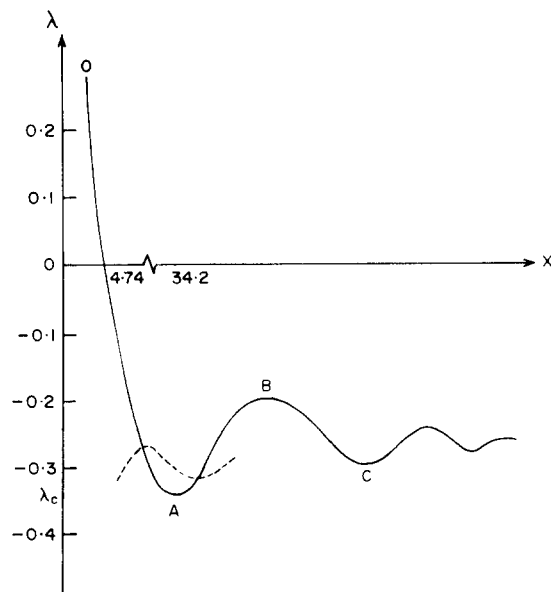


Fig. 4.3. The behaviour of λ as a function of the dimensionless radius x for an isothermal sphere. The diagram is purely schematic and not drawn to scale. The dashed line suggests the behaviour of the entropy as a function of some suitably chosen generalised coordinate.

from infinity (at $x=0$) to zero, and continues to decrease taking negative values until it reaches the minimum value -0.335 at some $x = x_c$. As x increases further, λ increases and oscillates. In fact, the $\lambda(x)$ curve has an infinite number of turning points; asymptotically (as $x \rightarrow \infty$, $u \rightarrow 1$, $v \rightarrow 2$), λ takes the value -0.25 . From this discussion, we obtain a further characterisation of the point x_c : it is the first turning point for $\lambda(x)$, i.e., the *smallest* root of the equation $d\lambda/dx = 0$.

The $u-v$ curve also implies a bound on the temperature of the system [Lynden-Bell and Wood 1968]. It is clear that for any isothermal sphere v is bounded from above, i.e., $v < v_{\max}$, where $v_{\max} \approx 2.5$. Since

$$v = m/x = (M/M_0)/(R/L_0) = (GM/R)\beta, \tag{4.31}$$

we immediately get

$$T > T_{\min}, \quad T_{\min} = 0.4GM/R. \tag{4.32}$$

The $T(E)$ curve for the isothermal sphere can be determined by combining the relations

$$T = (GM^2/R)v, \quad E = (GM^2/R)v^{-1}(u - \frac{3}{2}). \tag{4.33}$$

The form of the curve is shown in fig. 4.4. The specific heat is positive along AB (which represents a high temperature phase) and is negative along BC. The onset of negative specific heat occurs at about $x \approx 9$ corresponding to a density contrast of $n_0^{-1} \approx 32$. As we shall see later, the branch CD is unstable and is not physically realisable. (Point C corresponds to a density contrast of 709; thus isothermal spheres with a density contrast in the range (32, 709) have a negative specific heat.) Since we have not introduced any short distance cutoff in the system we do not have a low temperature phase with positive specific heat. (We shall see in section 4.5 that a short distance cutoff makes the curve turn to the left on the CD branch and produces a second region of positive specific heat.) Figure 4.4 is analogous to fig. 3.3, which describes the $T(E)$ curve for a binary model *in the absence* of a short distance cutoff. The modification of fig. 4.4 in the presence of a short distance cutoff will be very similar to the modification

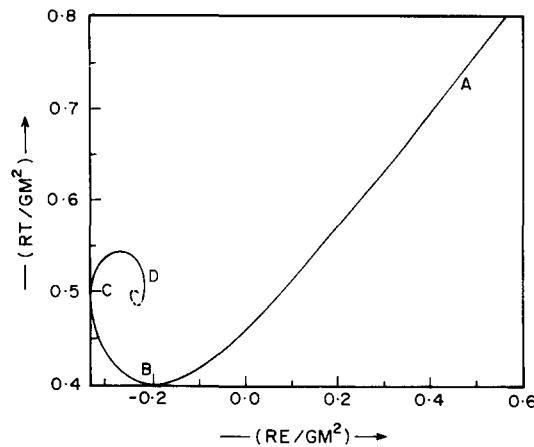


Fig. 4.4. The $T(E)$ curve for the isothermal sphere without any short distance cutoff. The specific heat is positive along AB and negative along BC. (We shall see later that the portion after C is unstable and not physically realisable.)

of fig. 3.3 into fig. 3.2. These results have interesting implications for the mean field thermodynamics. We shall take up this discussion in section 4.6, after first settling two other important issues.

4.4. Stability of the mean field solution

We have seen that the variational problem has no solution if $\lambda < \lambda_c$. But even when $\lambda > \lambda_c$, the extremum solution need not be a local maximum.

If $\lambda > \lambda_c$, then we will definitely have one point of intersection of the straight line (4.30) and the $u-v$ curve; in fact, we can have many points of intersection for a certain range of parameters. They all represent extrema of S . To know which of these are “stable” – i.e. real maxima of S – it is necessary to study the second variation of entropy. This analysis was originally performed by Antonov; he showed that *all* isothermal spheres located to the right of A on the $\lambda(x)$ curve of fig. 4.3 are unstable. These correspond to configurations with $RE/GM^2 > -0.335$ and $\rho(0)/\rho(R) > 709$. This result can also be proved using a general theorem due to Katz [1978, 1979]. Katz has extended Poincaré’s method of linear series to study the stability beyond the second turning point. He shows how the stability questions can be decided from an inspection of the equilibrium curves in the $\beta-E$ plane. This general method, however, cannot provide detailed information on the nature and form of the modes which actually trigger the instability. To obtain such information, we have to study the variational problem and explicitly construct the mode which leads to an increase in entropy [Antonov 1962; Padmanabhan 1989c].

Consider two density distributions, $\rho_e(\mathbf{x})$ and $\rho_e(\mathbf{x}) + \delta\rho(\mathbf{x})$; we take $\rho_e(\mathbf{x})$ as an equilibrium isothermal sphere with energy E , mass M and radius R (with, of course, $RE/GM^2 > -0.335$, since such a solution will not exist otherwise). We put $\delta\rho(\mathbf{x}) = \varepsilon a(\mathbf{x})$, where ε is an infinitesimal parameter, and assume that $\rho + \delta\rho$ also has the same M and E . The entropy difference between these two configurations is computed in appendix 4.1. There we show that

$$S(\rho_e + \varepsilon a) - S(\rho_e) = \varepsilon^2 \delta^2 S(a) + O(\varepsilon^3) = -\frac{\varepsilon^2}{3MT^2} \left(\int d\mathbf{x} \phi_e a \right)^2 - \varepsilon^2 \int d\mathbf{x} \left(\frac{a\psi}{2T} + \frac{a^2}{2\rho_e} \right). \quad (4.34)$$

Here ϕ_e is the background potential due to ρ_e and ψ is the potential due to $a(\mathbf{x})$,

$$\nabla^2 \psi = 4\pi G a, \quad \psi(\mathbf{x}) = -G \int \frac{a(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|} d\mathbf{y} \quad (4.35)$$

[ψ is called $\delta\phi$ in appendix 4.1]. This expression already takes account of the fact that ρ and $\rho + \delta\rho$ have the same energy E . The constraint that the total mass should not change under the perturbation implies that

$$\int d^3\mathbf{x} a(\mathbf{x}) = 0. \quad (4.36)$$

It is enough to confine our attention to spherically symmetric perturbations [$a(\mathbf{x}) = a(|\mathbf{x}|) = a(r)$], for which it is convenient to introduce the “mass perturbation” $q(r)$ by the relation

$$a(r) = \frac{1}{4\pi r^2} \frac{dq}{dr}. \quad (4.37)$$

The force due to the perturbed distribution is $\psi' = Gq(r)/r^2$. Since this force has to be finite everywhere including at $r = 0$, we see that $q(r)$ should grow at least as r^3 near $r = 0$; in particular $q(0) = 0$. The constraint (4.36) implies that $q(R) = q(0)$; so, altogether, we get the boundary conditions on $q(r)$ to be

$$q(0) = q(R) = 0. \quad (4.38)$$

In terms of $q(r)$, (4.34) becomes

$$\begin{aligned} [S(\rho_e + \delta\rho) - S(\rho_e)] &\equiv A[a] = -\frac{1}{3MT^2} \left(\int_0^R dr \phi_e \frac{dq}{dr} \right)^2 - \int_0^R dr \left(\frac{\psi q'}{2T} + \frac{q'^2}{8\pi\rho_e r^2} \right) \\ &= \frac{1}{3MT^2} \left(\int_0^R dr \phi_e' q \right)^2 + \frac{1}{2T} \int_0^R \psi' q dr + \int_0^R dr q \frac{d}{dr} \left(\frac{q'}{8\pi\rho_e r^2} \right). \end{aligned} \quad (4.39)$$

In arriving at the above form we have integrated by parts, used (4.38) and set the dummy variable ϵ to zero. Now using the fact that $\psi' = Gq/r^2$ we get

$$\begin{aligned} A[a] &= -\frac{1}{3MT^2} \left(\int_0^R dr \phi_e' q \right)^2 + \frac{G}{2T} \int_0^R \frac{q^2}{r^2} dr + \int_0^R dr q \frac{d}{dr} \left(\frac{q'}{8\pi\rho_e r^2} \right) \\ &= -\frac{1}{3MT^2} \left(\int_0^R dr \phi_e' q \right)^2 + \frac{1}{2} \int_0^R dr q \left(\frac{G}{Tr^2} + \frac{d}{dr} \frac{1}{4\pi\rho_e r^2} \frac{d}{dr} \right) q \\ &\equiv -\int_0^R \int_0^R dr_1 dr_2 q(r_1) K(r_1, r_2) q(r_2), \end{aligned} \quad (4.40)$$

where the kernel K is given by

$$K(r_1, r_2) = -\frac{\phi_e'(r_1)\phi_e'(r_2)}{3MT^2} + \frac{1}{2}\delta(r_1 - r_2) \left[\frac{G}{Tr_1^2} \frac{d}{dr_1} \left(\frac{1}{4\pi\rho_e r_1^2} \frac{d}{dr_1} \right) \right]. \quad (4.41)$$

To decide about the stability of the configuration $\rho_e(r)$ we ask the question: Can one find a $q(r)$ with $q(0) = q(R) = 0$ which will make A positive? Or, equivalently, can the expression

$$-A = \int_0^R dr_1 \int_0^R dr_2 q(r_1) K(r_1, r_2) q(r_2) \quad (4.42)$$

attain negative values?

It is well known that expression (4.42) can reach negative values if and only if the ‘‘matrix’’ $K(r_1, r_2)$ has at least one negative eigenvalue. If that happens then A will be positive; there will exist a neighbouring configuration $(\rho_e + \delta\rho)$ with higher entropy and ρ_e will not be a true local maximum. We,

therefore, have to look at the eigenvalue equation for K ,

$$\int_0^R dr_1 K(r, r_1) F_\xi(r_1) = \xi F_\xi(r), \quad (4.43)$$

and decide when ξ can be negative.

As before, let us consider a class of background isothermal spheres of fixed ρ_c and β – with different R . We know that for sufficiently small R , the configurations should be stable. This is because, for sufficiently small R , kinetic energy dominates over gravity and we are effectively dealing with an ideal gas (we demonstrate this result more rigorously in appendix 4.2). Let ξ_{\min} be the smallest eigenvalue; then for small enough R , $\xi_{\min} > 0$.

Now we increase R . Let us suppose that the instability sets in for some large enough R . Then ξ_{\min} must have become negative for this large R . It follows from continuity that the marginal stability occurs for a critical R_c at which $\xi_{\min} = 0$. For any R slightly larger than R_c we expect $\xi_{\min} < 0$. It can be easily seen that *all* configurations with $R > R_c$ are unstable.

To determine the critical radius R_c we have to solve the equation

$$0 = \int_0^R dr_1 K(r, r_1) F(r_1) = -\frac{\phi'_c(r)}{3MT^2} \int_0^R dr_1 \phi'_c(r_1) F(r_1) + \frac{1}{2} \left(\frac{G}{Tr^2} + \frac{d}{dr} \frac{1}{4\pi\rho_c r^2} \frac{d}{dr} \right) F(r), \quad (4.44)$$

or equivalently

$$\left[\frac{d}{dr} \left(\frac{1}{4\pi\rho_c r^2} \frac{d}{dr} \right) + \frac{G}{Tr^2} \right] F = \frac{2\phi'_c}{3MT^2} V = \frac{2V}{3MT^2} \frac{GM(r)}{r^2}, \quad (4.45)$$

$$V = \int_0^R dr_1 \phi'_c(r_1) F(r_1). \quad (4.46)$$

Note that, though V depends on F , V is a constant as far as (4.45) is concerned. We rescale the parameters in (4.45) by putting

$$\rho_c = \rho_c n, \quad M(r) = M_0 m, \quad r = L_0 x, \quad F = M_0 f(x), \quad (4.47)$$

where all the scaling variables have been defined previously. We now get

$$\hat{L}f(x) \equiv \left(\frac{d}{dx} \frac{1}{nx^2} \frac{d}{dx} + \frac{1}{x^2} \right) f(x) = -\Lambda \frac{m}{x^2}, \quad (4.48)$$

$$\Lambda = -\frac{1}{3m(x_0)} \frac{2V}{GM_0^2/L_0}. \quad (4.49)$$

This equation can be solved by the following method: We first note the action of the operator \hat{L} on

$m(x)$ and nx^3 ,

$$\hat{L}m(x) = m/x^2, \tag{4.50}$$

$$\hat{L}(nx^3) = \frac{d}{dx} \left(\frac{1}{nx^2} (3nx^2 + n'x^3) \right) + nx = \frac{d}{dx} \left(-\frac{m}{x} \right) + nx = \frac{m}{x^2}, \tag{4.51}$$

where we have used the relation $n' = -mn/x^2$. This suggests the ansatz

$$f(x) = c_1 nx^3 + c_2 m. \tag{4.52}$$

Substituting this expression in (4.48) we get $c_2 = -(c_1 + \Lambda)$, so that the solution is

$$f(x) = c_1 (nx^3 - m) - \Lambda m. \tag{4.53}$$

Given one solution to (4.48), in the form of (4.53), the second independent solution, $g(x)$, can be easily found from the Wronskian condition,

$$g = f \int \frac{nx^2 dx}{f^2}. \tag{4.54}$$

It is easily verified that $g(0) \neq 0$; therefore g is unacceptable. We only need to work with f in (4.53).

The solution $f(x)$ behaves as x^3 near $x = 0$ satisfying our requirement of a finite force at origin. The condition $f(x_0) = 0$ [corresponding to the constraint $q(R) = 0$] determines c_1

$$c_1 = \frac{m(x_0)\Lambda}{nx_0^3 - m(x_0)} = \frac{\Lambda}{u_0 - 1}. \tag{4.55}$$

We now only have to satisfy (4.46). The right hand side of (4.46) is

$$\int_0^R F\phi'_c dr = \frac{GM_0^2}{L_0} \int f(x) \frac{m}{x^2} dx = \frac{GM_0^2}{L_0} \left(\int_0^{x_0} c_1 nx^3 \frac{m}{x^2} dx - (\Lambda + c_1) \int_0^{x_0} m \frac{m}{x^3} dx \right). \tag{4.56}$$

Note that using (4.23) we can write

$$\int_0^{x_0} nx^3 \frac{m}{x^2} dx = - \int_0^{x_0} n'x^3 dx = -n_0x_0^3 + 3m_0 = 3m_0(1 - \frac{1}{3}u_0), \tag{4.57}$$

$$\begin{aligned} \int_0^{x_0} m \frac{m}{x^2} dx &= -\frac{m_0^2}{x_0} + \int_0^{x_0} \frac{2mm'}{x} dx = -\frac{m_0^2}{x_0} + 2 \int_0^{x_0} mnx dx \\ &= -m_0v_0 - 2 \int_0^{x_0} n'x^3 dx = -m_0v_0 + 2[3m_0(1 - \frac{1}{3}u_0)] = m_0(6 - 2u_0 - v_0). \end{aligned} \tag{4.58}$$

Substituting (4.57) and (4.58) into (4.56) we get

$$\frac{V}{GM_0^2/L_0} = c_1 \cdot 3m_0(1 - \frac{1}{3}u_0) - (c_1 + A)m_0(6 - 2u_0 - v_0). \quad (4.59)$$

Using expression (4.55) for c_1 and (4.49) for A this becomes

$$\frac{V}{GM_0^2/L_0} = \frac{\Lambda m_0}{u_0 - 1} (2u_0^2 + u_0v_0 - 7u_0 + 3) = -\frac{2}{3} \frac{V}{GM_0^2/L_0} \frac{2u_0^2 + u_0v_0 - 7u_0 + 3}{u_0 - 1}. \quad (4.60)$$

This equation determines the critical value x_c for stability. Since $V \neq 0$, we must have

$$1 + \frac{2}{3} \frac{2u_0^2 + u_0v_0 - 7u_0 + 3}{u_0 - 1} = \frac{4u_0^2 + 2u_0v_0 - 11u_0 + 3}{u_0 - 1} = 0. \quad (4.61)$$

This expression has a simple interpretation. Consider the derivative of the function $\lambda(x)$ defined in (4.30). We have

$$\begin{aligned} \frac{d\lambda}{dx} &= \frac{d}{dx} \left(\frac{1}{v} \left(u - \frac{3}{2} \right) \right) = \frac{u}{v} \frac{3-u-v}{x} - \frac{1}{v^2} \frac{v}{x} (u-1)(u-\frac{3}{2}) \\ &= \frac{1}{vx} (3u - u^2 - uv - u^2 + \frac{5}{2}u - \frac{3}{2}) = -\frac{1}{2vx} (4u^2 + 2uv - 11u + 3). \end{aligned} \quad (4.62)$$

Comparing (4.61) and (4.62) we notice that the critical point x_c is the smallest x at which $d\lambda/dx$ vanishes. From our previous discussion, we know that RE/GM^2 decreases from $x=0$ to $x=x_A$ after which it begins to increase. Thus the critical point of stability x_c is exactly the same point as A .

We can, of course, solve (4.61) numerically. Figure 4.5 shows the curve C_1 determined by the equation

$$4u^2 + 2uv - 11u + 3 = 0, \quad (4.63)$$

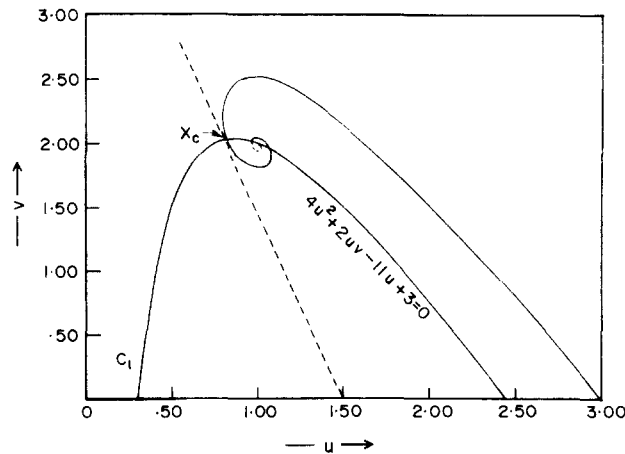


Fig. 4.5. Graphical construction determining the location of the critical radius at which the isothermal spheres become unstable.

superposed on the $u-v$ curve. Notice that the smallest x at which the $u-v$ curve and C_1 intersect is at $x_c = 34.2$.

