1

Astrophysics and the three-body problem

1.1 About the three-body problem

The three-body problem arises in many different contexts in nature. This book deals with the classical three-body problem, the problem of motion of three celestial bodies under their mutual gravitational attraction. It is an old problem and logically follows from the two-body problem which was solved by Newton in his *Principia* in 1687. Newton also considered the three-body problem in connection with the motion of the Moon under the influences of the Sun and the Earth, the consequences of which included a headache.

There are good reasons to study the three-body gravitational problem. The motion of the Earth and other planets around the Sun is not strictly a two-body problem. The gravitational pull by another planet constitutes an extra force which tries to steer the planet off its elliptical path. One may even worry, as scientists did in the eighteenth century, whether the extra force might change the orbital course of the Earth entirely and make it fall into the Sun or escape to cold outer space. This was a legitimate worry at the time when the Earth was thought to be only a few thousand years old, and all possible combinations of planetary influences on the orbit of the Earth had not yet had time to occur.

Another serious question was the influence of the Moon on the motion of the Earth. Would it have long term major effects? Is the Moon in a stable orbit about the Earth or might it one day crash on us? The motion of the Moon was also a question of major practical significance, since the Moon was used as a universal time keeping device in the absence of clocks which were accurate over long periods of time. After Newton, the lunar theory was studied in the eighteenth century using the restricted problem of three bodies (Euler 1772). In the restricted problem, one of the bodies is regarded as massless in comparison with the other two which are in a circular orbit relative to each other. At about the same time, the first special solution of the general three-body problem was discovered, the Lagrangian equilateral triangle
solution (Lagrange 1778). The theory of the restricted three-body problem was further developed by Jacobi (1836), and it was used for the purpose of identifying comets by Tisserand (1889, 1896) and reached its peak in the later nineteenth century with the work of Hill (1878) and Delaunay (1860). The ‘classical’ period reached its final phase with Poincaré (1892–1899).

In spite of these successes in special cases, the solution of the general three-body problem remained elusive even after two hundred years following the publication of *Principia*. In the general three-body problem all three masses are non-zero and their initial positions and velocities are not arranged in any particular way. The difficulty of the general three-body problem derives from the fact that there are no coordinate transformations which would simplify the problem greatly. This is in contrast to the two-body problem where the solutions are found most easily in the centre of mass coordinate system. The mutual force between the two bodies points towards the centre of mass, a stationary point in this coordinate system. Thus the solution is derived from the motion in the inverse square force field. Similarly, in the restricted three-body problem one may transfer to a coordinate system which rotates at the same rate around the centre of mass as the two primary bodies. Then the problem is reduced to the study of motions in two stationary inverse square force fields. In the general problem, the lines of mutual forces do not pass through the centre of mass of the system. The motion of each body has to be considered in conjunction with the motions of the other two bodies, which made the problem rather intractable analytically before the age of powerful computers.

At the suggestion of leading scientists, the King of Sweden Oscar II established a prize for the solution of the general three-body problem. The solution was to be in the form of a series expansion which describes the positions of the three bodies at all future moments of time following an arbitrary starting configuration. Nobody was able to claim the prize for many years and finally it was awarded in 1889 to Poincaré who was thought to have made the most progress in the subject even though he had not solved the specific problem. It took more than twenty years before Sundman completed the given task (Sundman 1912). Unfortunately, the extremely poor convergence of the series expansion discovered by Sundman makes this method useless for the purpose of calculating the orbits of the three bodies. Now that the orbits can be calculated quickly by computer, it is quite obvious why this line of research could not lead to a real solution of the three-body problem: the orbits are good examples of chaos in nature, and deterministic series expansions are utterly unsuitable for their description. Poincaré was on the right track in this regard and with the current knowledge was thus a most reasonable recipient of the prize.

At about the same time, a new approach began which has been so successful in recent years: the integration of orbits step by step. In orbit integration, each body,
in turn, is moved forward by small steps. In the most basic scheme, the step is calculated on the basis of the accelerations caused by the two other bodies during that step while they are considered to remain fixed. There is an error involved when only one body is moved at a time, and others move later, but this can be minimised by taking short steps and by other less obvious means.

Burrau (1913) considered a well defined, but in no way special, initial configuration of three bodies which has since become known as the Pythagorean problem since the three bodies are initially at the corners of a Pythagorean right triangle. The masses of the three bodies are 3, 4 and 5 units, and they are placed at the corners which face the sides of the triangle of the corresponding length (Fig. 1.1). In the beginning the bodies are at rest. Burrau’s calculation revealed the typical behaviour of a three-body system: two bodies approach each other, have a close encounter, and then recede again. Subsequently, other two-body encounters were calculated by Burrau until he came to the end of his calculating capacity. Only after the introduction of modern computers and new orbit integration methods was the celestial dance in the Pythagorean problem followed to its conclusion.