Thus we reach the following conclusion: Isothermal spheres with $x < x_A$ are true maxima of entropy; those with $x > x_A$ are saddle points. In order to express this result in terms of physical variables, we use the fact that the density contrast $\rho_e(r)/\rho_c = n_0$ is a monotonic function of x_0 ; for $x = x_A$, $n(x_A)^{-1} = 709$. We can now state the final results:

(i) Systems with $RE/GM^2 < -0.335$ cannot evolve into isothermal spheres. Entropy has no extremum for such systems.

(ii) Systems with $RE/GM^2 > -0.335$ and $\rho_c > 709\rho_e(R)$ can exist in a metastable (saddle point state) isothermal sphere configuration. The entropy extrema exist but they are not local maxima.

(iii) Systems with $RE/GM^2 > -0.335$ and $\rho_c < 709\rho_e(R)$ can form isothermal spheres which are local maxima of entropy.

It would be interesting to see which configuration induces the instability [Padmanabhan 1989c]. Clearly $A[\delta\rho] = A[-\delta\rho]$. So we will be able to fix $\delta\rho$ only up to a sign (i.e., if $\delta\rho$ induces instability so does $-\delta\rho$). From our solution (4.53) we get

$$\begin{aligned}
 a(x)x^2 &= df/dx = c_1(2nx^2 - mnx) - \Lambda nx^2 \\
 &= \Lambda nx^2 \left(\frac{2 - m/x}{u_0 - 1} - 1 \right) = nx^2 \left(\frac{3 - m/x - u_0}{v_0 - 1} \right) = \frac{\Lambda nx^2}{u_0 - 1} \left((3 - u_0) - \frac{m}{x} \right).
 \end{aligned}
 \tag{4.64}$$

Thus the density contrast $\delta\rho/\rho = a/n$ is proportional to $(3 - u_0) - m/x$. The pattern of $\delta\rho/\rho$ will be determined by the number of times $\delta\rho$ changes sign, i.e., the number of times $\delta\rho$ becomes zero in the interval $0 < x < x_0$.

These zeros can be easily determined by the following geometrical construction, shown in fig. 4.6. Suppose x_0 is in the portion LM (stable solution). We draw a vertical line through x_0 cutting the line $u + v = 3$ (LM) at N. We then draw a horizontal line through N cutting the $u-v$ curve at x_1 . Then $a(x_1) = 0$. [This is obvious because the horizontal line is just $(3 - u_0) = v = (m/x)_{x_1}$.] Clearly for x_0 in the range LM there is only one solution; $\delta\rho$ will be like the one in the inset in fig. 4.6.

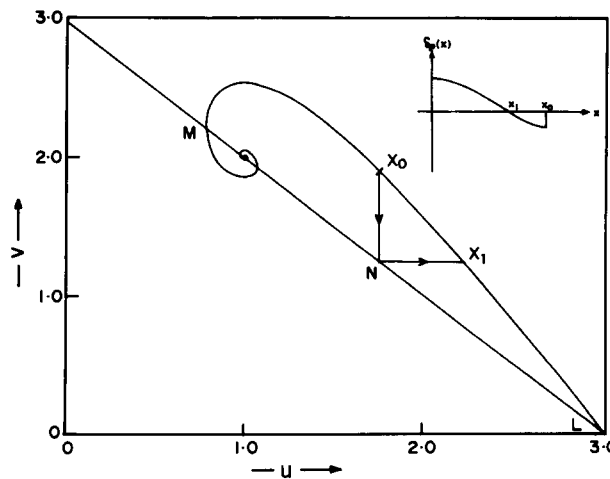


Fig. 4.6. Geometrical construction determining the nature of the perturbation mode. In this case, the mode does not induce any instability.

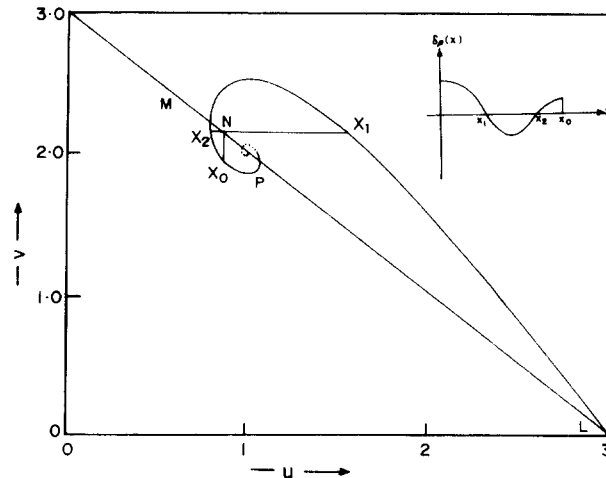


Fig. 4.7. Geometrical construction determining the nature of the perturbation mode. In this case, the mode has a core-halo structure and induces an instability.

Suppose x_0 is in the range MP of the curve (fig. 4.7). Then the same construction gives *two* points x_1 and x_2 at which $a(x) = 0$. This is, of course, the range at which our critical solution occurs. In other words, the mode which destabilises the isothermal sphere will have a pattern like the one in the inset in fig. 4.7. We may interpret this as having a “core-halo” structure.

4.5. Isothermal sphere with a short distance cutoff

The discussion in the last two sections shows that the isothermal solution exists only for a restricted range of parameters; even in this restricted domain, the solution may not be stable. It is therefore necessary to ask the question: How can we best describe systems for which the isothermal solution does not exist? Do they have some other equilibrium state?

It is clear that such an equilibrium state can exist if we take the finite size of the particles into account. In the above discussion, we have treated the constituents of the system as point particles and have assumed the validity of the $-r^{-1}$ potential all the way up to $r = 0$. In any realistic stellar system, this assumption breaks down at very short distances because of the finite size of the constituent particles. Our earlier study of the toy models showed that this feature may change the qualitative behaviour of the system.

Incorporating the finite size effects in an exact manner is a difficult task. However, an understanding of the effects due to the short distance cutoff can be obtained by a simplified analysis first performed by Aronson and Hansen [1972]. Their key idea is based on the following observation: The results obtained above are applicable to any system described by eqs. (4.11). In particular, they are applicable to a gaseous isothermal sphere, i.e. to a self-gravitating ideal gas kept at constant temperature, described by the equation of state $P = \rho kT$ [Bonner 1956, 1958]. This ideal gas equation ignores the finite size of the constituent particles. However, we can easily incorporate the effects due to the finite size of the constituent particles by modifying the equation of state to a Van der Waals type of equation,

$$P = \frac{\rho T}{1 - \rho/\rho_m}, \quad (4.65)$$

where ρ_m is a constant determined by the mass and size of the individual particles, which represents the maximum density permissible for the system. Such a self-gravitating, non-ideal gas in isothermal equilibrium will be described by the equation

$$\frac{1}{r^2} \frac{d}{dr} \left(\frac{r^2}{\rho} \frac{d}{dr} \frac{\rho T}{1 - \rho/\rho_m} \right) = -4\pi G\rho. \tag{4.66}$$

This equation replaces (4.11) in the study of effects due to a short range cutoff. It clearly reduces to (4.11) when ρ_m tends to infinity.

We will assume, as before, that the system is confined by a sphere of radius R and that the total mass of the system is M . The above equation can then be integrated to give a unique solution $\rho(r)$ once the parameters ρ_m , T , M , R are specified. In particular, the central density $\rho(0) = \rho_c$ is determined by these parameters.

Let us suppose that we fix the values of ρ_m , M and R but vary the temperature T . Aronson and Hansen found the following interesting feature in their numerical study of the system: As the temperature falls below a critical temperature T_c , the central density increases abruptly; so does the density contrast between the centre and the edge. In other words, the system develops a very compact, high density, core surrounded by a thin halo. This phase transition to a core-halo structure is extremely sharp.

This phenomenon can be understood by examining the nature of the solutions to (4.66) [Narasimha and Padmanabhan 1989]. We first convert (4.66) into dimensionless form by introducing the scalings

$$n = \rho/\rho_c, \quad t = k(r/R), \quad k = (bq)^{1/2}, \quad b = (4\pi GR^2\rho_m/T), \quad q = \rho_c/\rho_m, \tag{4.67}$$

in terms of which (4.66) becomes

$$\frac{1}{t^2} \frac{d}{dt} \left(\frac{t^2}{n} \frac{n'}{(1 - qn)^2} \right) = -n. \tag{4.68}$$

The boundary conditions for (4.68) are the following: $n(0) = 1$, $n'(0) = 0$. We can integrate (4.68) with these boundary conditions, treating q as a free parameter. This will allow us to determine the dimensionless density $n(t; q)$ and the dimensionless mass $m(t; q)$ contained within the radius t ,

$$m(t) = \int_0^t dx x^2 n(x), \tag{4.69}$$

These functions provide the complete solution to the problem if we are given the central density ρ_c and temperature T as free parameters. However, we are interested in the situation in which the total mass is specified and not the central density. Therefore, we have to determine the value of the central density, or, equivalently, the value of q , by imposing the constraint that the total mass inside the radius R is M . This condition can be written as

$$a = \frac{M}{4\pi\rho_m R^3} = \frac{1}{b} \frac{m(\sqrt{bq})}{\sqrt{bq}}. \tag{4.70}$$

Given the parameters of the system, a is fixed; this equation, therefore provides an implicit relationship between the (dimensionless) inverse temperature b and the (dimensionless) central density q . As we decrease the temperature, b increases. If there is a phase transition in the system, then we expect a rapid increase in q for a relatively small increase in b near a critical temperature. Or, equivalently, we expect b to be a slowly varying function of q near the phase transition. Thus a phase transition is signalled if we find that b is approximately constant for a range of values of q .

That such is indeed the case can be ascertained by the asymptotic analysis of the solution of (4.68). The function $m(x)/x^3$ starts with a value $\frac{1}{3}$ near $x = 0$ and decreases with increasing x . It can be easily seen that near $x = 0$,

$$m(x)/x \approx \frac{1}{3}x^2 - \frac{1}{30}(1-q)^2x^5, \quad (4.71)$$

while for somewhat larger x ,

$$m(x)/x \approx 2. \quad (4.72)$$

Using these asymptotic forms in (4.70) it is easy to determine the form of the function $b(q)$. We have b nearly equal to zero (corresponding to a very high temperatures) for $q \approx 3a$; near this value

$$b(q) \approx \frac{10}{9} \frac{(q-3a)}{a^2(1-3a)^2}. \quad (4.73)$$

As we increase q the value of b increases steadily and we can use the asymptotic form (4.72) in (4.70) to determine the functional form of $b(q)$. It is then easy to see from (4.72) that in this intermediate range, b is essentially a constant,

$$b(q) \approx 2/a. \quad (4.74)$$

This is precisely the region at which b is a very slowly varying function of q ; as we reasoned before this feature can cause a sharp phase transition making q approach rapidly the limiting value of unity. It is difficult to obtain good analytic approximations for $b(q)$ when q is close to unity but numerical study shows that b increases rapidly (and without bound) as q approaches unity.

The situation can be now described better using b as the independent variable: At very high temperatures ($b \approx 0$), $q \approx 3a$, which corresponds to a uniform distribution of matter inside the spherical volume. (In this case, the central density is the same as the uniform mean density $3M/4\pi R^3$, and the density contrast is zero.) As we decrease the temperature, b increases and so does q ; this corresponds to a steady increase in the central density and the density contrast, which is to be expected. However, when b approaches the critical value $2/a$, q starts to increase very rapidly with b ; that is, when the temperature drops below the critical temperature $T_c \approx GM/2R$, which corresponds to $b = 2/a$, the central density increases by a few orders of magnitude. The system will now have a very compact core and a large density contrast between the centre and the edge. The form of the function $q(b)$ is indicated in fig. 4.8. Figure 4.9 shows two density profiles at temperatures slightly above and below the critical temperature; the sudden emergence of the core-halo structure is obvious.

We saw earlier that the isothermal sphere without a short distance cutoff cannot exist at temperatures below a critical value. The introduction of the short distance cutoff allows the system to exist at any temperature. At low temperatures, it exists in the core-halo phase. The modification of the

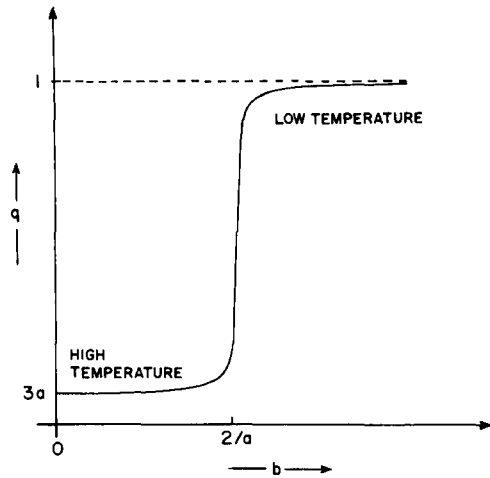


Fig. 4.8. Schematic diagram showing the behaviour of the function $q(b)$. Since $m(x)/x$ for the isothermal solution approaches 2 in an oscillatory manner, the vertical branch $b(q) \approx 2/a$ has small oscillations around the line shown.

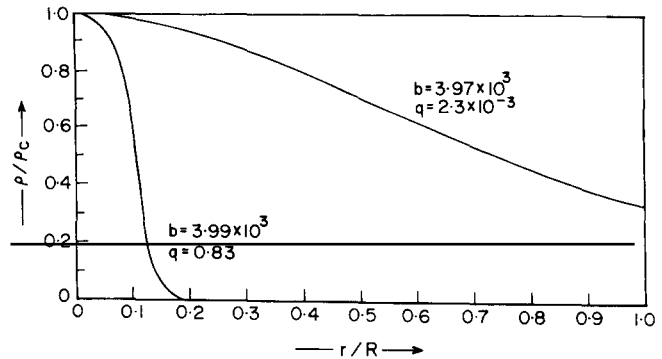


Fig. 4.9. The density of an isothermal sphere with a short distance cutoff as a function of the radius. For $b = 3.97 \times 10^3$, the density varies very little between the centre and the edge; when b is increased slightly to $b = 3.99 \times 10^3$, the density drops to very low values at the edge, indicating a sharp transition to a core-halo structure. The value of q increases from 2.3×10^{-3} to 0.83.

properties of the isothermal sphere due to the introduction of the short distance cutoff can be easily seen in the $u-v$ diagram. Figure 4.10 shows the behaviour of the $u-v$ curves as a function of the parameter a ; $a = 0$ corresponds to the usual isothermal sphere and as we increase the value of a the deviation from the isothermal sphere becomes more and more pronounced. Note that the nature of the curves is very different for $a = 0$ and for $a \neq 0$ irrespective of how small a is.

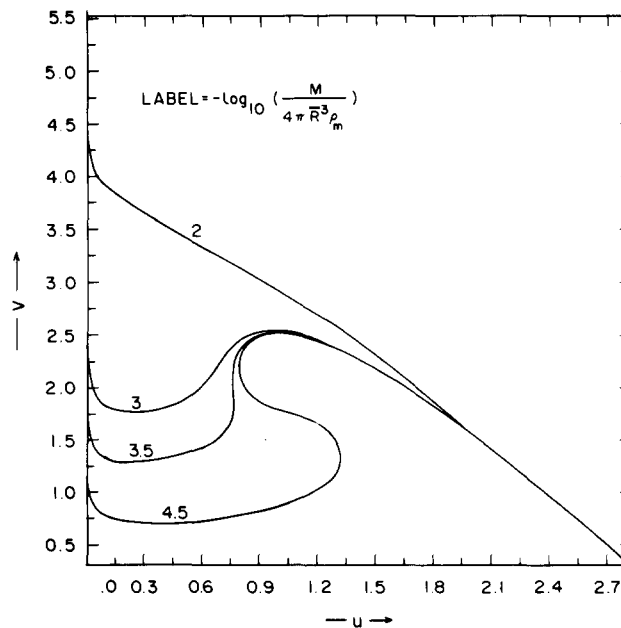


Fig. 4.10. The $u-v$ curve for the isothermal sphere with short distance cutoff; compare with fig. 4.1.

The above discussion was presented in the language of the canonical description (constant temperature, no constraint on energy) for ease of visualisation. Rigorously speaking, we must consider the system in the microcanonical description with constant energy. (In other words, we should determine the temperature parameter b in terms of E just as we determined the parameter q in terms of the mass M .) From the expression for the total energy

$$E = \frac{3}{2}MT - 4\pi G \int_0^R r dr \rho(r)M(r) \quad (4.75)$$

one can obtain the dimensionless combination

$$\frac{RE}{GM^2} = \frac{k}{2m_0^2} \int_0^k (3nt^2 - 2mnt) dt. \quad (4.76)$$

Eliminating k between this expression and the relation $RT/GM^2 = [m(k)/k]^{-1}$, one can obtain the $T(E)$ curve. This is shown in fig. 4.11 for various values of the cutoff parameter a . Notice that we have now obtained two regions of positive specific heat: a high temperature region along AB and a low temperature phase along CD. They are connected by a region of negative specific heat along BC. Comparing with fig. 4.4 – which describes the $T(E)$ curve in the absence of a short distance cutoff – we realise that the cutoff is crucial to stabilise the system by introducing the low temperature phase. It is also clear that the phase transition in the canonical description is replaced by a region of negative specific heat in the microcanonical description.

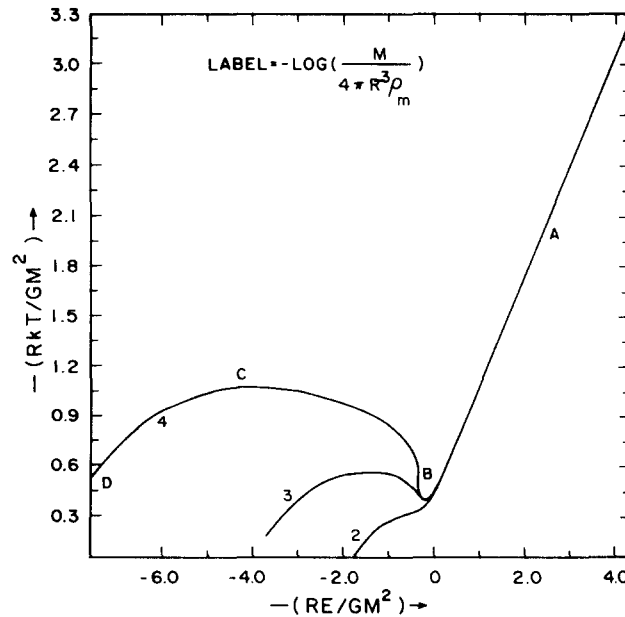


Fig. 4.11. The $T(E)$ curve for an isothermal sphere with short distance cutoff. Notice the similarity with the $T(E)$ curve for the toy model shown in fig. 3.1. The cutoff has changed the nature of the $T(E)$ curve from that shown in fig. 4.4.

4.6. Mean field thermodynamics and comparison with toy models

In section 4.4 we obtained the equations describing the isothermal sphere by approximating the exact expression for the phase volume. This approximation involves replacing a sum of terms in (4.3) by the *largest* one. Clearly, the approximation breaks down unless the term which is retained is the local *maximum*.

Such a breakdown occurs for all systems with $\lambda < \lambda_c$; in this range there is not even an extremum solution to the variational problem, let alone a maximum solution. Even when $\lambda > \lambda_c$, we do not obtain a local maximum if $n_0^{-1} > 709$. In either of these cases, the mean field approximation fails and we have to tackle the exact expression for phase volume.

The exact expression (4.1) is divergent for point particles but can be made finite by introducing a short distance cutoff. (Such a cutoff is anyway needed to give meaning to the entire exercise.) Configurations with a closely packed core will then contribute significantly to (4.1). If no other local maximum exists, then the equilibrium configuration will be one with such a compact core.

When $\lambda > \lambda_c$ and $n_0^{-1} < 709$, the mean field analysis remains valid. (In fact this is the *only* region in which it is valid; the portion of the $u-v$ curve beyond the point of stability is not physically relevant.) However, this validity rests on the tacit assumption that (4.1) is made finite by a suitable short distance cutoff. Thus in (4.3) two kinds of configurations make large contributions: the isothermal solution and the configurations in which part of the matter exists as a very compact core. One cannot decide a priori which of these two occupy the larger phase volume but it is likely that the latter one does. In other words, the local maximum of the entropy is not the global maximum. (This result is, of course, obvious in the limit of zero cutoff; and hence, it should be valid for sufficiently small cutoff as well.) In such a case, the isothermal solution will represent a valid metastable configuration though we expect the system to eventually slide back to the compact core configuration.

The thermodynamics of this system can now be easily analysed using the $T(E)$ curves of figs. 4.4 and 4.11. From our discussion above it is clear that the portion of the curve in fig. 4.4 beyond the point C, for which $x_A = 34.2$ and $n_0^{-1} = 709$, is irrelevant because the mean field approximation breaks down there. In fact, this region gets modified when the short distance cutoff is introduced, as should be clear from fig. 4.11. These curves in fig. 4.11 are very similar to what we have seen in the toy models: The short distance cutoff introduces a low energy phase with positive specific heat; the two regions of positive specific heat are connected by a branch with negative specific heat. The characteristic energies for the phases are set by the cutoff sizes: R and a short distance cutoff a . Most astrophysical systems will be found with energies in the middle range, with negative specific heat.

We found – in our study of the toy models – that the negative specific heat region of the microcanonical description is replaced by a phase transition in the canonical description. The same feature arises in the model studied in the last section. If we start with the high temperature, high energy phase of the system and reduce the temperature it will soon reach a point at which the $T-E$ relation becomes multivalued; i.e., we have different E for the same T . At this moment, the system releases the excess energy as latent heat and makes a transition to the low temperature phase. This transition is extremely sharp.

The mean field thermodynamics therefore confirms our earlier claim that the behaviour of self-gravitating systems is analogous to the behaviour of laboratory systems undergoing phase transitions. The universal behaviour of the $T(E)$ curve for gravitating systems is reminiscent of the universality of behaviour seen in normal phase transitions.

Lastly, we comment on the implications of the above result for the dynamical evolution of the

systems. As we mentioned earlier, the equations describing the isothermal sphere can also be derived based on certain intuitive considerations related to the dynamical equation (4.14) governing the evolution. In this approach, the following question becomes relevant: How does the evolution of the system depend on the value of λ ? For the sake of definiteness we will assume that the $C(f)$ in (4.14) is given by the Fokker–Planck term. (This equation will be discussed in detail in chapter 5.) Though analytic solutions are not available, the following qualitative description can be given. (For the analysis of the dynamical evolution see Inagaki [1980], Lynden-Bell and Eggeleton [1980], Lynden-Bell and Inagaki [1983], Larson [1970 a,b], Hachisu et al. [1978], Cohn [1980]; also see the review by Elson et al. [1987] and the book by Spitzer [1987].)

Consider the case with $RE/GM^2 < 0.335$ first. Here the evolution proceeds without ever reaching an isothermal configuration. We expect a very inhomogeneous “core–halo” final state to be reached in a time scale governed by the collisional processes operating in the system. The two-body relaxation time t_R determined by $C(f)$ provides a naive estimate for this process. In reality, the relaxation proceeds much more slowly because the temperature gradient in the central power law zone of the star cluster is extremely slight.

Now consider the case for which $RE/GM^2 > -0.335$ and $\rho_c/\rho(R) > 709$. This situation is somewhat involved. In the first phase of the evolution, $C(f)$ can drive the system into the isothermal sphere state. Once it is reached, no first order perturbation can increase S . (This is in contrast to the previous case, in which a first order increase of S is possible for suitable perturbations.) What is more, the Fokker–Planck term $C(f)$ vanishes when f is a Maxwellian distribution. [This is a feature shared by several other reasonable $C(f)$'s.] Only a second order perturbation of a particular kind (discussed in the last section) can drive the system along the valley of the saddle point. It seems likely that the timescale will now increase and will be some $t_{1R} > t_R$. It must be emphasised that this argument presupposes the existence of a local description for the collisional processes. In reality, the collisional processes operating in star clusters are highly non-local and the estimate of the timescale t_{1R} is a far more difficult problem than indicated by the simple arguments presented above.

Lastly, consider the case for which $RE/GM^2 > -0.335$ but $\rho_c < 709\rho(R)$. The system will again reach the isothermal configuration in the first phase of evolution. But now even second order perturbations cannot increase S ; $C(f)$ is also zero because f is Maxwellian. Only a large random fluctuation can drive the system out of this local maximum of entropy. The timescale for this phenomenon could be exponentially large.

Appendix 4.1. The variational problem for the entropy

Consider first the class of all $f(\mathbf{x}, \mathbf{v})$ with a fixed E , M and $\rho(\mathbf{x})$. Fixing $\rho(\mathbf{x})$ fixes M , $\phi_c(\mathbf{x})$ and U . Since E is also fixed, $K = E - U$ is fixed. Thus extremising S subject to constant $\rho(\mathbf{x})$ and E is the same as extremising S with constant $\rho(\mathbf{x})$ and K . Introducing two lagrange multipliers $\lambda(\mathbf{x})$ and β , we have the variation

$$\begin{aligned} \delta S &= - \int d^3\mathbf{x} d^3\mathbf{v} [\delta(f \ln f) + \lambda(\mathbf{x}) \delta f + \frac{1}{2}\beta v^2 \delta f] \\ &= - \int d^3\mathbf{x} d^3\mathbf{v} \left((\ln ef) \delta f + \frac{(\delta f)^2}{2f} + \lambda \delta f + \frac{1}{2}\beta v^2 \delta f \right). \end{aligned} \quad (4.77)$$

Equating the coefficient of f to zero, we get

$$f = \exp\{-[\lambda(\mathbf{x}) + 1] - \frac{1}{2}\beta v^2\}. \quad (4.78)$$

The multipliers $\lambda(\mathbf{x})$ and β have to be fixed in terms of $\rho(\mathbf{x})$ and K . Doing this, we can write f as

$$f(\mathbf{x}, \mathbf{v}) = (2\pi T)^{-3/2} \rho(\mathbf{x}) \exp(-v^2/2T), \quad (4.79)$$

with $T = 2K/3M$. From the coefficient of $(\delta f)^2$, we see that this is a true maximum for S . The maximum value of S can be obtained by substituting (4.79) into the expression for the entropy. We find

$$S_{\max}(\rho(\mathbf{x}); K) = \frac{3}{2}M \ln T - \int d^3\mathbf{x} \rho \ln \rho + \text{constant}. \quad (4.80)$$

Equation (4.80) gives the maximum value of the entropy possible for given $\rho(\mathbf{x})$ and K ; or, equivalently, for a given $\rho(\mathbf{x})$ and E . This maximum is achieved when the velocity distribution is Maxwellian.

We next vary K and $\rho(\mathbf{x})$ keeping E and M fixed, so as to find the $\rho(\mathbf{x})$ for which (4.80) is extremum. From the constraint

$$0 = \delta E = \delta K + \frac{1}{2} \int d^3\mathbf{x} (\delta\rho \phi + \rho \delta\phi + \delta\rho \delta\phi) = \frac{3}{2}M \delta T + \int d^3\mathbf{x} (\phi \delta\rho + \frac{1}{2}\delta\rho \delta\phi) \quad (4.81)$$

we find that

$$\delta T = -\frac{2}{3M} \int d^3\mathbf{x} (\phi \delta\rho + \frac{1}{2}\delta\rho \delta\phi). \quad (4.82)$$

In arriving at (4.82) we have used the fact that $\int \phi \delta\rho d^3\mathbf{x} = \int \rho \delta\phi d^3\mathbf{x}$. From (4.80) we have

$$\delta S = \frac{3M}{2T} \delta T - \int d^3\mathbf{x} \delta\rho \ln \rho e - \frac{3M}{4T^2} (\delta T)^2 - \int d^3\mathbf{x} \frac{(\delta\rho)^2}{2\rho}; \quad (4.83)$$

substituting for δT from (4.82) and imposing the $\delta M = 0$ constraint by a lagrange multiplier α , we have the variation

$$\begin{aligned} \delta S + \alpha \delta M &= - \int d^3\mathbf{x} \delta\rho [\ln(\rho e) - \alpha + \phi/T] - \int d^3\mathbf{x} \left(\frac{\delta\rho \delta\phi}{2T} + \frac{(\delta\rho)^2}{2\rho} \right) \\ &\quad - \frac{3M}{4T^2} \left(-\frac{2}{3M} \int d^3\mathbf{x} (\phi \delta\rho + \frac{1}{2}\delta\rho \delta\phi) \right)^2 \\ &\equiv - \int d^3\mathbf{x} \delta\rho [\ln(\rho e) - \alpha + \phi/T] + \delta^2 S, \end{aligned}$$

where the second variation of S is

$$\begin{aligned}\delta^2 S &= - \int d^3 \mathbf{x} \left(\frac{\delta \rho \delta \phi}{2T} + \frac{(\delta \rho)^2}{2\rho} \right) - \frac{1}{3MT^2} \left(\int d^3 \mathbf{x} \phi \delta \rho \right)^2 + O(\delta \cdot)^3 \\ &= - \int d^3 \mathbf{x} \left(\frac{\delta \rho \delta \phi}{2T} + \frac{(\delta \rho)^2}{2\rho} \right) - \frac{1}{3MT^2} \left(\int d^3 \mathbf{x} \phi \delta \rho \right)^2.\end{aligned}\quad (4.84)$$

This is our basic result. The extremum is decided by setting the coefficient of $\delta \rho$ to zero; whether the extremum is a true maximum is decided by the sign of $\delta^2 S$.

Appendix 4.2. Stability of small isothermal spheres

We will prove here that for small R (i.e. for $R < R_1$ with some specified R_1), $A(a)$ is negative definite. To do this, we will show that for small R , the second and third terms on the right hand side of (4.34) are negative definite.

Consider the function

$$Q = M_0 c (n x^3 - m) = c m M_0 (u - 1), \quad (4.85)$$

which is the same as $F = M_0 f$ with $\Lambda = 0$. Therefore, Q satisfies the equation [see (4.45)]

$$\frac{d}{dr} \left(\frac{1}{\rho r^2} \frac{dQ}{dr} \right) Q = - \frac{4\pi G}{T r^2} Q. \quad (4.86)$$

We know from our previous analysis that $Q \propto m(u-1)$ behaves like $2m$ near $x=0$ ($u=3$), and becomes zero when $u=1$. The first zero (after the trivial $x=0$) occurs at $x = x_1 \cong 8.2$. Let us call this radius R_1 .

Thus for $R = R_1$, Q may be considered the eigenfunction of the eigenvalue equation

$$\frac{d}{dr} \left(\frac{1}{\rho r^2} \frac{dQ}{dr} \right) = -\varepsilon \frac{4\pi G}{T r^2} Q, \quad (4.87)$$

with eigenvalue $\varepsilon = 1$. Since Q has no modes it is the first eigenvalue. But the eigenvalue problem above corresponds to minimising the expression

$$H = \int_0^{R_1} \frac{1}{r^2 \rho} \left(\frac{dq}{dr} \right)^2 dr / \int_0^{R_1} \frac{4\pi G}{T} \frac{q^2}{r^2} dr. \quad (4.88)$$

Therefore we conclude: The minimum of H is $\varepsilon_{\min} = 1$. From (4.34) we get

$$A[a] = - \frac{1}{3MT^2} \left(\int_0^{R_1} dr \phi'_c q \right)^2 - \int_0^{R_1} dr \left((-1) \frac{\psi' q}{2T} + \frac{q'^2}{8\pi \rho r^2} \right)$$

$$\begin{aligned}
&= -\frac{1}{3MT^2} \left(\int_0^{R_1} dr \phi'_e q \right)^2 + \int_0^{R_1} dr \frac{1}{2} \left(\frac{Gq^2}{Tr^2} - \frac{(q')^2}{4\pi\rho r^2} \right) \\
&= -\frac{1}{3MT^2} \left(\int_0^{R_1} dr \phi'_e q \right)^2 - \frac{1}{8\pi} \int_0^{R_1} dr \left(\frac{1}{\rho r^2} q'^2 - \frac{4\pi G}{T} \frac{q^2}{r^2} \right) \\
&= -\frac{1}{3MT^2} \left(\int_0^{R_1} dr \phi'_e q \right)^2 - \frac{1}{8\pi} \int_0^{R_1} dr \frac{4\pi Gq^2}{Tr^2} (H-1). \tag{4.89}
\end{aligned}$$

Since $H > 1$, $A[a]$ is negative definite. Thus all isothermal spheres with $R < R_1$ are stable.

5. The approach to equilibrium: collisional relaxation

5.1. Introduction

The study of the equilibrium properties of gravitating systems presented in the last three sections rests on the following tacit assumption: If the system is set up at some initial instant in an arbitrary configuration, then it will approach a suitable equilibrium configuration of maximum entropy at sufficiently late times. The time taken by the system to approach the equilibrium configuration (within some reasonable accuracy) is usually called the relaxation time, t_R , for the system. To understand how the system approaches the equilibrium configuration we need to study the dynamical evolution of the system for $t < t_R$.

This is a consequently tougher task than the study of equilibrium properties and often requires a very different description. For systems dominated by gravity there are some new difficulties, not encountered in other systems: (1) In the non-gravitating systems, like neutral gases and plasmas with Debye shielding, the relaxation usually proceeds through ‘‘collisions’’. The concept of collisions is well defined for such systems with short range forces. As we shall see, this is not the case for gravitating systems. (2) It is known that self-gravitating systems relax through a process other than the collisional relaxation; a process usually called ‘‘violent relaxation’’ [Lynden-Bell 1967]. Since violent relaxation operates at a timescale much shorter than the collisional relaxation timescale, this could be the dominant process in several astrophysical systems. The dynamics of this process, unfortunately, is not yet well understood. (3) The collisional relaxation time for several astrophysical systems is larger than the age of the universe. This implies that, from a practical point of view, we need a detailed description of the evolution prior to the relaxation. In contrast, most of the properties of laboratory systems can be understood in terms of equilibrium or quasi-equilibrium configurations because of their short relaxation times.

In this chapter we will discuss the collisional relaxation of gravitating systems; violent relaxation will be taken up in chapter 6. The emphasis in both these chapters will be on conceptual issues and on questions which are still open.

5.2. The description of dynamical evolution

The *equilibrium* properties of the system can be directly related to the volume occupied by the system in phase space. To study the *approach* to this equilibrium configuration we need more detailed information. It is usual to provide such information in terms of the distribution functions in the phase space. In particular a one particle distribution function $f(\mathbf{x}, \mathbf{v}, t)$ can be defined by the relation

$$dm = f(\mathbf{x}, \mathbf{v}, t) d\mathbf{x} d\mathbf{v}, \quad (5.1)$$

where dm is the mass contained inside the phase cell $(\mathbf{x}, \mathbf{v}; \mathbf{x} + d\mathbf{x}, \mathbf{v} + d\mathbf{v})$. One can equivalently look upon $f(\mathbf{x}, \mathbf{v}, t)$ as proportional to the probability for finding a particle in the specified phase cell.

If no further restrictions are put on f , then it is possible to write down distribution functions which contain the *complete* information about the system. For example, consider the function

$$f_{\text{exact}}(\mathbf{x}, \mathbf{v}, t) = m \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i(t)) \delta(\mathbf{v} - \mathbf{v}_i(t)), \quad (5.2)$$

where $(\mathbf{x}(t), \mathbf{v}(t))$ is the actual phase space trajectory of the system obtained, say, by integrating the equations of motion. This function satisfies the equation

$$\frac{\partial f_{\text{ex}}}{\partial t} + \mathbf{v} \cdot \frac{\partial f_{\text{ex}}}{\partial \mathbf{x}} - \nabla \phi_{\text{ex}} \cdot \frac{\partial f_{\text{ex}}}{\partial \mathbf{v}} = 0, \quad (5.3)$$

where $\phi_{\text{ex}}(\mathbf{x}, t)$ is the exact gravitational potential of the system determined by the equation

$$\nabla^2 \phi_{\text{ex}}(\mathbf{x}, t) = 4\pi G \int f_{\text{ex}}(\mathbf{x}, \mathbf{v}, t) d\mathbf{v}. \quad (5.4)$$

The above description, while being technically correct, is useless for any practical purposes. To write down (5.2) we either need the exact trajectories or we need the solution to (5.3) and (5.4) for an arbitrary initial condition [corresponding to (5.2) evaluated at $t = 0$]. Either of these tasks is impossibly difficult for a realistic system.

If we have to make any progress using distribution functions, then it is necessary to disregard such an exact description in terms of delta functions and confine our attention to functions which are smooth. This approximation involves the following physical assumption about the system: It is possible to divide the phase space into a large number of cells in such a way that each cell is (a) large enough to contain a macroscopic number of particles but (b) small enough for all the particles in the cell to be assumed to possess the same average characteristics of the cell. If such an intermediate scale exists for the cell size, then we can define smooth distribution functions which are useful. In particular, such a distribution function will be smooth over distances larger than the mean interparticle distance $l \equiv n^{-1/3} \equiv (4\pi R^3/3N)^{1/3}$, where R is the size of the system containing N particles.

We can associate with such a smooth distribution function $f_{\text{sm}}(\mathbf{x}, \mathbf{v}, t)$ a smooth gravitational potential $\phi_{\text{sm}}(\mathbf{x}, t)$ through the equation

$$\nabla^2 \phi_{\text{sm}}(\mathbf{x}, t) = 4\pi G \int f_{\text{sm}}(\mathbf{x}, \mathbf{v}, t) d\mathbf{v}. \quad (5.5)$$

The central issue, of course, is the following: What is the equation satisfied by f_{sm} ? If we ignore the granularity of the system, then we know that f_{sm} and ϕ_{sm} are good approximations to f_{ex} and ϕ_{ex} ; in this limit, we expect f_{sm} to satisfy an equation similar to (5.3). It is, therefore, convenient to write the equation satisfied by f_{sm} in the following form:

$$\frac{\partial f_{sm}}{\partial t} + \mathbf{v} \cdot \frac{\partial f_{sm}}{\partial \mathbf{x}} - \nabla \phi_{sm} \cdot \frac{\partial f_{sm}}{\partial \mathbf{v}} = C(f). \quad (5.6)$$

The right hand side describes the effects of granularity in the system and it is conventional to call $C(f)$ the ‘‘collision’’ term. To study collisional relaxation we need to obtain the form of $C(f)$ in some more suitable approximation.