Later work has shown that the solution of the Pythagorean problem is quite typical of initially bound three-body systems. After many close two-body approaches, a configuration arises which leads to an escape of one body and the formation of a binary by the other two bodies (see Fig. 1.2, Szebehely and Peters 1967). A theoretical treatment of a three-body system of this kind is given in Chapter 7. In the following chapter, situations are discussed where a third body comes from a large distance, meets a binary, and perhaps takes the place of one of the binary members which escapes. Such orbits were calculated already in 1920 (Becker 1920). Sometimes the third body is always well separated from the binary; then the situation is best described by perturbations on the binary caused by the third body. Some examples of these systems are discussed in Chapter 10.
In recent years there has been increasing demand for solutions of the general three-body problem in various astrophysical situations. For example, binaries and their interactions with single stars play a major role in the evolution of star clusters (Aarseth 1973). Triple stellar systems are another obvious astrophysical three-body
1.2 The three-body problem in astrophysics

For more than 300 years there have been many different motivations to solve the three-body problem and many different techniques have been applied to it. In this book we have the rather limited purpose of looking at solutions of astrophysical significance. At the present time we can solve any given three-body problem, starting from the known positions, velocities and masses of the three bodies, by using a computer. There is of course the limitation of the accuracy of calculation which may be quite significant in some cases. But notwithstanding the accuracy, the solution of an astrophysical problem usually involves much more than a calculation of a single orbit. Typically we have to sample three-body orbits in a phase space of up to eleven dimensions. Then the calculation of orbits is only one tool; one has to have a deeper understanding of the three-body process to make sense of the limited amount of information that is derivable from orbit calculations.

Therefore we do not deal with the mathematical three-body problem. Fortunately, there are excellent books by Marchal (1990) and Hénon (1997, 2001) which deal with the mathematical aspects very thoroughly. As an example, periodic orbits of the general three-body problem are of great mathematical interest, but there are very few examples where they are important in astrophysics.

In problems of astrophysical importance, one may almost always identify a binary and a third body. A binary can be treated as a single entity with certain ‘internal’ properties (like a molecule). It is described by its component masses, by its energy and angular momentum, as well as by its orientation in space. This binary entity interacts with a third body once, or more frequently, which changes the internal properties of the binary. At the same time, the third body absorbs whatever energy is given out from the binary, in order to conserve the total energy. Similarly, conservation of angular momentum between the binary and the third body has to be satisfied.

Before we can take up the discussion of the three-body problem, we have to be familiar with binaries, i.e. the two-body problem. The two-body problem is treated in basic courses of mechanics and celestial mechanics. Therefore the discussion of Chapter 3 may appear as unnecessary repetition to some readers, and they
Astrophysics and the three-body problem

may want to skip much of the first chapters. Also the Hamiltonian techniques of Chapter 4 are commonly studied in courses of advanced mechanics. They form such an essential part of the treatment of weak perturbations of binaries that it is necessary to introduce the Hamiltonian concepts also in this book.

A planet going around the Sun is an example of a binary. The third body could then be another planet, a moon, an asteroid or a comet. Because of the large differences between the masses of the bodies, from the dominant Sun down to asteroids of kilometre size or even smaller, Solar System dynamics is special in many ways. A very up to date treatment exists in this area (Murray and Dermott 1999). We discuss only a small class of Solar System problems which are related to stellar dynamics and therefore form a suitable introduction to later studies of three-body systems with more equal masses. Thus Chapter 5 contains many topics which readers may have encountered earlier.

At the present time, three-body astrophysics is primarily motivated by the need to understand the role of binaries in the evolution of stellar systems. For most of the time, a binary acts just like a single star in a stellar system. The distances between stars are large compared with the sizes of the stars and even compared with the sizes of close binary orbits. For a relatively brief moment a binary and a third star interact strongly, a ‘new’ binary forms, and a ‘new’ third star leaves the scene. The importance of the process lies in its ability to redistribute energy and angular momentum efficiently; the population of binaries may become more and more tightly bound as time and three-body scattering go on, while the population of single stars may gain speed and become ‘heated’. This will have profound consequences on the structure and evolution of a star cluster; for example single stars and sometimes also binaries escape from the cluster; binary orbits may shrink to form contact binaries, and also triple stars may form where the third body remains bound to the binary. The end products of the three-body process may appear as sources of radio jets, X-rays, gamma rays or as other kinds of ‘exotic’ objects (Hut et al. 2003).