While the intuitive ideas involved in the above approach are certainly correct, it is not easy to translate them into a rigorous derivation of $C(f)$. In fact, we would expect the nature of $C(f)$ to depend very much on the manner in which f_{sm} is defined in terms of f_{ex} or, equivalently, on the detailed physical assumptions governing the length and time scales over which f_{ex} varies. For systems governed by short range forces with range λ , the quantity $C(f)$ can be derived by an iterative expansion in the ‘‘diluteness’’ factor (λ/l) (see, e.g., Lifshitz and Pitaevskii [1982], section 16); for systems interacting by long range forces this method cannot be carried over in a straightforward manner. An alternative (but related) method of derivation, which starts with the N -dimensional distribution function and truncates the resulting BBGKY hierarchy of equations, also fails because of lack of a clear separation of the various timescales. One may even question whether it is at all possible to obtain sensible equations for a smoothed out distribution f_{sm} because, in the collisionless limit, Hamiltonian evolution tends to convert any smooth distribution into a fractal-like structure.

Many of these issues are still unsettled and hence the discussion in this chapter necessarily is less definitive compared to earlier chapters. In order to highlight the controversies, it seems worthwhile to attack this problem from two very different angles. In the next three sections, we will discuss the derivation of $C(f)$ by analysing the physical processes which contribute to it from first principles. In section 5.6, we will discuss a formal derivation of $C(f)$ by actually computing it as due to the difference between f_{ex} and f_{sm} .

5.3. Relaxation time for ‘‘hard’’ and ‘‘soft’’ collisions

In the absence of the collision term $C(f)$, the evolution of f_{sm} is governed entirely by the smooth mean field ϕ_{sm} and the particles in the system follow orbits available in this potential. The collision term C affects this smooth evolution in two different ways: (1) The actual potential felt by any particle at any instant is ϕ_{ex} and not ϕ_{sm} . As the configuration changes in time we expect the potential felt by the particle to fluctuate around the mean potential. Comparing (5.3) and (5.6) we see that this fluctuation between the actual potential and the smoothed out potential contributes to $C(f)$. This random force induces a diffusion in velocity space. (2) Every once in a while two particles in the system will come very close to each other and scatter with a significant change in their momentum. Such effects, which clearly agree with our intuitive notion of ‘‘collisions’’, must also contribute to $C(f)$.

It is clear that the relaxation of the system will proceed essentially through the process which acts at the shorter timescale. For gaseous systems with short range interactions, the first process described above (usually called ‘‘soft collisions’’) does not make any contribution. Since the forces are of short range, the smooth mean field vanishes for such systems. The relaxation proceeds entirely through the

second process (“hard collisions”). Two gas molecules will be scattered with a significant change in their momentum if they are closer than the range λ of the molecular interaction. This gives the relaxation time for gaseous systems to be $t_R \approx (n\lambda^2 v)^{-1}$, where n is the number density and v is the mean velocity. The relaxation time for “hard” collisions of gravitating systems can be estimated in an almost identical manner. When two particles pass each other with an impact parameter b , the velocity change induced in the particles will be about

$$\Delta v \approx \frac{Gm}{b^2} \frac{2b}{v} = \frac{2Gm}{bv}. \quad (5.7)$$

Clearly “hard” collisions ($\Delta v \sim v$) occur when the impact parameter is less than

$$b_{\text{hard}} \approx R/N. \quad (5.8)$$

The rate of such collisions will be $\sim n v b_{\text{hard}}^2$, so that the relaxation time for hard collisions is

$$t_{R,\text{hard}} = \frac{1}{n v b_{\text{hard}}^2} \approx \frac{R^3}{N} \frac{1}{v} \frac{N^2}{R^2} = N \frac{R}{v}. \quad (5.9)$$

Over time scales $t < t_{R,\text{hard}}$, hard collisions *can* indeed relax a gravitating system. Since this process is similar to the one in molecular kinetics, it can be studied by a collisional term similar to the usual Boltzmann collision term.

It turns out, however, that the soft collisions can relax the gravitating system at a shorter timescale, thereby rendering the hard collisions described above a secondary effect. We shall see in the next section that the fluctuating force which acts on each particle is dominantly of magnitude $Gm^2 N^{2/3}/R^2$ and that the timescale governing the fluctuations is about $t_{\text{fluc}} \sim N^{-1/3} R/v$. Such a random force will give random “kicks” to the particle inducing the particle to perform a random walk in velocity space. (Each encounter lasts for about a time $t_{\text{fluc}} \sim N^{-1/3} R/v$, which is quite small compared to the dynamical timescale $t_{\text{dyn}} \approx R/v$.) The cumulative effect of all the random kicks which occur in a time interval Δt (chosen such that $t_{\text{fluc}} \ll \Delta t \ll t_{\text{dyn}}$) is to produce a mean square displacement in velocity space of the order of

$$\begin{aligned} \langle (\Delta v)^2 \rangle &= \{(\Delta v)^2 \text{ in a soft collision}\} \times \{\text{no of encounters in a time } \Delta t\} \\ &= \int \left(\frac{2Gm}{bv} \right)^2 f(v) 2\pi b db v dv \Delta t, \end{aligned} \quad (5.10)$$

where $f(v) dv$ is the number of particles in the range $(v, v + dv)$. The integration over v can be performed giving

$$\langle (\Delta v)^2 \rangle \approx G^2 m^2 \int_0^\infty \frac{f(v) dv}{v} \int \frac{db}{b} \Delta t = G^2 m^2 \frac{n}{\bar{v}} \Delta t \int \frac{db}{b}, \quad (5.11)$$

where we have defined a mean inverse velocity \bar{v} . The integral over b diverges for both small and large values of the impact parameter – a difficulty directly related to the long range nature of gravity.

Introducing two cutoffs b_{\max} and b_{\min} , this integral can be expressed as

$$\langle (\Delta v)^2 \rangle = G^2 m^2 \frac{n}{\bar{v}} \ln(b_{\max}/b_{\min}) \Delta t. \quad (5.12)$$

The value of b_{\min} and b_{\max} has to be decided by extraneous physical considerations, which are necessarily adhoc. [In a rigorous derivation of $C(f)$ this problem should not arise.] Since the assumption of ‘‘softness’’ breaks down at $b < b_{\text{hard}} \approx R/N$ it is physically reasonable to take $b_{\min} \approx R/N$. The choice of b_{\max} is plagued by several conceptual problems (see, e.g., Lee [1968], Davidsen and Ostriker [1968], Kandrup [1980], Prigogine and Severne [1966]). A study of the nature of the fluctuating forces, discussed in the next section, suggests that it is reasonable to take $b_{\max} \approx R/N^{1/3}$. There is, however, another school of thought, which maintains that it is more appropriate to take $b_{\max} \approx R/N$. We will comment on this issue at the end of section 5.4; right now we shall take $b_{\max} \approx R/N^{1/3}$. Then we get

$$\frac{(\Delta v)^2}{\bar{v}^2} \approx \frac{G^2 m^2}{\bar{v}^3} n \Delta t \ln\left(\frac{R/N^{1/3}}{R/N}\right) = \left(\frac{\ln N}{N}\right) \frac{\Delta t}{R/v} \left(\frac{GM}{Rv^2}\right)^2 \approx \frac{\Delta t}{R/v} \frac{\ln N}{N}. \quad (5.13)$$

The relaxation time for soft collisions can be estimated by calculating the time it takes for the mean square velocity $\langle (\Delta v)^2 \rangle$ to grow to a value $\sim v^2$. This takes a time

$$t_{\text{R,soft}} = \frac{N}{\ln N} \frac{R}{v}, \quad (5.14)$$

which is smaller by a factor $\ln N$ than $t_{\text{R,hard}}$ in (5.9). Thus the dominant collisional effect in gravitating systems is slow diffusion in velocity space. This is the process we should describe by the $C(f)$ term [Chandrasekhar 1943, 1944; Chandrasekhar and Von Neumann 1942].

The qualitative discussion presented above identifies three important length scales of the system: (i) R/N , which is the impact parameter for hard collisions; at distances smaller than this, the correlation of the particles is significant; (ii) $R/N^{(1/3)}$, which is the mean interparticle distance; this is the distance scale which contributes most to the fluctuating force. This is also the scale over which we have smoothed out the exact distribution function to obtain our f_{sm} ; (iii) R , the size of the system. Roughly speaking, the hard collisions dominate in the range $(0, R/N)$, the soft collisions operate in the range $(R/N, R/N^{1/3})$ and the mean, smoothed out potential operates in the range $(R/N^{1/3}, R)$.

5.4. *The nature of the fluctuating forces*

The tentative estimates made in the last section suggest that the crucial relaxation effect in gravitating systems arises due to the fluctuating force acting on the particles. The magnitude and timescale of this fluctuating force can be estimated by simple probability considerations [Chandrasekhar 1941].

Consider a system of N particles, located within a region of radius R . Let the mass of each particle be m and let the position of the i th particle be r_i . To understand the nature of the fluctuating forces it is best to concentrate on a particle located at the origin. Since the smoothed out mean force on this particle vanishes due to symmetry, any force acting on the particle will be due to a fluctuation away

from spherical symmetry. The instantaneous force felt by a particle located at the origin will be

$$\mathbf{F} = \sum_i \frac{Gm^2}{r_i^3} \mathbf{r}_i \equiv \sum_i \mathbf{F}_i. \quad (5.15)$$

We are interested in the probability $p(\mathbf{F}) d\mathbf{F}$ that this force \mathbf{F} has a particular value in the range $(\mathbf{F}, \mathbf{F} + d\mathbf{F})$ at any given instant. This $P(\mathbf{F})$ can be related to probability $p(\mathbf{r}_i)$ that the i th particle is at \mathbf{r}_i in a simple manner,

$$P(\mathbf{F}) d\mathbf{F} = d\mathbf{F} \int \prod_{i=1}^N p(\mathbf{r}_i) d\mathbf{r}_i \delta\left(\mathbf{F} - \sum_i \mathbf{F}_i(\mathbf{r}_i)\right). \quad (5.16)$$

[We have assumed that the positions of the particles are uncorrelated, so that $\prod_i p(\mathbf{r}_i)$ is the probability for the configuration $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ to occur.] The easiest way to evaluate (5.16) is to use the Fourier transform of $P(\mathbf{F})$,

$$\begin{aligned} \tilde{P}(\mathbf{k}) &\equiv \int P(\mathbf{F}) \exp(i\mathbf{k} \cdot \mathbf{F}) d\mathbf{F} \\ &= \int \left(\prod_{i=1}^N p(\mathbf{r}_i) d\mathbf{r}_i \right) \exp\left(i\mathbf{k} \cdot \sum \mathbf{F}_i\right) \\ &= \prod_{i=1}^N \int p(\mathbf{r}_i) e^{i\mathbf{k} \cdot \mathbf{F}_i} d\mathbf{r}_i = \left(\int p(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{F}(\mathbf{r})} d\mathbf{r} \right)^N. \end{aligned} \quad (5.17)$$

To proceed further we need an estimate for $p(\mathbf{r})$. We expect this quantity to be proportional to the mean density $\rho(\mathbf{r})$. The simplest description will correspond to a constant mean density implying that p is a constant for $|\mathbf{r}| < R$; i.e., we take

$$P(\mathbf{r}) = \text{constant} = 3/(4\pi R^3). \quad (5.18)$$

We are ultimately interested in the limit in which $R \rightarrow \infty$ and $N \rightarrow \infty$ keeping $n = 3N/4\pi R^3$ constant. With this limit in mind we write

$$\left(\int p e^{i\mathbf{k} \cdot \mathbf{F}} d\mathbf{r} \right)^N = \left(1 - \frac{3}{4\pi R^3} \int (1 - e^{i\mathbf{k} \cdot \mathbf{F}}) d\mathbf{r} \right)^{(4/3)\pi R^3 n} \quad (5.19)$$

As $R \rightarrow \infty$ this expression has the following limit:

$$p(\mathbf{k}) = \exp\left(-n \int (1 - e^{i\mathbf{k} \cdot \mathbf{F}}) d\mathbf{r}\right) \equiv \exp[-nC(\mathbf{k})]. \quad (5.20)$$

We can evaluate $C(\mathbf{k})$ by standard methods; we get

$$C(\mathbf{k}) = \frac{4}{15} (2\pi Gm^2)^{3/2} |\mathbf{k}|^{3/2} \equiv (F_0 k)^{3/2} / n, \quad (5.21)$$

where we have defined F_0 as

$$F_0 = 2\pi(4/15)^{2/3} \frac{Gm^2}{n^{-2/3}} \approx 2.6 \frac{Gm^2}{(n^{-1/3})^2}. \quad (5.22)$$

Note that F_0 is typically the force exerted by a particle located at the mean interparticle distance $n^{-1/3}$ from the origin, i.e., the force due to a typical “nearest neighbour” located at distance $R/N^{1/3}$. Using (5.21), (5.20) and (5.17) we get

$$P(F) = \int \frac{d\mathbf{k}}{(2\pi)^3} \exp[-i\mathbf{k} \cdot \mathbf{F} - (F_0 k)^{3/2}]. \quad (5.23)$$

This integral can be represented in a more convenient form after some transformations [Chandrasekhar 1941, 1943]. If we write

$$P(F) dF \equiv 4\pi P(F) F^2 dF \equiv W(F) dF, \quad F = |\mathbf{F}|, \quad (5.24)$$

then $W(F)$ can be expressed as

$$W(F) = F_0^{-1} H(F/F_0), \quad (5.25)$$

$$H(y) \equiv \frac{2}{\pi y} \int_0^\infty dx (x \sin x) \exp[-(x/y)^{3/2}]. \quad (5.26)$$

This function $H(y)$ has simple asymptotic forms,

$$h(y) \cong \begin{cases} 4y^2/3\pi, & \text{as } y \rightarrow 0, \\ (\frac{15}{8})(2/\pi)^{1/2} y^{-5/2}, & \text{as } y \rightarrow \infty. \end{cases} \quad (5.27)$$

Correspondingly, $W(F)$ has the asymptotic forms

$$W(F) \cong \begin{cases} (4/3\pi F_0^3) F^2 & \text{as } F \rightarrow 0, \\ (\frac{15}{8})(2/\pi)^{1/2} F_0^{3/2} F^{-5/2} & \text{as } F \rightarrow \infty. \end{cases} \quad (5.28)$$

The function $W(F)$ starts at zero; increases, reaches a maximum at $1.6F_0$ and dies down for large F . In other words, most of the force comes from the nearest neighbour located at the mean distance $R/N^{1/3}$.

The calculation of $W(F)$ was based on the assumption that the particles in the system are distributed in an uncorrelated manner. This assumption clearly breaks down at distances smaller than R/N corresponding to forces greater than $F_{\max} \sim Gm^2/(R/N)$. Thus one should not trust $W(F)$ for $F > F_{\max}$.

The above discussion demonstrates two important characteristics of F_{random} : (i) It originates because of the graininess in the system. Note that as $n \rightarrow \infty$, $F_0 \rightarrow \infty$ and $P(\mathbf{k})$ vanishes. (ii) Most of the contribution to the fluctuating force arises from within the range of distances $(R/N, R/N^{1/3})$. At distances smaller than R/N , we cannot treat particles as uncorrelated; for distances much larger than $R/N^{1/3}$, $P(F)$ becomes negligible.

This result can be generalised in different ways. For example, $P(F)$ can be computed in a similar manner even if the mean density is not a constant but falls as r^{-p} (with $p < 3$); the qualitative nature of the results is not changed [Kandrup 1980; Ahmed and Cohen 1973]. Most of the fluctuating force is still

contributed by the nearest neighbour. The result is not sensitive to the total number of particles in the system either; numerical evaluation shows that the result for $N = 50$ differs from the result for $N = \infty$ only by ten percent; for $N \sim 1000$ the agreement between the results is excellent.

The above facts suggest that the conclusions we arrived at regarding the nature of the fluctuating forces are not sensitive to the detailed assumptions made in the calculation. In particular, the results will be true even for a particle located away from the origin, as long as there are a reasonable number of particles (about 50 or so) surrounding it.

To have a complete picture of the fluctuating force, we also need to estimate the timescale over which the fluctuation lasts. That is, we need to know the ‘‘lifetime’’ of each configuration. This issue, however, is more complicated and still remains somewhat controversial. An exact calculation of the lifetimes of the fluctuations would require a computation in phase space, which is difficult to perform. But one can evaluate the probability $P(\mathbf{F}, d\mathbf{F}/dt)$ for the simultaneous occurrence of a force \mathbf{F} and the rate of change of the force ($d\mathbf{F}/dt$) by doing the calculation in configuration space itself. From this probability, one can define a timescale $T(\mathbf{F})$ by the relation

$$T(\mathbf{F}) = \frac{|\mathbf{F}|}{\sqrt{\langle |d\mathbf{F}/dt|^2 \rangle}}. \quad (5.29)$$

Detailed computation shows that $T(\mathbf{F})$ has the following simple asymptotic forms [Chandrasekhar and Von Neumann 1942]:

$$T(\mathbf{F}) = \begin{cases} t_{\text{fluc}}\beta, & \beta \ll 1, \\ \sqrt{15/8} t_{\text{fluc}}\beta^{-1/2}, & \beta \gg 1, \end{cases} \quad (5.30)$$

$$\beta = F/F_0, \quad t_{\text{fluc}} = \frac{1}{(30\pi^3)^{1/6}} \frac{1}{n^{1/3}v}. \quad (5.31)$$

[These asymptotic forms are very good approximations to the exact expression. If we use the first form for $\beta \leq (15/8)^{1/3}$ and the second form for $\beta \geq (15/8)^{1/3}$ then we will incur only a maximum error of about eight percent compared to the exact result.] We notice that the lifetime for both small and large fluctuations in the force is negligibly small. The longest time any particular configuration lasts is about t_{fluc} , which is about $R/N^{1/3}v$.

If this were to be the whole story, then we may conclude that the typical distance scale for dominant encounters is about $R/N^{1/3}$. Unfortunately, the timescale $T(\mathbf{F})$ computed above may not be the relevant timescale determining the effect of the encounters. The reason is as follows: Even though most of the power of the fluctuating force may be concentrated around the frequency $T(\mathbf{F})^{-1}$, the stars may not be able to absorb and respond to this high frequency power. Thus we should actually investigate the frequency band to which the stars *can* respond. (This may be thought of as a resonance effect in phase space.) Such an investigation, however, is much more complicated and the results, in general, depend on the specific system which is being studied. (For one example of such a study, see Binney and Lacey [1988].) The effects arising from frequency matching in phase space can easily make the timescale $T(\mathbf{F})$ discussed above irrelevant.