These various scenarios can be reproduced by numerical orbit calculations. The orbits of thousands of stars can be calculated in a simulation of a star cluster. Even though these simulations have now reached a great level of complication and trustworthiness (e.g. Heggie and Hut 2003, Aarseth 2003a), it is still useful to examine the three-body process to see how much is understood from elementary principles. Together with the simulations of large numbers of bodies one may attain a deeper understanding of the evolutionary process.

Chapter 7 starts with the discussion of initially strongly interacting three-body systems. We will learn that such systems have a limited lifetime, and we do not expect to find very many of them in nature. But they are important in the description of the intermediate state between the impact of a third body on a binary and
1.2 The three-body problem in astrophysics

the departure of a ‘new’ third body from a ‘new’ binary. The techniques used in Chapter 7, assuming a complete reshuffle of the positions and velocities of the three bodies in the phase space, work surprisingly well.

By ‘work well’ we mean that a large body of numerical orbit calculations can be described in a statistical sense by simple physical principles. That this should be so is not at all obvious at the outset. Therefore we devote considerable space to comparing numerical calculations with the theory. The beauty of it is that once the theory is established for certain parts of the parameter space, we have good reasons to expect that it applies more generally. Also we do not need to go back to calculating millions of orbits when a slightly different astrophysical problem arises but we can use the theory directly with a fair amount of confidence.

In addition to these practical considerations, it gives the reader a certain satisfaction to learn that simple analytical solutions of the three-body problem exist, even though only in a statistical sense and for a limited part of the parameter space. It will also become clear that these are the only solutions of any significance in large parts of the phase space due to the chaotic nature of the problem.

Chapters 9 and 10 try to cover the remaining parameter space, i.e. when the binary and the third body are so well separated throughout the interaction that the principle of complete chaos is not productive. At one extreme there is the very slow and gentle perturbation of the binary known as the Kozai mechanism. Then only the binary eccentricity and inclination change periodically while the orbital sizes are unaffected. At the other extreme we have a high speed intruder which gives the binary a ‘shock treatment’ during its brief encounter with the binary. In between, a binary is strongly perturbed at close encounters but is able to maintain its identity and not break up or exchange members. A stability boundary is derived which tells us where the perturbation treatment ends and the chaos theory begins.

In this way we can give a rather complete coverage of the astrophysical three-body problem. As in the case of the chaos theory of Chapters 7 and 8, also in the perturbation theory of Chapters 9 and 10, a great deal of space is dedicated to comparing numerical results with the theory. This is necessary since it is not always obvious, in the absence of exact theory, what approximations should lead to the best understanding of the experiments. Often we even find that in the final steps we just have to accept the guidance of the numerical experiments without clear justification of the theory. This is not because the theories could not be pushed any further but more because we like to keep the theory at a rather simple level (and it may appear quite complicated to some readers already as it is). But also we have to remember that in the general three-body problem with strong interactions no exact theory exists, and we should not spend too much effort towards this elusive goal.

Throughout the development of the theory we will look at some small astrophysical problems which are easily solved at this stage. In the final chapter a couple of
larger issues are discussed which require solutions from different parts of the text as well as other astrophysical information. To give the reader an idea of what sorts of problems we are dealing with we outline a couple of astrophysical examples in the next sections.

1.3 Short period comets

The origin of short period comets is one of the oldest three-body problems. Lexell (1778, 1779) studied the motion of the comet found in 1770 by Messier, and suggested that the orbit had become elliptical with a period of 5.6 years, when the comet passed close to Jupiter. Later, Laplace (1799–1825, 1805) and Leverrier worked on the capture hypothesis. Tisserand (1889) and H. A. Newton (1891), among others, discussed details of the capture process. The idea, already put forward by Lexell, was that short period comets are created from long period comets which pass near Jupiter and lose energy during the encounter. Everhart (1969) carried out a major survey of close encounter orbits between comets and planets using computers. In spite of all these and later efforts, the origin of the short period comets is still an open question to a large extent.

There are more than two hundred known short period comets, even though it is well documented that comets fade away after $10^2$–$10^3$ revolutions around the Sun and that they escape the solar system after $10^4$–$10^5$ revolutions (Fig. 1.3). There must be a source of comets which constantly (or from time to time) replenishes the population. Two such sources have been suggested: the Oort Cloud of comets and the Kuiper Belt of asteroids and comets. Sometimes also other source regions, such as interstellar comets, have been mentioned. The processes which may keep the short period comet population intact are mostly related to the three-body problem.