We argued in the last section that the fluctuating forces induce a random walk in velocity space leading to $(\Delta v)^2 \propto t$. It may seem that the above calculation justifies this assumption because $t_{\text{fluc}} \ll t_{\text{orbit}} = R/v$; we can indeed consider a time scale Δt such that $t_{\text{fluc}} \ll \Delta t \ll t_{\text{orbit}}$ and use the

diffusion approximation. This argument, however, is not quite correct because t_{fluc} , defined through (5.29), may not be the relevant quantity to characterise the random nature of the force. This difficulty is revealed by the following calculation: The velocity increment $\Delta \mathbf{v}$ acquired by a star in a time interval T is

$$\Delta \mathbf{v} = \int_0^T dt \mathbf{F}(t), \quad (5.32)$$

where $\mathbf{F}(t)$ is the random force acting on the star at time t . From (5.32), it is easy to obtain the relation

$$\langle (\Delta v)^2 \rangle = 2 \int_0^T dt \int_0^T dt' \langle \mathbf{F}(t) \cdot \mathbf{F}(t') \rangle. \quad (5.33)$$

If the correlation function $\langle \mathbf{F}(t) \cdot \mathbf{F}(t') \rangle$ depends only on the time difference $s = t' - t$, then (5.33) can be converted to the form

$$\langle (\Delta v)^2 \rangle = 2 \int_0^T ds (T - s) \langle \mathbf{F}(0) \cdot \mathbf{F}(s) \rangle. \quad (5.34)$$

Further progress depends on the s -dependence of $\langle \mathbf{F}(0) \cdot \mathbf{F}(s) \rangle$. If the correlation decreases faster than s^{-1} , then we can take the limit ($T \rightarrow \infty$) in the above expression and obtain the standard result

$$\langle (\Delta v)^2 \rangle = 2T \int_0^\infty ds \langle \mathbf{F}(0) \cdot \mathbf{F}(s) \rangle \propto T. \quad (5.35)$$

Thus, in order to verify that our diffusion approximation is valid, we should actually compute $\langle \mathbf{F}(0) \cdot \mathbf{F}(s) \rangle$ rather than $T(F)$. Exact evaluation of this function is difficult. It can be evaluated if we make the approximation that each star follows a linear trajectory in the time interval $(0, s)$ [Lee 1968]. Then we get

$$\langle \mathbf{F}(0) \cdot \mathbf{F}(s) \rangle = \int d\mathbf{v} f(\mathbf{v}) \int d\mathbf{r} \frac{Gm^2 \mathbf{r}}{r^3} \cdot \frac{Gm^2(\mathbf{r} + \mathbf{v}s)}{|\mathbf{r} + \mathbf{v}s|^3}. \quad (5.36)$$

If the \mathbf{r} -integration in (5.36) is confined to a bounded region of space then $\langle \mathbf{F}(0) \cdot \mathbf{F}(s) \rangle$ will behave as s^{-2} for large s and we are justified in using the diffusion result. But if we integrate over the whole space then we get an s^{-1} behaviour. Writing

$$\int d\mathbf{r} \frac{\mathbf{r}}{r^3} \cdot \frac{(\mathbf{r} + \mathbf{v}s)}{|\mathbf{r} + \mathbf{v}s|^3} = - \int d\mathbf{r} \nabla \cdot \left(\frac{\mathbf{r}}{r^3} \frac{1}{|\mathbf{r} + \mathbf{v}s|} \right) + \int d\mathbf{r} \frac{1}{|\mathbf{r} + \mathbf{v}s|} \nabla \cdot \left(\frac{\mathbf{r}}{r^3} \right), \quad (5.37)$$

it is easy to show that

$$\langle \mathbf{F}(0) \cdot \mathbf{F}(s) \rangle = \frac{4\pi G^2 m^2}{\langle v \rangle} \frac{1}{s}. \quad (5.38)$$

In this case, we cannot obtain (5.35) from (5.34) and the diffusion approximation fails.

The fact that the force correlations die down too slowly for (5.34) to converge suggests that our arguments based on the time of fluctuations are rather naive. This fact has led to the point of view that the validity of the diffusion approximation rests on the \mathbf{r} -integral in (5.36) being restricted to the size of the system. It is easy to see that this corresponds to the choice $b_{\max} = R$ in the earlier calculation.

In a rigorous derivation of the collisional equation this difficulty should disappear. We shall discuss this point again in section 5.6; right now we will proceed with the derivation of $C(f)$ based on the diffusion approximation.

5.5. Fokker–Planck description of the collisional evolution

The discussion in the last two sections suggests that the effect of soft collisions may be described by a diffusion process in velocity – or, equivalently, in momentum – space. Such a diffusion process is best described by a (diffusion) current $\mathbf{J}(\mathbf{p})$ in momentum space. This current characterises the flow of particles from one region of momentum space to another. The evolution will then be described by the equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} - \nabla \phi_{\text{mean}} \cdot \frac{\partial f}{\partial \mathbf{v}} = C(f) = -\nabla_{\mathbf{p}} \cdot \mathbf{J}(\mathbf{p}) = -\frac{\partial J^{\alpha}}{\partial p^{\alpha}}. \quad (5.39)$$

In a pure diffusion process, the current will be proportional to the concentration gradient; in the present case, we will therefore expect a term in J^{α} , which is proportional to, say, $\partial f / \partial v_{\alpha}$. However, it is easy to see that J^{α} should also contain some kind of damping term, proportional to f . In the absence of such a damping mechanism the diffusion term will lead to a monotonic increase in mean square velocity of *all* the particles, which is clearly unphysical. The second term, which leads to the damping of the velocities is usually called the “dynamical friction” term.

The relative role of these two terms and the nature of the evolution predicted by the Fokker–Planck equation can be understood from a simplified version of (5.39). This simple model possesses an exact mathematical solution and thus illustrates clearly the effect of the $\partial J^{\alpha} / \partial p^{\alpha}$ term. Consider the equation

$$\partial f(v, t) / \partial t = (\partial / \partial v)[(\alpha v) f + \frac{1}{2} \sigma^2 \partial f / \partial v] \equiv -\partial J / \partial v, \quad (5.40)$$

in which the form chosen for J has two terms; the second term $\frac{1}{2} \sigma^2 \partial f / \partial v$ has the standard form of a “diffusion current” proportional to the gradient. As time goes on this term will cause the mean space velocities of particles to increase in proportion to t , inducing the “random walk” in velocity space. Under the effect of this term, *all* the particles in the system will have their $\langle v^2 \rangle$ increasing without bound. This unphysical situation is avoided by the presence of the first term $(\alpha v f)$ in J . This term acts as a friction term (“dynamical friction”). The combined effect of the two terms is to drive f to a Maxwellian distribution with $\beta = (kT)^{-1} = \alpha / \sigma^2$. In such a Maxwellian distribution the increase in Δv^2 due to diffusion is exactly balanced by the losses due to dynamical friction. When two particles scatter, one gains the energy lost by the other; the one which has lost energy has undergone “dynamical friction” while the one which gained energy has undergone “diffusion” to higher v^2 . The cumulative effect of such phenomena is described by the two terms in $J(v)$. We shall see below that these two terms arise *together* when $J(v)$ is derived from first principles.

The above points can be easily illustrated by solving (5.40). Suppose we take a initial distribution

$f(v, 0) = \delta(v - v_0)$ peaked at a velocity v_0 . The solution of (5.40) with this initial condition is

$$f(v, t) = \left(\frac{\alpha}{\pi\sigma^2(1 - e^{-2\alpha t})} \right)^{1/2} \exp\left(-\frac{\alpha(v - v_0 e^{-\alpha t})^2}{\sigma^2(1 - e^{-2\alpha t})} \right), \quad (5.41)$$

which is a Gaussian with the mean

$$\langle v \rangle = v_0 e^{-\alpha t} \quad (5.42)$$

and dispersion

$$\langle v^2 \rangle - \langle v \rangle^2 = (\sigma^2/\alpha)(1 - e^{-2\alpha t}). \quad (5.43)$$

At late times ($t \rightarrow \infty$), the mean velocity $\langle v \rangle$ goes to zero while the velocity dispersion becomes σ^2/α . Thus the equilibrium configuration is a Maxwellian distribution of velocities with this particular dispersion, for which $J = 0$. To see the effect of the two terms individually on the initial distribution $f(v, 0) = \delta(v - v_0)$, we can set α or σ to zero. When $\alpha = 0$, we get pure diffusion,

$$f_{\alpha=0}(v, t) = \left(\frac{1}{2\pi\sigma^2 t} \right)^{1/2} \exp\left(-\frac{(v - v_0)^2}{2\sigma^2 t} \right). \quad (5.44)$$

Nothing happens to the steady velocity v_0 ; but the velocity dispersion increases in proportion to t representing a random walk in velocity space. On the other hand, if we set $\sigma = 0$, then we get

$$f_{\sigma=0}(v, t) = \delta(v - v_0 e^{-\alpha t}). \quad (5.45)$$

Now there is no spreading in velocity space (no diffusion); instead the friction steadily decreases $\langle v \rangle$.

We shall now take up the rigorous computation of $J(v)$ and show that it has two terms – one proportional to $\partial f/\partial v$ representing diffusion and the other proportional to f representing dynamical friction [Lifshitz and Pitaevskii 1982; Padmanabhan 1988].

To do this we have to essentially compute the diffusive flux current $J^\alpha(p^\alpha)$ in momentum space due to soft collisions. We shall first parametrise the basic collision between two particles in a convenient manner. Consider a collision in which two particles with momenta p_1 and p_2 scatter to momenta p'_1 and p'_2 . This event can be characterised by the set of vectors (p_1, p_2, p'_1, p'_2) . However, since scatterings conserve momenta we can do away with one variable by introducing the momentum transfer q and use the set $(p_1, p_2; p_1 + q, p_2 - q)$. It turns out, however, that a more convenient representation is in terms of the average momentum \bar{p} of each particle $\bar{p}_1 \equiv p_1 + \frac{1}{2}q$, $\bar{p}_2 \equiv p_2 - \frac{1}{2}q$, and the momentum transfer q . In terms of these variables we introduce a transition rate $W(\bar{p}_1, \bar{p}_2; q)$, which represents the rate of transition from $(\bar{p}_1 - \frac{1}{2}q, \bar{p}_2 + \frac{1}{2}q)$ to the state $(\bar{p}_1 + \frac{1}{2}q, \bar{p}_2 - \frac{1}{2}q)$. This parametrization has the following advantage: changing q to $-q$ interchanges the initial and final states. From the time reversibility of the scattering we immediately see that

$$W(\bar{p}_1, \bar{p}_2; q) = W(\bar{p}_1, \bar{p}_2; -q). \quad (5.46)$$

More precisely, we will define W by the following relation: The quantity $W(p + \frac{1}{2}q, p' - \frac{1}{2}q; q)f(p)f(p') dp' dq$ will represent the number of collisions which take place in unit time interval

between some particle of momentum \mathbf{p} and particles with momentum in the range $(\mathbf{p}', \mathbf{p}' + d\mathbf{p}')$, such that the momentum transfer is in the range $(\mathbf{q}, \mathbf{q} + d\mathbf{q})$. We will now relate \mathbf{J} to \mathbf{W} and f and then compute \mathbf{W} from the basic process of inverse square law scattering.

Consider an infinitesimal area element centred at p_{1x} and perpendicular to the p_x direction in momentum space. The flux of particles from left to right (i.e., the number of particles per unit area which cross this surface from left to right, in a unit of time) is given by

$$I_L = \int_{q_x > 0} d\mathbf{q} \int_{p_{1x} - q_x}^{p_{1x}} dk_x \int_{\text{all}} d\mathbf{p}_2 W(\mathbf{k} + \frac{1}{2}\mathbf{q}, \mathbf{p}_2 - \frac{1}{2}\mathbf{q}) f(\mathbf{p}_2) f(\mathbf{k}). \quad (5.47)$$

The integrand describes the scattering from $(\mathbf{k}, \mathbf{p}_2)$ to $(\mathbf{k} + \mathbf{q}, \mathbf{p}_2 - \mathbf{q})$ in which the particle with momentum \mathbf{k} moves across the surface. To achieve this with a given momentum transfer \mathbf{q} , the value of k_x must be in the range $(p_{1x} - q_x, p_{1x})$. We integrate over the momentum transfer \mathbf{q} making sure it is a flow from left to right (i.e., $q_x > 0$).

By a similar reasoning, the flux of particles moving from right to left across this surface is given by

$$I_R = \int_{q_x < 0} d\mathbf{q} \int_{p_{1x}}^{p_{1x} - q_x} \int_{\text{all}} d\mathbf{p} W(\mathbf{k} + \frac{1}{2}\mathbf{q}, \mathbf{p}_2 - \frac{1}{2}\mathbf{q}; \mathbf{q}) f(\mathbf{p}_2) f(\mathbf{k}). \quad (5.48)$$

The integrand is the same as in (5.47) but the range is different. We are interested in the net flux $J_x \equiv I_L - I_R$. By a series of simple transformations, we can make the domain of integration in both I_R and I_L the same and express J_x as

$$J_x = \int_{\text{all}} dl \int_{q_x > 0} d\mathbf{q} \int_{p_{1x} - q_x}^{p_{1x}} dl_x W(l + \frac{1}{2}\mathbf{q}, l' - \frac{1}{2}\mathbf{q}, \mathbf{q}) [f(l)f(l') - f(l + \mathbf{q})f(l - \mathbf{q})]. \quad (5.49)$$

To proceed further we will make the assumption that the collisions are ‘‘soft’’ so that most of the contribution to this integral comes from processes with small momentum transfer, i.e., from small \mathbf{q} . In this case, we can (i) replace W in (5.49) by $W(l, l', \mathbf{q})$, (ii) Taylor expand $f(l + \mathbf{q})f(l' - \mathbf{q})$ in \mathbf{q} retaining only up to linear terms in \mathbf{q} , and (iii) replace the integral over l_x by the multiplication by the range of integration, q_x . This brings \mathbf{J} to the form

$$J_\alpha = \frac{1}{2} \int dl' \int d\mathbf{q} q_\alpha q_\beta W(f \partial f' / \partial l'_\beta - f' \partial f / \partial l_\beta) = \int dl B_{\alpha\beta}(l, l') (f \partial f' / \partial l'_\beta - f' \partial f / \partial l_\beta), \quad (5.50)$$

where we have defined

$$B_{\alpha,\beta}(l, l') = \frac{1}{2} \int q_\alpha q_\beta W(l, l'; \mathbf{q}) d\mathbf{q}. \quad (5.51)$$

This quantity can be easily evaluated. Note that $B_{\alpha\beta}$ can only depend on $\mathbf{k} \equiv l - l'$; further, for soft collisions $B_{\alpha\beta}$ must be ‘‘transverse’’ to k^α : $B_{\alpha\beta} k^\alpha = 0$. The most general second rank symmetric tensor,

constructible from k_α and transverse to it must have the form

$$B_{\alpha\beta} = \frac{1}{2}B(\delta_{\alpha\beta} - k_\alpha k_\beta/k^2), \quad k = |\mathbf{k}|. \quad (5.52)$$

We, therefore, only need to compute

$$B = B_\alpha^\alpha = \frac{1}{2} \int q^2 W(\mathbf{l}, \mathbf{l}'; \mathbf{q}) d\mathbf{q}. \quad (5.53)$$

Since W represents the rate of transitions and \mathbf{q} is the momentum transfer per soft collision, the integral in (5.53) is essentially the same as the one in (5.10). Therefore

$$B = \frac{1}{2} \int q^2 W d\mathbf{q} = \frac{1}{2} m \int \left(\frac{2Gm^2}{bk} \right)^2 k 2\pi b db = 4\pi G^2 m^5 L/k, \quad (5.54)$$

where

$$L = \int \frac{db}{b} = \ln(b_{\max}/b_{\min}) = \ln(N^{2/3}) = \frac{2}{3} \ln N \quad (5.55)$$

is the Coulomb logarithm encountered before. We will write B in momentum space as

$$B = B_0/|\mathbf{k}|, \quad B_0 = 4\pi G^2 m^5 L. \quad (5.56)$$

Substituting (5.53), (5.52) into (5.50) we get

$$J_\alpha = \frac{1}{2} B_0 \int d\mathbf{l}' (f \partial f' / \partial l'_\beta - f' \partial f / \partial l_\beta) (\delta_{\alpha\beta} / k - k_\alpha k_\beta / k^3), \quad (5.57)$$

where \mathbf{k} is the momentum change ($\mathbf{l} - \mathbf{l}'$). This form clearly shows that for any Maxwellian distribution [with $f(\mathbf{l}) \propto \exp(-\sigma l^2)$] for which

$$f \partial f' / \partial l'_\beta - f' \partial f / \partial l_\beta = f(\mathbf{l})(-2\sigma l'_\beta) f(\mathbf{l}') - f(\mathbf{l}')(-2\sigma l_\beta) f(\mathbf{l}) = 2\sigma f(\mathbf{l}) f(\mathbf{l}') (\mathbf{l} - \mathbf{l}')_\beta \propto k_\beta, \quad (5.58)$$

J_α vanishes because of the relation $k_\beta B^{\alpha\beta} = 0$. Thus our collision term does admit the Maxwellian distribution as an equilibrium solution.

It is also clear that J_α will have one term proportional to f and one term proportional to $\partial f / \partial l_\beta$. There is a way of rewriting (5.57) which brings out the diffusion and dynamical friction terms clearly. Noticing that

$$\frac{\delta_{\alpha\beta}}{k} - \frac{k_\alpha k_\beta}{k^3} = \frac{\partial^2 |\mathbf{k}|}{\partial k_\alpha \partial k_\beta} = \frac{\partial^2 k}{\partial k_\alpha \partial k_\beta}, \quad (5.59)$$

we can rewrite (5.57) as

$$J_\alpha = \frac{1}{2} B_0 f(\mathbf{l}) \int d\mathbf{l}' \frac{\partial f}{\partial l'_\beta} \frac{\partial^2 k}{\partial k_\alpha \partial k_\beta} - \frac{1}{2} B_0 \frac{\partial f}{\partial l_\beta} \int d\mathbf{l}' \frac{\partial^2 k}{\partial k_\alpha \partial k_\beta}. \quad (5.60)$$

Using

$$\partial/\partial k_\alpha = \partial/\partial l_\alpha = -\partial/\partial l'_\alpha, \quad (5.61)$$

we can rewrite this expression in the form

$$J_\alpha = \frac{1}{2} B_0 f \int d\mathbf{l}' f'_\beta \frac{\partial^2 k}{\partial k_\alpha \partial k_\beta} - \frac{1}{2} B_0 f_\beta \frac{\partial^2 \psi}{\partial k_\alpha \partial k_\beta}, \quad (5.62)$$

$$\psi(\mathbf{l}) = \int d\mathbf{l}' f(\mathbf{l}') |\mathbf{l} - \mathbf{l}'|, \quad f'_\beta \equiv \partial f(\mathbf{l}') / \partial l'_\beta. \quad (5.63)$$

The two terms in (5.62) can be transformed into a more conventional form by integrating them by parts and using (5.61) repeatedly,

$$\begin{aligned} \int d\mathbf{l}' \frac{\partial f}{\partial l'_\beta} \frac{\partial^2 k}{\partial k_\alpha \partial k_\beta} &= - \int d\mathbf{l}' f(\mathbf{l}') \frac{\partial}{\partial l'_\beta} \frac{\partial^2 k}{\partial k_\alpha \partial k_\beta} = + \int d\mathbf{l}' f(\mathbf{l}') \frac{\partial}{\partial k_\beta} \frac{\partial^2 k}{\partial k_\alpha \partial k_\beta} \\ &= \int d\mathbf{l}' f(\mathbf{l}') \frac{\partial}{\partial k_\alpha} \frac{\partial^2 k}{\partial k_\beta \partial k_\beta} = \frac{\partial}{\partial l_\alpha} \frac{\partial}{\partial l^\beta \partial l_\beta} \int d\mathbf{l}' f(\mathbf{l}') |\mathbf{l} - \mathbf{l}'| = \frac{\partial}{\partial l_\alpha} \nabla_i^2 \psi(\mathbf{l}), \end{aligned} \quad (5.64)$$

$$f_\beta \frac{\partial^2 \psi}{\partial k_\alpha \partial k_\beta} = \frac{\partial}{\partial k_\beta} \left(f \frac{\partial^2 \psi}{\partial k_\alpha \partial k_\beta} \right) - f \frac{\partial}{\partial k_\beta} \frac{\partial^2 \psi}{\partial k_\alpha \partial k_\beta} = \frac{\partial}{\partial k_\beta} \left(f \frac{\partial^2 \psi}{\partial k_\alpha \partial k_\beta} \right) - f \frac{\partial}{\partial k_\alpha} \nabla^2 \psi, \quad (5.65)$$

where ∇^2 stands for the Laplacian in *momentum* space ($\partial^2/\partial k^\alpha \partial k_\alpha$). Defining

$$\eta(\mathbf{l}) \equiv \nabla^2 \psi(\mathbf{l}) = \frac{\partial^2}{\partial l^\alpha \partial l_\alpha} \int d\mathbf{l}' f(\mathbf{l}') |\mathbf{l} - \mathbf{l}'| = 2 \int d\mathbf{l}' \frac{f(\mathbf{l}')}{|\mathbf{l} - \mathbf{l}'|}, \quad (5.66)$$

and substituting (5.64) and (5.65) in (5.62) we get

$$\begin{aligned} J_\alpha(\mathbf{l}) &= \frac{1}{2} B_0 f(\mathbf{l}) \frac{\partial \eta}{\partial l_\alpha} - \frac{1}{2} B_0 \left[\frac{\partial}{\partial l_\beta} \left(f \frac{\partial^2 \psi}{\partial l_\alpha \partial l_\beta} \right) - f \frac{\partial \eta}{\partial k_\alpha} \right] \\ &= B_0 f(\mathbf{l}) \frac{\partial \eta}{\partial l_\alpha} - \frac{1}{2} B_0 \frac{\partial}{\partial l_\beta} \left(f \frac{\partial^2 \psi}{\partial l_\alpha \partial l_\beta} \right) \equiv a_\alpha(\mathbf{l}) f(\mathbf{l}) - \frac{1}{2} \frac{\partial}{\partial l_\beta} (\sigma_{\alpha\beta}^2 f), \end{aligned} \quad (5.67)$$

where

$$\nabla^2 \psi = \eta, \quad (5.68)$$

$$\nabla_i^2 \eta(\mathbf{l}) = \nabla_i^2 \left(2 \int d\mathbf{l}' \frac{f(\mathbf{l}')}{|\mathbf{l} - \mathbf{l}'|} \right) = -8\pi f(\mathbf{l}). \quad (5.69)$$

This brings J into the form discussed earlier. We can identify $a_\alpha(l)$ with the process of dynamical friction and $\sigma_{\alpha\beta}^2$ with diffusion [Rosenbluth et al. 1957].

The “potentials” η and ψ have very interesting properties. We note that (5.69) is essentially Coulomb’s law in velocity space, with $f(l)$ as source. Suppose that velocity distribution is spherically symmetric, i.e., $f(l) = f(|l|)$. We then know that the “force” ($\partial\eta/\partial l_\alpha$ at any given l is only due to the particles inside the velocity sphere of radius $|l|$. In other words, the dynamical friction force on a particle with momentum l is caused *only* by the particles with lower momentum if $f(l) = f(|l|)$. This result can be generalised to ellipsoidal distributions as well.

Finally, (5.67) allows the computation of the coefficient of dynamical friction if f is known. For a Maxwellian distribution of velocities

$$f(v) = A e^{-v^2/q^2} \quad (5.70)$$

we get, for small v ,

$$a(v) \cong -\frac{32}{3}\sqrt{\pi}n(Gm)^2L(v/q^3)(1 - \frac{3}{5}v^2/q^2 + \dots), \quad (5.71)$$

so that

$$\alpha(v) = -\frac{a(v)}{v} \cong \frac{16}{3\sqrt{\pi}} \frac{\ln N}{N} \frac{q}{R}. \quad (5.72)$$

Thus our estimate of $t_{R,\text{soft}}$ earlier is correct except for a factor of $3\sqrt{\pi}/16$. This is the timescale specified by $C(f)$.

It is straightforward to verify that the collision term derived above conserves the total energy. The mean field entropy

$$S(t) = -\int d\mathbf{v} d\mathbf{x} f(\mathbf{v}, \mathbf{x}, t) \ln f(\mathbf{v}, \mathbf{x}, t), \quad (5.73)$$

on the other hand, satisfies the inequality $dS/dt \geq 0$ during the Fokker–Planck evolution. These facts are relevant for the discussion of mean field stability presented in chapter 4.

5.6. A critical look at the derivation of the Fokker–Planck equation

In the last three sections we presented a derivation of $C(f)$ along conventional lines. During this derivation we came across several anomalies which require closer examination. We shall discuss some of these aspects in this section and attempt a more rigorous derivation of the collision term $C(f)$.

The most unsatisfactory feature about the derivation presented in the previous sections is the following: This derivation treats the left and right hand sides of (5.6) quite differently. The collision term is obtained by a completely independent reasoning and is attached to the evolution equation in an ad hoc manner. In any correct derivation of the dynamical equation the mean field and collision terms should arise together.

Even if we accept the adhocness of the derivation we are not entirely free from difficulties. The diffusion approximation, worked out in the previous sections, suffers from several technical shortcom-

ings: (i) This approximation treats the collisional transition as localised in configuration space; a particle located at \mathbf{x} and having a velocity \mathbf{v} is supposed to acquire a velocity $\mathbf{v} + \Delta\mathbf{v}$ after a collision without changing its configuration co-ordinate \mathbf{x} . This assumption is not easy to justify in systems with long range forces. (ii) Similar to the above assumption is the belief that the fluctuating forces act randomly. Such a randomness requires the forces at different times to be uncorrelated or, equivalently, the temporal correlation function of the force to die down rapidly with time. As we saw earlier, an approximate evaluation of this function suggests a very weak t^{-1} decay in time, casting doubt on the randomness of the fluctuating force. In fact, the rigorous derivation of the Fokker–Planck approximation from the fundamental principles of statistical mechanics fails if the force correlation function does not decrease faster than t^{-1} . Indirect evidence for this breakdown of the formalism is the logarithmic divergence we encountered in the derivation. In a rigorous derivation, we should not require any ad hoc prescriptions for making this logarithm finite.

Such a derivation can be attempted along the lines of quasi linear approximations used in plasma physics [Lifshitz and Pitaevskii 1982]. We shall first discuss this approximation in the context of a homogeneous plasma and then indicate the generalisation to gravitating systems [Kadomtsev and Pogutse 1970]. This analysis will be also useful in understanding the concept of collisionless relaxation in the next section.

Consider a distribution function $f_{\text{tot}}(\mathbf{x}, \mathbf{v}, t)$ describing the electrons in the plasma. Let us suppose that f_{tot} obeys the collisionless equation

$$\partial f_{\text{tot}}/\partial t + \mathbf{v} \cdot \nabla f_{\text{tot}} = (e/m)\mathbf{E} \cdot \partial f_{\text{tot}}/\partial \mathbf{v}, \quad \mathbf{E} = -\nabla\phi, \quad (5.74)$$

where \mathbf{E} is the electric field governed by Gauss' law,

$$\nabla \cdot \mathbf{E} = -4\pi e \left(\int f_{\text{tot}} d\mathbf{v} - n_0 \right). \quad (5.75)$$

Here n_0 is the constant density of background ions. The quasilinear approximation can be used when the distribution function f_{tot} can be separated as $f_{\text{tot}} = f_0 + f$, where f_0 is a slowly varying, smooth distribution and f is a rapidly varying component. The electric field is produced essentially by the fluctuating component f .

We shall *choose* this decomposition in such a way that the rapidly varying f satisfies the equation

$$\partial f/\partial t + \mathbf{v} \cdot \nabla f = (e/m)\mathbf{E} \cdot \partial f_0/\partial \mathbf{v}. \quad (5.76)$$

Since f_{tot} satisfies (5.74), it follows that

$$\partial f_0/\partial t = (e/m)\mathbf{E} \cdot \partial f/\partial \mathbf{v}. \quad (5.77)$$

This equation is inconsistent as it stands; we have assumed that f_0 , which appears on the left hand side, has no strong dependence on space or time, while f , which occurs on the right hand side, is likely to vary rapidly in space and time. The equality can be maintained only in some average sense. We shall, therefore, rewrite this equation as

$$\partial f_0/\partial t = (e/m)\langle \mathbf{E} \cdot \partial f/\partial \mathbf{v} \rangle, \quad (5.78)$$

where the averaging is performed over the rapidly varying E and f , to obtain a smoothed out contribution. The right hand side of (5.78) represents the collision term $C(f_0)$; our system is therefore described by eqs. (5.75), (5.76) and (5.78).

These equations can be brought into a more tractable form by Fourier transforming $f(\mathbf{x}, \mathbf{v}, t)$ and $\phi(\mathbf{x}, t)$ in space and time. If $f_{k\omega}(\mathbf{v})$ and $\phi_{k\omega}$ denote their Fourier transforms, then (5.76) can be solved to obtain

$$f_{k\omega}(\mathbf{v}) = \frac{e}{m} \left(\frac{1}{\omega - \mathbf{k} \cdot \mathbf{v}} \right) \left(\mathbf{k} \cdot \frac{\partial f_0}{\partial \mathbf{v}} \right) \phi_{k\omega} + g_{k\omega}(\mathbf{v}), \quad (5.79)$$

where $g_{k\omega}$ is any solution to the equation

$$(\omega - \mathbf{k} \cdot \mathbf{v})g_{k\omega} = 0, \quad (5.80)$$

and is related to the initial perturbation. Substituting (5.79) in (5.75) we find $\phi_{k\omega}$,

$$\phi_{k\omega} = -\frac{1}{\varepsilon(\mathbf{k}, \omega)} \frac{4\pi e}{k^2} \int g_{k\omega}(\mathbf{v}) d\mathbf{v}, \quad (5.81)$$

where the dielectric constant ε is given by

$$\begin{aligned} \varepsilon(\mathbf{k}, \omega) &= 1 + \frac{4\pi e^2}{mk^2} \int \frac{\mathbf{k} \cdot (\partial f_0 / \partial \mathbf{v})}{\omega - \mathbf{k} \cdot \mathbf{v}} d\mathbf{v} \\ &= 1 + \frac{4\pi e^2}{mk^2} P \int \frac{\mathbf{k} \cdot (\partial f_0 / \partial \mathbf{v})}{\omega - \mathbf{k} \cdot \mathbf{v}} d\mathbf{v} - i \frac{4\pi^2 e^2}{mk^2} \int \delta(\omega - \mathbf{k} \cdot \mathbf{v}) \mathbf{k} \cdot \frac{\partial f_0}{\partial \mathbf{v}} d\mathbf{v}. \end{aligned} \quad (5.82)$$

(Here P stands for the principal value of the integral.) We can now substitute the value of $E_{k\omega} = -i\mathbf{k}\phi_{k\omega}$ and (5.79) into the right hand side of (5.78). We get

$$E_{k\omega} \cdot \frac{\partial f_{k\omega}}{\partial \mathbf{v}} = +i\mathbf{k} \cdot \frac{\partial}{\partial \mathbf{v}} \left(\frac{e}{m} \frac{(\mathbf{k} \cdot \partial f_0 / \partial \mathbf{v}) \phi_{k\omega}}{\omega - \mathbf{k} \cdot \mathbf{v}} + g_{k\omega} \right) \frac{4\pi e}{k^2 \varepsilon_{k\omega}} \int g_{k\omega} d\mathbf{v}. \quad (5.83)$$

It is to be noted that this expression contains the full information about the propagation of disturbances in the plasma. It is convenient to consider separately the “wave” region (for which $k < \omega_0/v_{\text{thermal}}$, where ω_0 is the plasma frequency of the medium and v_{thermal} is some typical velocity) and the “non-wave” region (for which $k > \omega/v_{\text{thermal}}$). In the wave region, one can ignore the effects due to $g_{k\omega}$ at sufficiently late times and one gets the usual quasilinear theory (see, e.g., Lifshitz and Pitaevskii [1982], section 51). We are interested in the “non-wave” region, where the field fluctuations are produced by $g_{k\omega}$. In this case we can resort to the following approximations: (i) assume that the electric field is essentially produced by $g_{k\omega}$; (ii) replace the real part of $\varepsilon_{k\omega}$ by unity, and (iii) retain expressions only up to quadratic order in e . A detailed calculation [Kadomtsev and Pogutse 1970] now gives

$$\begin{aligned} C(f_0) &= \pi \left(\frac{e}{m} \right)^2 \int \frac{(4\pi e)^2}{k^4} k^a \frac{\partial}{\partial v^a} \left(\langle g(\mathbf{v}')g(\mathbf{v}'')^* \rangle_{k\omega} \delta(\omega - \mathbf{k} \cdot \mathbf{v}) \mathbf{k} \cdot \frac{\partial f_0}{\partial \mathbf{v}} \right. \\ &\quad \left. - \langle g(\mathbf{v}'')g^*(\mathbf{v}') \rangle_{k\omega} \delta(\omega - \mathbf{k} \cdot \mathbf{v}') \mathbf{k} \cdot \frac{\partial f_0}{\partial \mathbf{v}'} \right) d^3\mathbf{v}' d^3\mathbf{v}'' d^3\mathbf{k} d\omega, \end{aligned} \quad (5.84)$$

where $\langle g(\mathbf{v}')g(\mathbf{v}'') \rangle_{k\omega}$ etc. are defined as the Fourier transforms of the two point correlation functions,

$$\int \langle g(\mathbf{v}')g(\mathbf{v}'') \rangle_{k\omega} \exp(-i\omega\tau + i\mathbf{k} \cdot \boldsymbol{\xi}) d\mathbf{k} d\omega = \langle g(\mathbf{x} + \boldsymbol{\xi}, \mathbf{v}', t + \tau)g(\mathbf{x}, \mathbf{v}'', t) \rangle, \quad (5.85)$$

etc. The entire physics of the collision term is contained in these correlation functions. Notice that $g(\mathbf{x}, \mathbf{v}, 0)$ is related to the initial perturbation; at any later instant the dynamics will induce a certain amount of correlation even into a random initial perturbation. The nature of the collision term depends on our assumptions regarding the development of these correlations.

The Fokker–Planck equation arises from (5.84) if we assume that

$$\langle g(\mathbf{x} + \boldsymbol{\xi}, \mathbf{v}', t + \tau)g(\mathbf{x}, \mathbf{v}'', t) \rangle = f_0(\mathbf{v}')\delta(\mathbf{v}' - \mathbf{v}'')\delta(\boldsymbol{\xi} - \mathbf{v}'\tau). \quad (5.86)$$

The delta functions signify the propagation of perturbations along linear trajectories with constant velocities; the presence of $f_0(\mathbf{v}')$ implies that the probability of occurrence of any velocity is the same as that dictated by the background distribution. The Fourier transform of the correlation function will be

$$\langle g(\mathbf{v}')g(\mathbf{v}'') \rangle_{k\omega} = \frac{f_0(\mathbf{v}')}{(2\pi)^3} \delta(\mathbf{v}' - \mathbf{v}'')\delta(\omega - \mathbf{k} \cdot \mathbf{v}'). \quad (5.87)$$

Substituting this into (5.84) we find that

$$\frac{\partial f_0}{\partial t} = 2 \left(\frac{e}{m} \right)^2 e^2 \frac{\partial}{\partial v_a} \left\{ \int \frac{k^a k^b}{k^4} \left[\left(\frac{\partial}{\partial v^b} - \frac{\partial}{\partial v'^b} \right) f_0(\mathbf{v})f_0(\mathbf{v}') \right] \delta(k^i(v_i - v'_i)) d\mathbf{v}' d\mathbf{k} \right\}. \quad (5.88)$$

Performing integration over \mathbf{k} using the relation

$$\int \frac{d^3k}{k^4} k^a k^b \delta(\mathbf{k} \cdot \mathbf{p}) = -\frac{1}{2} \left(\frac{p^a p^b}{p^3} - \frac{\delta^{ab}}{p} \right) \int_{k_{\min}}^{k_{\max}} \frac{dk}{k} = -\frac{1}{2} \left(\frac{p^a p^b - p^2 \delta^{ab}}{p^3} \right) L, \quad (5.89)$$

we can see that (5.88) is the same as the Fokker–Planck equation derived earlier. As usual, L stands for the divergent Coulomb logarithm.

The above derivation depended on the following three crucial factors: (1) The total distribution function f_{tot} satisfied eq. (5.74). (2) A natural separation of f_{tot} into two parts with widely different timescales of evolution was possible. (3) A suitable averaging procedure could be defined to interpret the right hand side of (5.78). The overall homogeneity of the plasma simplified the mathematical details but was not essential for the derivation.

In the case of gravitating systems conditions (1) and (2) can easily be satisfied. The exact distribution function $f_{\text{ex}}(\mathbf{x}, \mathbf{v}, t)$ does satisfy the collisionless Boltzmann equation; the analysis in the previous sections shows that the timescale for fluctuating forces is much smaller than the dynamical timescale. Therefore, we can attempt a rigorous derivation of the Fokker–Planck equation provided a suitable averaging procedure can be defined.

Such a derivation was attempted by Kandrup along the following lines [Kandrup 1988; 1981]. We first define a quantity δf by the relation

$$f_{\text{ex}} = f_{\text{sm}} + \delta f \equiv f + \delta f, \quad (5.90)$$

in which we have dropped the subscript “sm” from f_{sm} to simplify the notation. We know that f_{ex} satisfies eq. (5.3); but, at this stage, we are free to choose the equations satisfied by f and δf individually. We will now choose the split (5.90) in such a way that δf satisfies the equation

$$\partial_t \delta f + D \delta f = \nabla_a \delta \phi \partial^a f, \tag{5.91}$$

where ϕ and $\delta \phi$ represent the potentials associated with f and δf , respectively, and we have used the notation

$$\partial_t \equiv \partial / \partial t, \quad \partial^a \equiv \partial / \partial v_a, \quad \nabla_a \equiv \partial / \partial x_a, \quad D \equiv v^a \nabla_a - \nabla_a \phi \partial^a. \tag{5.92}$$

The formal solution to this equation can be immediately written down,

$$\delta f(t) = \mathcal{G}(t, t_0) \delta f(t_0) + \int_0^{t-t_0} dt' \mathcal{G}(t, t-t') \nabla^a \delta \phi(t-t') \partial_a f(t-t'), \tag{5.93}$$

where \mathcal{G} is the Green’s function for the D -operator,

$$\mathcal{G}(t_2, t_1) \equiv T \exp \int_{t_1}^{t_2} dt D(t), \tag{5.94}$$

and $f(t_0)$ is the initial configuration given as a boundary condition. Notice that the Green’s function \mathcal{G} induces a flow determined by the background potential ϕ . Since f_{ex} satisfies (5.3) it is easy to verify that f must satisfy the equation

$$\partial_t f + Df = \nabla_a \delta \phi \partial^a \delta f. \tag{5.95}$$

We would like to interpret the right hand side of (5.95) as the collision term. This is not possible in the present form of (5.95) because it is not expressed in terms of a smooth function like f . To obtain an expression similar to the Fokker–Planck term it is essential that we replace the right hand side of (5.95) by a “suitably averaged” expression $\langle \nabla_a \delta \phi \partial^a \delta f \rangle$. Once we settle on the averaging process, an equation in closed form can be obtained by iterating (5.93) in the coupling constant G . Such an iteration is necessary because the right hand side of (5.93) involves $\delta \phi$, which, in turn, involves δf . We will need terms up to G^2 to produce a term similar to the Fokker–Planck term. Stopping the iteration at quadratic order corresponds to truncation of velocity increments at quadratic order in the earlier derivation of the Fokker–Planck equation.