The Oort Cloud (Oort 1950) is a collection of as many as $10^{12}$ comets loosely bound to the Solar System. The orbits of the comets are such that they generally do not enter the planetary region, and their orbits are mainly affected by the Sun, passing stars, gas clouds and the tidal field of the Galaxy as a whole. As a result of these influences, there is a more or less steady flux of ‘new’ comets which enter the planetary system for the first time. The flux may be rather uniform per pericentre interval (closest approach distance to the Sun, in AU) up to the distance of Jupiter; beyond that the flux is expected to rise but the difficulty of observing distant comet passages prevents observational confirmation of this expected trend.

Since we have some idea of the Oort Cloud flux of comets we may ask how much of this flux is captured to short period orbits and what are the orbital properties of the captured comets. An Oort Cloud comet comes in a highly eccentric orbit, and in the absence of planets it would return back to the Oort Cloud in the same elliptical orbit. But the orbit is influenced by one or more planets, and this influence is likely
1.3 Short period comets

Figure 1.3 Halley’s comet, the most famous periodic comet, photographed in May 13, 1910. The big round object is Venus and the stripes are city lights of Flagstaff. Image Lowell Observatory.

to change the orbit either to a hyperbolic escape orbit or to a more strongly bound short period elliptical orbit. Basically we need to solve the three-body problem consisting of the Sun, a planet and a comet. Since the comet has much smaller mass than the other two bodies, the problem is restricted only to the question of the motion of the comet. Generally we may assume that the planet goes around the Sun in a circular orbit, and remains in this orbit independent of what happens to the comet. This is an example of the restricted circular three-body problem which will be discussed in Chapter 5.

It is a straightforward procedure to make use of an orbit integrator, a computer code which calculates orbits, and to calculate the orbit of an Oort Cloud comet through the Solar System, past various planets, perhaps through millions of orbital cycles, until the comet escapes from the Solar System, or until it collides with the Sun or one of the planets, or until the comet disintegrates. The calculation can be long but it is possible with modern computers. However, questions remain: how representative is this orbit and how accurate is the solution? Indeed, we do not know the exact starting conditions for any Oort Cloud comets and neither can we carry out the calculation over the orbital time of millions of years without significant loss of accuracy.

To some extent these problems can be avoided by using the method of sampling. We take a large sample of orbits with different initial conditions and calculate them through the necessary period of time. We may then look at observational samples
and compare them with samples obtained by orbit calculation. In practice we need millions of orbits in order to get satisfactory statistics of the captured short period comets. It becomes a major computational challenge.

We can learn quite a lot by studying the interaction of a comet with only one of the planets. A typical Oort Cloud comet has a high inclination relative to the plane of the Solar System. It comes from one side of the plane (say, above), dives through the plane, turns around below the plane, and crosses the plane again. A strong three-body interaction takes place only if a planet happens to be close to one of the crossing points at the right time. Most likely no planet is there at all, but when a close encounter happens, almost always it is only with one planet only at one of the two crossing points. Therefore we have a three-body problem.

But even then the exact nature of the three-body encounter is unclear. The best we can do is to develop a statistical theory of how the comet is likely to react to the presence of the planet close to the crossing point. Ōpik (1951) developed such a theory where the comet is assumed to follow an exact two-body orbit around the Sun until it comes to the sphere of gravitational influence of the planet. At this point the comet starts to follow an exact two-body orbit relative to the planet. After leaving the sphere of influence of the planet, the orbit is again a (different) two-body orbit around the Sun. A theory along these lines will be discussed in Chapter 6, together with some more recent work. It is the most basic form of a solution to the three-body problem. Notice that we will be discussing probabilities; this is the recurring theme of the solutions of astrophysical three-body problems.

The statistical properties of comets which we will have to confront are primarily the distribution of their orbital sizes (semi-major axes), their perihelion distances and the distribution of orbital inclinations relative to the Solar System plane. In terms of orbital sizes, the comets can be classified as being either Jupiter family comets (orbital period below 20 years), Halley type comets (period above 20 but below 200 years) or long period comets (longer periods than 200 years). Oort Cloud comets have orbital periods in excess of a million years. Within these groups, the inclination distributions vary (Fig. 1.4). One of the aims of a successful theory of short period comets is to explain their observed numbers (28 Halley type, 183 Jupiter family comets; Marsden and Williams 1999) in relation to the rate of comets coming from the Oort Cloud. Also it should explain how the differences in the inclination distributions originate.

If it turns out that a decrease in the semi-major axis of a comet is associated with a decrease in its inclination (and this will be shown in Chapter 6) then we might propose that comets evolve from Oort Cloud comets to Halley type comets to Jupiter family comets by successively decreasing the semi-major axis and the inclination of the orbit. How this works out in detail will be discussed in Chapter 11.