The iteration of (5.93) is straightforward though messy. Once such an iteration is done the right hand side of (5.95) will be a quadratic expression in δf . The collision term we obtain from this equation is going to depend on the averaging that is used on the right hand side. A particularly simple expression emerges if we use the following prescription for evaluating the average of any variable \mathcal{A} defined in phase space [Kandrup 1988]:

$$\langle \mathcal{A} \rangle = \int \prod_k \frac{d\mathbf{x}_k d\mathbf{v}_k}{N} f(\mathbf{x}_k, \mathbf{v}_k, t) \mathcal{A}. \tag{5.96}$$

This is equivalent to assuming that the delta functions in f_{ex} are distributed with a mean density of f . Using this as our definition for averaging we can evaluate the collision term as

$$C(f) \equiv \langle \nabla_a \delta \phi \partial^a \delta f \rangle. \quad (5.97)$$

A detailed computation leads to the following result:

$$C(f) = \int_0^{t-t_0} N dt' \partial_a \int d\mathbf{x}_1 d\mathbf{v}_1 F^a(1, 0; t) \\ \times [F^b(1, 0; t-t') \partial_b f(\mathbf{x}, \mathbf{v}, t) f(\mathbf{x}_1, \mathbf{v}_1, t) + F^b(0, 1; t-t') \partial_b^1 f(\mathbf{x}, \mathbf{v}, t) f(\mathbf{x}_1, \mathbf{v}_1, t)], \quad (5.98)$$

where

$$F^a(1, 0; t) = Gm \nabla^a \int \frac{d\mathbf{x}_1 d\mathbf{v}_1}{N} \frac{\delta f(\mathbf{x}_1, \mathbf{v}_1, t)}{|\mathbf{x} - \mathbf{x}_1|} \\ = Gm \nabla^a \left(\frac{1}{|\mathbf{x} - \mathbf{x}_1|} - \int \frac{d\mathbf{x}_1 d\mathbf{v}_1}{N} \frac{f(\mathbf{x}_1, \mathbf{v}_1, t)}{|\mathbf{x} - \mathbf{x}_1|} \right), \quad (5.99)$$

etc. This is the fluctuating field at \mathbf{x} produced by any single field star. In writing down this expression we have set $N-1 \approx N$.

This collision term contains several interesting features. There is an integration over the time co-ordinate t' from zero to $t-t_0$ incorporating the correlation between the forces at different times. There is also an integration over the entire range of \mathbf{x}_1 taking into account the non-local nature of the force. If the distribution functions die down rapidly enough at large distances, then we will not face the logarithmic divergences encountered earlier.

It is worth investigating the limit in which (5.98) reduces to the Fokker-Planck term derived earlier. For this to happen we will need to make the following approximations: (i) Ignore the spatial dependence of f assuming homogeneity over the range from which most of the contributions occur; i.e., take $f(\mathbf{x}, \mathbf{v}, t)$ to be $f(\mathbf{v}, t)$ in the collision term. (ii) Ignore the contribution from the mean field force – which is the second term on the right hand side of (5.99) – in the collision term; this will also allow us to approximate the evolution described by the Green's function \mathcal{G} as along rectilinear trajectories: $\mathbf{x}(t) = \mathbf{x}(t_0) + \mathbf{v}(t_0)(t-t_0)$; $\mathbf{v}(t) = \mathbf{v}(t_0)$. In that case, we will get

$$\partial_t f(\mathbf{v}, t) = \int_0^{t-t_0} dt' (Gm)^2 n \partial^a \int d\mathbf{r} \int d\mathbf{v}_1 \nabla_a \frac{1}{|\mathbf{r}|} \nabla_b \frac{1}{|\mathbf{r} - \mathbf{u}t'|} (\partial^b - \partial_1^b) f(\mathbf{v}) f(\mathbf{v}_1), \quad (5.100)$$

where n is the number of density and $\mathbf{r} = \mathbf{x} - \mathbf{x}_1$; $\mathbf{u} = \mathbf{v} - \mathbf{v}_1$. If we do the t' integration using two cutoffs we get the Fokker-Planck result,

$$\partial_t f(\mathbf{v}) = \partial^a \int d\mathbf{v}_1 K_{ab} (\partial^b - \partial_1^b) f(\mathbf{v}) f(\mathbf{v}_1), \quad (5.101)$$

$$K_{ab} = 2\pi(Gm)^2 n (\delta_{ab}/u - u_a u_b / u^3) \ln(t_{\text{max}}/t_{\text{min}}). \quad (5.102)$$

It is easy to verify that this result is the same as the one derived in the last section but expressed in a different notation.

The above comparison reveals several things. Rigorously speaking, we must apply any approximation to both sides of the evolution equation; if we ignore the mean field and the \mathbf{x} dependence of f in the collision term, we should also ignore them on the left hand side. This is what is done in (5.101). But as it stands, (5.101) is fairly useless for studying the evolution of any realistic, inhomogeneous, system. What we require is an approximation in which the mean field and \mathbf{x} dependence of f are retained on the *left* hand side of the evolution equation but are ignored in the collision term. This can be done only in very special circumstances in which the two sides differ widely in scale lengths; in general, one has to use (5.98). In the earlier derivation we obtained a neat separation between the diffusion term and the dynamical friction term. This separation is not easy in the exact expression. The following point needs to be made regarding the dynamical friction: In general, there can exist collective processes which produce a damping at the same timescale as $t_{R, \text{soft}}$ [Marochnik 1968]. These processes are necessarily non-local in configuration space. Our derivation in section 5.5 does not take these processes into account though the error incurred is only sublogarithmic. Expression (5.98) takes these effects into account properly.

The main drawback in this approach seems to be the rather arbitrary nature of the averaging which we resorted to in (5.96). Further investigations are needed to justify this definition and clarify the dependence of our results on this assumption.

6. The approach to equilibrium: collisionless relaxation

6.1. Introduction

We found in the last section that the relaxation time t_R for collisional processes is about $N(\ln N)^{-1}$ times larger than the dynamical timescale. For most of the astrophysical systems, this timescale t_R turns out to be larger than the age of the universe. Therefore collisional processes could not have relaxed the astrophysical systems within their lifetime. On the other hand, observations suggest that most of the astrophysical systems have relaxed velocity distributions. It is therefore worth investigating whether some other relaxation mechanism could be operating in these systems.

One such mechanism – usually called “violent relaxation” – was suggested by Lynden-Bell [1967]. This process can take place within the dynamical timescale of the system and hence can relax the system very quickly. We shall examine some aspects of this process in this chapter.

6.2. The process of violent relaxation

Consider a system of N particles, described by the distribution function $f(\mathbf{x}, \mathbf{v}, t)$. If collisional effects are ignored (i.e., if we consider only timescales $t \ll t_R$), then the evolution of the system is described by the equations

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} - \nabla \phi \cdot \frac{\partial f}{\partial \mathbf{v}} = C(f) \cong 0, \quad (6.1)$$

$$\nabla^2 \phi(\mathbf{x}, t) = 4\pi G \int f(\mathbf{x}, \mathbf{v}, t) d^3 \mathbf{v}. \quad (6.2)$$

In writing these equations, we have already assumed that the granularity in the system may be ignored; that is, the distribution function $f(\mathbf{x}, \mathbf{v}, t)$ is assumed to be smooth over the mean interparticle distance $R/N^{1/3}$ of the system.

We are interested in two specific features of the evolution described by (6.1) and (6.2). The first feature to note is the following: These equations predict an infinite number of conserved quantities for the system. Since f satisfies (6.1), any functional of f of the form

$$Q[f] \equiv \int A[f(\mathbf{x}, \mathbf{v}, t)] \, d\mathbf{x} \, d\mathbf{v} \quad (6.3)$$

will be conserved in the course of the evolution because

$$\frac{dQ}{dt} = \int \frac{dA}{dt} \, d\mathbf{x} \, d\mathbf{v} = \int \frac{\partial A}{\partial f} \frac{df}{dt} \, d\mathbf{x} \, d\mathbf{v} = 0. \quad (6.4)$$

It is, therefore, clear that there will be severe constraints on the form of the distribution functions which can be obtained from a given initial distribution function through collisionless evolution.

The second important feature concerns the tendency of the phase elements to distribute themselves at finer and finer scales in the course of the collisionless evolution. Let us suppose that at some chosen time ($t = t_1$, say) the system occupies a compact region of phase space. As the evolution proceeds, the “area” of this region remains constant but its shape gets violently distorted. The reason for this distortion is easy to see: In a general potential, the orbital period will depend on the size (amplitude) of the orbit. Thus the synchronisation between particles populating different orbits will be progressively lost as the evolution proceeds. Thus, as time goes on, the phase elements will get mixed at finer and finer scales. To monitor this evolution we have to use a progressively larger and larger resolution in phase space.

In practice, however, we do not adopt such a procedure. We are usually interested in the average value of f within cells of *fixed* size in phase space. Let us suppose that the phase space is divided into cells of given volume μ . We can arrange matters in such a way that at $t = t_1$, k of these cells (in some region of phase space) are occupied by the system and rest of the cells are empty. As time goes on the phase density will spread into cells which were originally empty. If we wait long enough, almost all the cells will contain some amount of phase element. If we now average the phase density over the size of the cell, we will lose most of the details of the further evolution. In other words, the distribution function, averaged over the cells of size μ , has reached a steady state even though the evolution is proceeding at still finer and finer scales.

Such an average distribution function – called “coarse-grained” distribution function – may be formally defined by the relation

$$f_c(\mathbf{x}, \mathbf{v}, t) = \int K(\mathbf{x}, \mathbf{v}; \mathbf{x}', \mathbf{v}') f(\mathbf{x}', \mathbf{v}', t) \, d\mathbf{x}' \, d\mathbf{v}', \quad (6.5)$$

where the weight function K satisfies the following three conditions: (i) $K \geq 0$; (ii) $\int K \, d\mathbf{x} \, d\mathbf{v} = 1$; (iii) $\int K \, d\mathbf{x}' \, d\mathbf{v}' = 1$. The simplest kind of weight function is the one which is constant over a cell of finite volume μ and zero outside. In that case

$$f_c(\mathbf{x}, \mathbf{v}, t) = \int_{\mu} f(\mathbf{x} + \mathbf{x}', \mathbf{v} + \mathbf{v}', t) \frac{d\mathbf{x}' \, d\mathbf{v}'}{\mu}, \quad (6.6)$$

where the integration is over a cell of volume μ centred at (\mathbf{x}, \mathbf{v}) . After a certain time interval, the evolution in f_c becomes negligible; one can then say that the coarse-grained distribution function f_c has reached a steady state or “equilibrium”.

The nature of this “equilibrium” must be clear from the above discussion. It originates because of the finite size of the cell over which coarse-graining is performed. As the system evolves the changes in the distribution function occur at smaller and smaller scales. Once this scale at which changes occur becomes smaller than the size of the coarse-graining cell, f_c ceases to exhibit any evolution.

There exists another *distinct* process which contributes to the collisionless relaxation. This arises from the fact that, during the evolution, the particles in the system are experiencing a time dependent gravitational potential, $\phi(\mathbf{x}, t)$. The energy E of an individual particle is, therefore, not constant during the evolution (in fact, it is easy to see that $\dot{E} = \partial\phi/\partial t$). The change in the energy of any given particle, during the course of the evolution, will clearly depend in a complex way on the initial conditions of the particle; but the overall effect of this process is to widen the range of energies of the particles. In this respect, the time dependent gravitational field provides a relaxation process similar to the collisions in a gaseous system, with one important distinction: Since the mass of the individual particle does not appear in the equations, this relaxation process changes the energy per unit mass of a given particle in a way that is independent of the mass of the particle; normal collisional relaxation, on the other hand, leads to equipartition of *energy*. It should be emphasised that the two processes discussed above – mixing in phase space and non-conservation of energies of individual particles – are quite distinct and their roles in relaxing a collisionless system are conceptually different.

In the spirit of statistical mechanics, we may ask the question: What is the most probable f_c which is consistent with the given initial distribution function? In the course of a normal statistical evolution, we expect the final state to have the same constants of the motion as the initial state. These constants of the motion will be finite in number. Here, the collisionless evolution conserves an infinite set of quantities. As a result, we expect the final state to depend on the details of the initial state more strongly. To write down f_c , we will require far more information about the initial state than is contained in the knowledge of the total mass and energy alone.

Lynden-Bell attempted to obtain the form of f_c based on the following assumptions: (a) The evolution proceeds in a such a way that the volume of the region with a given phase density is conserved, (b) the total energy of the system is conserved, and (c) the relaxation proceeds to its completion. Under these circumstances, it is possible to determine the form of f_c as a weighted superposition of Fermi–Dirac distributions with different temperatures. The weight factors depend on the initial distribution. Since the process which leads to the relaxation is governed by the dynamical timescale, we expect this process to be considerably faster than collisional relaxation.

The detailed properties of this distribution function f_c are discussed by Lynden-Bell [1967]. It may be noted that in the conditions appropriate for astrophysical systems, the distribution function reduces to a superposition of several Maxwellian distributions. Thus violent relaxation, *if it is completely efficient*, as assumed by Lynden-Bell, can produce a coarse-grained distribution function which is approximately Maxwellian.

The central difficulty in accepting the distribution function derived by Lynden-Bell is the following: it predicts infinite mass for the system. In other words, our variational problem which determines the most probable coarse grained distribution function possesses no solution for any finite total mass, contradicting the original assumption. It is usually believed that this difficulty arises because of our insistence on complete relaxation. In realistic systems, the outer regions do not relax completely. Thus the distribution function will be significantly different from f_c in the outer regions leading to a finite total

mass. Alternatively, one may attempt to derive the most probable f_c after imposing more constraints on the system. In either case, however, we lose our ability to predict any simple form for f_c .

The exact form of f_c , therefore, depends on the details of the assumptions we are prepared to make. It is, however, possible to put some interesting constraints on the evolution of the distribution function, based on the nature of the collisionless relaxation. We shall consider the origin of such constraints in the next section.

6.3. Constraints on collisionless evolution

To study the constraints imposed on the distribution function during collisionless evolution, we will require the concept of a ‘‘convex’’ function. We define a function $C(x)$ to be convex if

$$C\left(\frac{\sum p_i x_i}{\sum p_i}\right) \leq \frac{\sum p_i C(x_i)}{\sum p_i}, \quad (6.7)$$

for all $(x_i, p_i) > 0$. This is equivalent to saying that $C(\langle x \rangle)$ is less than or equal to $\langle C(x) \rangle$, where the angular brackets denote some averaging process with positive definite weights. Two other, equivalent, definitions of convex functions which we will use are the following: (i) If $C(x)$ is twice differentiable then it is convex if $d^2C/dx^2 \geq 0$ in the region of interest. (ii) If $C(x)$ is differentiable once then it is convex if $C(x_2) - C(x_1) \geq (x_2 - x_1)C'(x_1)$ for all (x_1, x_2) in the region of interest.

Consider any functional H of the coarse-grained distribution function f_c defined by the integral

$$H[f_c] = - \int C(f_c) \, dx \, dv, \quad (6.8)$$

where C is a convex function. Let the original (‘‘fine-grained’’) distribution function corresponding to f_c be f ; we shall also assume that at some initial moment $t = t_1$, $f_c(t_1) = f(t_1)$. It can then be easily shown that $H(t_2) \geq H(t_1)$ for all $t_2 > t_1$ [Tremaine et al. 1986]. We notice that

$$\begin{aligned} H(t_2) - H(t_1) &= \int dx \, dv [C(f_c(t_1)) - C(f_c(t_2))] = \int dx \, dv [C(f(t_1)) - C(f_c(t_2))] \\ &= \int dx \, dv [C(f(t_2)) - C(f_c(t_2))]. \end{aligned} \quad (6.9)$$

[The second equality follows from the fact that at $t = t_1$, $f_c(t_1) = f(t_1)$; the third from the conservation of $C(f)$ under time evolution; see (6.4).] But the coarse-grained distribution function f_c is obtained by averaging f over the phase cells. Therefore, using our definition of a convex function (6.7), we may conclude that

$$\int dx \, dv C(f_c) \leq \int dx \, dv C(f), \quad (6.10)$$

and hence

$$H(t_2) > H(t_1). \quad (6.11)$$

It should be stressed that $H(t)$, in general, is *not* a monotonically increasing function of t . The instant t_1 is special in the sense that we set $f_c = f$ at that instant. The above argument only shows that H -functions will always have values higher than the value taken at this special instant. For two arbitrary instants $t_2, t_3 > t_1$, nothing can be said about the relative values of $H(t_2)$ and $H(t_3)$.

The above condition shows that collisionless evolution must proceed in such a manner that *all* convex H -functionals increase during the evolution. Suppose we are given two distribution functions $f_c(t_2)$ and $f_c(t_1)$ with $t_2 > t_1$. Collisionless evolution could have evolved $f_c(t_1)$ to $f_c(t_2)$ only if all H -functionals satisfy the condition $H[f_c(t_2)] > H[f_c(t_1)]$.

It is, of course, not possible to test this inequality for an infinite number of H -functionals. Neither is it necessary; a much more useful form of the above constraint can be easily developed along the following lines [Tremaine et al. 1986].

Given a coarse-grained distribution function f_c we define the volume of phase space with phase density larger than q by the relation

$$V(q) = \int dx dv \theta(f_c(x, v) - q), \quad (6.12)$$

where $\theta(z)$ is unity for $z > 0$ and is zero otherwise. Similarly, the mass contained in the region with phase density larger than q is defined as

$$M(q) = \int dx dv f_c \theta(f_c - q). \quad (6.13)$$

Clearly

$$\frac{dM}{dq} = - \int dx dv f_c \delta(f_c - q) = -q \int dx dv \delta(f_c - q) = +q \frac{dV}{dq}, \quad (6.14)$$

so that

$$M(V) = \int_0^V q(V') dV', \quad (6.15)$$

where $q(V)$ is the inverse function of $V(q)$. Given the distribution function f_c we can obtain $M(V)$ by eliminating q between relations (6.12) and (6.13).

Now suppose we are given two distribution functions f_1 and f_2 with corresponding $M_1(V)$ and $M_2(V)$. We can then show that f_1 could have evolved into f_2 if and only if $M_2(V) \leq M_1(V)$ for all V .

We shall first show that if $M_2(V) \leq M_1(V)$ then $H(f_2) > H(f_1)$ for all H -functions. Now

$$H(f_2) - H(f_1) = \int dx dv [C(f_1) - C(f_2)] = \int_0^\infty dV [C(f_1) - C(f_2)]. \quad (6.16)$$

Using the fact that for convex $C(f)$, $C(f_1) - C(f_2) \geq (f_1 - f_2)C'(f_2)$, we get

$$H_2 - H_1 \geq \int_0^\infty dV (f_1 - f_2)C'(f_2) = \int_0^\infty dV \frac{d}{dV} (M_1 - M_2)C'(f_2). \quad (6.17)$$

Integrating by parts,

$$H_2 - H_1 \geq [(M_1 - M_2)C'(f_2)]_0^\infty - \int_0^\infty dV (M_1 - M_2)C''(df_2/dV) dV. \quad (6.18)$$

The first term vanishes at both limits; the second term is non-negative because $C'' > 0$ and $df_2/dV \leq 0$. Thus $H_2 > H_1$.

We can also prove that if f_1 can evolve into f_2 then $M_2(V) < M_1(V)$. This can be easily done by using the convex function

$$C(f) = \begin{cases} 0, & f \leq 0, \\ f - \phi, & f > \phi, \end{cases} \quad (6.19)$$

where $\phi = f(V_0)$ with some arbitrary V_0 . The fact that the H -function constructed out of this particular $C(f)$ should be non-decreasing leads to the condition $M_2(V_0) \leq M_1(V_0)$.

One useful corollary of the above constraint is the following: The maximum value of the phase density can only decrease during collisionless evolution. Since $f_{\max} = f(V=0)$, the Taylor expansion of $M(V)$ gives $M(V) \approx f_{\max} V$; the condition $M_2 < M_1$ immediately leads to the conclusion: $f_{\max}^{(2)} < f_{\max}^{(1)}$.

The constraints derived above are useful in testing the hypothesis that astrophysical systems have undergone violent relaxation. Given a theoretical model for the initial distribution function we can test whether collisionless relaxation could have evolved the astrophysical systems to the presently observed form. Some such examples are discussed by Tremaine et al. in their paper cited above. It seems, however, that this aspect of the study has not yet received the attention it deserves.

6.4. Dynamical description of collisionless relaxation

We have now seen that two separate relaxation mechanisms work in gravitating systems. Our description of both processes has been quite ad hoc. A truly fundamental description, starting from first principles, must be able to derive both these processes in a similar manner. We shall discuss the ingredients of such a derivation in this section.

A fundamental distinction between the two processes must be emphasised right at the outset. The collisional relaxation – discussed in chapter 5 – produces a final state which is independent of the details of the initial state. In this sense, collisional evolution is quite similar to the relaxation seen in laboratory systems. This is not true as regards violent relaxation; the final state produced by violent relaxation does remember details of the initial state. In fact, the efficient operation of violent relaxation requires the initial state to be far away from equilibrium.

It should also be noted that we needed the concept of coarse-grained distribution function to introduce the concept of violent relaxation. To be precise, we have so far used two (different) levels of smoothing of the exact distribution function $f_{\text{ex}}(\mathbf{x}, \mathbf{v}, t)$: (i) We first smooth f_{ex} over scales large compared to the mean interparticle separation of the system, thereby obtaining f_{smooth} . We showed in chapter 5 that f_{smooth} satisfies, to a good approximation, the Fokker–Planck equation, while f_{ex} , which is a sum of delta functions, satisfies the collisionless Boltzmann equation. (ii) We next consider the evolution of f_{smooth} at times $t \ll t_R$. In this case, the Fokker–Planck collision term can be ignored and f_{smooth} itself satisfies the collisionless Boltzmann equation. The evolution of f_{smooth} produces a considerable amount of mixing in phase space. We now define a coarse-grained distribution function f_c by

smoothing f_{smooth} over cells of some fixed size μ in phase space. Violent relaxation can lead to a steady state for this f_c on a timescale which is small compared to the collisional relaxation time.

The above discussion shows that the evolution of f_c depends both on the initial state and the nature of the coarse-graining. Given the equation satisfied by f_{smooth} and the weight function K used for coarse-graining we can write down the (formal) equation satisfied by f_c . However, this will not lead to any fundamental description; it is clear that no relaxation will occur if we coarse-grain the distribution function over a cell size which is too small while no evolution will occur if we coarse-grain over cells which are too big. The entire process depends on the existence of a suitable intermediate scale and a sufficiently non-equilibrium initial state.

The most fruitful approach to study the evolution of f_c seems to be the quasilinear theory discussed in chapter 5. To adapt that discussion to the present context we may take for f_{total} the distribution f_{smooth} and for f_0 the coarse-grained distribution f_c . Let us assume, for a moment, that we can study a gravitating system ignoring the inhomogeneities. Then the derivation of eq. (5.84) can be translated for a gravitating system to obtain

$$\begin{aligned} \frac{\partial f_c}{\partial t} = \text{const.} \int \frac{(4\pi G)^2}{k^4} k^a \frac{\partial}{\partial v^a} \left(\langle g(\mathbf{v}')g(\mathbf{v}'')^* \rangle_{k\omega} \delta(\omega - \mathbf{k} \cdot \mathbf{v}) \mathbf{k} \cdot \frac{\partial f_c}{\partial \mathbf{v}} \right. \\ \left. - \langle g(\mathbf{v}'')g^*(\mathbf{v}') \rangle_{k\omega} \delta(\omega - \mathbf{k} \cdot \mathbf{v}') \mathbf{k} \cdot \frac{\partial f_c}{\partial \mathbf{v}'} \right) d^3 \mathbf{v}' d^3 \mathbf{v}'' d^3 \mathbf{k} d\omega . \end{aligned} \quad (6.20)$$

To proceed further we need information about the correlation functions. The assumption of linear trajectories with constant velocity implies that the correlation function must have the form

$$\langle g(\mathbf{v})g(\mathbf{v}') \rangle_{k\omega} = \delta(\omega - \mathbf{k} \cdot \mathbf{v}) \delta(\mathbf{v} - \mathbf{v}') A[f_c] , \quad (6.21)$$

where $A[f_c]$ is some functional of f_c . In chapter 5, we assumed that $A[f_c] = f_c$; this is equivalent to demanding complete lack of correlations. To obtain conditions favourable for violent relaxation we have to proceed differently and assume that strong correlations exist in certain regions of phase space. This is best done in the following manner [Kadomtsev and Pogutse 1970].

We note that the quantity A is related to the equal time correlation function via the equation

$$\langle g(\mathbf{r} + \boldsymbol{\xi}, \mathbf{v}, t)g(\mathbf{r}, \mathbf{v}', t) \rangle = (2\pi)^3 A \delta(\mathbf{v} - \mathbf{v}') \delta(\boldsymbol{\xi}) . \quad (6.22)$$

Consider a situation in which f_{tot} was initially equal to unity over certain regions and zero outside these regions. Since this condition is conserved in time we can write

$$\langle f_{\text{tot}}(\mathbf{x}, \mathbf{v}, t)f_{\text{tot}}(\mathbf{x}, \mathbf{v}, t) \rangle = \langle 1 \times f_{\text{tot}}(\mathbf{x}, \mathbf{v}, t) \rangle = f_c . \quad (6.23)$$

Therefore

$$\langle g(\mathbf{r}, \mathbf{v}, t)g(\mathbf{r}, \mathbf{v}, t) \rangle = \langle (f_{\text{tot}} - f_c)(f_{\text{tot}} - f_c) \rangle = f_c(1 - f_c) . \quad (6.24)$$

Thus the correlation is equal to $f_c(1 - f_c)$ when the two points (\mathbf{x}, \mathbf{v}) , $(\mathbf{x}', \mathbf{v}')$ coincide and is zero elsewhere. In reality we expect the correlation to be non-zero over a region of some size $(\Delta x)^3(\Delta v)^3 \equiv$

q^2 in phase space. We, therefore, write

$$(2\pi)^3 A = (\Delta x)^3 (\Delta v)^3 f_c (1 - f_c) = q^2 f_c (1 - f_c). \quad (6.25)$$

In this case, the collision term becomes

$$C(f_c) = \text{const.} (Gq)^2 \int \frac{\mathbf{k}}{k^4} \cdot \frac{\partial}{\partial \mathbf{v}} \left(\mathbf{k} \cdot \frac{\partial f_c}{\partial \mathbf{v}} f'_c (1 - f'_c) - \mathbf{k} \cdot \frac{\partial f'_c}{\partial \mathbf{v}'} f_c (1 - f_c) \right) \delta(\mathbf{k} \cdot \mathbf{v} - \mathbf{k} \cdot \mathbf{v}') d\mathbf{v}' d\mathbf{k}. \quad (6.26)$$

We see that the equilibrium solution to this equation [i.e. the distribution for which $C(f) = 0$] is a Fermi–Dirac distribution,

$$f_{c,\text{eq}}(v) = [\exp \beta(\frac{1}{2}mv^2 + \mu) + 1]^{-1}, \quad (6.27)$$

as predicted by violent relaxation. Also note that the effective coupling constant for the system is now Gq and not G . The relaxation time t_R will therefore be reduced by a factor $q^{1/2}$. The larger the domain over which the distribution is correlated, the quicker violent relaxation will be.

The above analysis suggests the reinterpretation of violent relaxation as collisional relaxation of macroscopically correlated domains. The relaxation time t_R has the dependence $N_{\text{eff}}(\ln N_{\text{eff}})^{-1}$, where N_{eff} is the *effective* number of particles in the system. If large regions of phase space are strongly correlated, these regions behave as “macroparticles” and the number N_{eff} should be taken as the number of such regions. (Quite clearly $N_{\text{eff}} \ll N$.) This will lead to a reduction of the relaxation time t_R .

The discussion above is based on treating the distribution function which obeys the collisionless Boltzmann equation as some sort of an average. We mentioned in section 5.2 that this procedure may not be entirely satisfactory. It is often suggested in the literature (see, e.g., Binney and Tremaine [1987], p. 194) that f is better defined as a probability density. This cautionary note must be borne in mind while interpreting the results discussed in this section.

7. Conclusions

The topics discussed in this review highlight the nontrivial – and somewhat counterintuitive – facets of gravitational statistical mechanics. It is probably only fair to say that we do not have at present a systematic understanding of this subject at a level similar to, say, the kinetic theory of plasmas. Based on what we know, we may say that the evolution of a set of N gravitating hard spheres confined inside a region of radius R , will proceed through three different stages. In the first stage, characterised by the dynamical timescale $t_{\text{dyn}} \approx R/v$, the system will relax violently and will (probably) reach an approximately Maxwellian distribution. In the second stage, the system is essentially governed by the mean potential and the constituent particles move along orbits in this mean potential. In the last stage, characterised by a timescale t_R , the system plunges to the configuration which is the global maximum for the entropy. This state is likely to be made of a few, hard, contact binaries surrounded by a halo of high velocity stars bouncing off the walls of the container. The timescale t_R will be the two body relaxation time if the system is *not* Antonov stable. But if it is, the timescale t_R can be considerably larger than the two-body relaxation time. If the system is not confined by the walls of the container then

there will be continuous evaporation of particles from the system. In general, this will accelerate the processes mentioned above. If the evaporation is allowed most of the system will ultimately disperse and we will be left with a few, very tightly bound, binaries.

It is clear that the final equilibrium state is characterised by an interplay between two phases of the system: the high velocity component dominated by the kinetic energy and a hard, compact core dominated by the potential energy. (In the absence of a confining wall, the high velocity component would have dispersed to infinity.) The models studied in chapters 3 and 4 provide a fair amount of insight into such equilibrium configurations. The essential physics seems to be represented by these models even though the details may differ from system to system.

The issue of approach to this equilibrium, unfortunately, is not so well understood. As we saw in chapters 5 and 6, it is not easy to provide a good dynamical model describing the relaxation process. There are several conceptual issues which are still unresolved. If the analogy between virialised gravitating systems and gaseous systems at the verge of a phase transition is really of fundamental significance, then it may be worthwhile to examine whether the techniques used to study the latter may be adopted to study the former. As a beginning, one should try to understand the dynamical evolution of the toy models discussed in part 3. This problem is under investigation.

Acknowledgements

I thank: Prof. D. Lynden-Bell for introducing me to this fascinating subject and for several illuminating comments; R. Nityananda, for several detailed discussions during the last two years, which have helped the evolution of many of the ideas expressed here; K. Subramanian, S. Sreedhar, D. Narasimha and T.P. Singh for many useful debates and comments; H.M. Antia for teaching me $\text{T}_{\text{E}}\text{X}$, letting me use the macros developed by him and for being very helpful in sorting out the $\text{T}_{\text{E}}\text{X}$ problems; and the referee for several constructive comments.

References

- Ahmed, A. and L. Cohen, 1973, *Astrophys. J.* 179, 885.
 Antonov, V.A., 1962, *Vestn. Leningrad Gos. Univ.* 7, 135 [English translation in: *Dynamics of Globular Clusters*, IAU Symp. 113, eds. J. Goodman and P. Hut (Reidel, Dordrecht, 1985)].
 Aronson, E.B. and C.J. Hansen, 1972, *Astrophys. J.* 177, 145.
 Binney, J.J. and C.G. Lacey, 1988, *Mon. Not. R. Astron. Soc.* 230, 597.
 Binney, J.J. and S.D. Tremaine, 1987, *Galactic Dynamics* (Princeton Univ. Press, Princeton, NJ).
 Bonner, W.B., 1956, *Mon. Not. R. Astron. Soc.* 116, 351.
 Bonner, W.B., 1958, *Mon. Not. R. Astron. Soc.* 118, 523.
 Chandrasekhar, S., 1939, *Introduction to the Study of the Stellar Structure* (Dover Publ., New York).
 Chandrasekhar, S., 1941, *Astrophys. J.* 94, 511.
 Chandrasekhar, S., 1943, *Rev. Mod. Phys.* 15, 1.
 Chandrasekhar, S., 1944, *Astrophys. J.* 99, 25.
 Chandrasekhar, S. and J. Von Neumann, 1942, *Astrophys. J.* 95, 489.
 Cohn, H., 1980, *Astrophys. J.* 242, 765.
 Davidsen, A.F. and J.P. Ostriker, 1968, *Astrophys. J.* 151, 679.
 Dyson, F.J., 1967, *J. Math. Phys.* 8, 1538.
 Dyson, F.J. and A. Lenard, 1967, *J. Math. Phys.* 8, 423.
 Elson, R., P. Hut and S. Inagaki, 1987, *Annu. Rev. Astron. Astrophys.* 25, 565.
 Fowler, R.H., 1936, *Statistical Mechanics* (Cambridge Univ. Press, Cambridge, UK).

- Gibbs, J.W., 1902, *Elementary Principles in Statistical Mechanics* (Yale Univ. Press, New Haven, CT); reprinted by Dover Publ. (New York, 1960).
- Hachisu, I. et al., 1978, *Prog. Theor. Phys.* 60, 393.
- Hertel, P. and W. Thirring, 1971, *Commun. Math. Phys.* 24, 22.
- Hertel, P. and W. Thirring, 1972, *Commun. Math. Phys.* 28, 159.
- Hertel, P., H. Narnhofer and W. Thirring, 1972, *Commun. Math. Phys.* 28, 159.
- Horowitz, G. and J. Katz, 1977, *Astrophys. J.* 211, 226.
- Inagaki, S., 1980, *Publ. Astron. Soc. Jpn.* 32, 213.
- Kadomtsev, B.B. and O.P. Pogutse, 1970, *Phys. Rev. Lett.* 25, 1155.
- Kandrup, H.E., 1980, *Phys. Rep.* 63, 1.
- Kandrup, H.E., 1981, *Astrophys. J.* 244, 316.
- Kandrup, H.E., 1988, *Mon. Not. R. Astron. Soc.* 235, 1157.
- Katz, J., 1978, *Mon. Not. R. Astron. Soc.* 183, 765.
- Katz, J., 1979, *Mon. Not. R. Astron. Soc.* 189, 817.
- Katz, J. and D. Lynden-Bell, 1978, *Mon. Not. R. Astron. Soc.* 184, 709.
- Landau, L.D. and E.M. Lifshitz, 1980, *Statistical Mechanics* (Pergamon, Oxford).
- Larson, R.B., 1970a, *Mon. Not. R. Astron. Soc.* 147, 323.
- Larson, R.B., 1970b, *Mon. Not. R. Astron. Soc.* 150, 93.
- Lee, E., 1968, *Astrophys. J.* 151, 687.
- Lenard, A. and F.J. Dyson, 1968, *J. Math. Phys.* 9, 698.
- Levy-Leblond, J.M., 1969, *J. Math. Phys.* 10, 806.
- Lieb, E.H. 1976, *Rev. Mod. Phys.* 48, 568.
- Lieb, E.H. and J.L. Lebowitz, 1972, *Adv. Math.* 9, 316.
- Lieb, E.H. and W.E. Thirring, 1975, *Phys. Rev. Lett.* 35, 687.
- Lifshitz, E.M. and L.P. Pitaevskii, 1982, *Physical Kinetics* (Pergamon, Oxford).
- Lynden-Bell, D., 1967, *Mon. Not. R. Astron. Soc.* 136, 101.
- Lynden-Bell, D. and P.P. Eggeleton, 1980, *Mon. Not. R. Astron. Soc.* 191, 483.
- Lynden-Bell, D. and S. Inagaki, 1983, *Mon. Not. R. Astron. Soc.* 205, 913.
- Lynden-Bell, D. and R.M. Lynden-Bell, 1977, *Mon. Not. R. Astron. Soc.* 181, 405.
- Lynden-Bell, D. and R. Wood, 1968, *Mon. Not. R. Astron. Soc.* 138, 495.
- Marochnik, L.S. 1968, *Sov. Astron. J.* 11, 873.
- Narasimha, D. and T. Padmanabhan, 1989, TIFR-TAP preprint.
- Ostriker, J., 1964, *Astrophys. J.* 140, 1056.
- Padmanabhan, T., 1988, *Gravitational Many Body Problem*, TIFR lecture notes, preprint TIFR-TAP-10/88 (unpublished).
- Padmanabhan, T., 1989a, *Astrophys. J.*, to appear.
- Padmanabhan, T., 1989b, TIFR-TAP preprint.
- Padmanabhan, T., 1989c, *Astrophys. J. Suppl.*, to appear.
- Penrose, O., 1970, *Foundations of Statistical Mechanics: a Deductive Treatment* (Pergamon, Oxford).
- Prigogine, I. and G. Severne, 1966, *Physica* 32, 1376.
- Rosenbluth, M., W. MacDonald and D. Judd, 1957, *Phys. Rev.* 107, 1.
- Salzberg, A.M., 1965, *J. Math. Phys.* 6, 158.
- Saslaw, W.C., 1985, *Gravitational Physics of Stellar and Galactic Systems* (Cambridge Univ. Press, Cambridge).
- Spitzer, L., 1987, *Dynamical Evolution of Globular Clusters* (Princeton Univ. Press, Princeton, NJ).
- Stodolkiewicz, J.S., 1963, *Acta Austr.* 13, 30.
- Ter Haar, D., 1954, *Elements of Statistical Mechanics* (Rinehart, New York).
- Ter Haar, D., 1955, *Rev. Mod. Phys.* 27, 289.
- Ter Haar, D., 1961, *Rep. Prog. Phys.* 24, 304.
- Thirring, W., 1970, *Z. Phys.* 235, 339.
- Tolman, R.C., 1936, *The Principles of Statistical Mechanics* (Oxford Univ. Press, Oxford).
- Tremaine, S., M. Henon and D. Lynden-Bell, 1986, *Mon. Not. R. Astron. Soc.* 219, 285.