Contents

Chapter 1

The two-body problem

1.1 Introduction

The roots of celestial mechanics are two fundamental discoveries by Isaac Newton. First, in any inertial frame the acceleration of a body of mass m subjected to a force $\mathbf{F}$ is

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \frac{\mathbf{F}}{m}. \tag{1.1}$$

Second, the gravitational force exerted by a point mass m_1 at position $\mathbf{r}_1$ on a point mass m_0 at $\mathbf{r}_0$ is

$$\mathbf{F} = \frac{\mathbb{G}m_0m_1(\mathbf{r}_1 - \mathbf{r}_0)}{|\mathbf{r}_1 - \mathbf{r}_0|^3}, \tag{1.2}$$

with $\mathbb{G}$ the gravitational constant.[1] Newton's laws have now been superseded by the equations of general relativity but remain accurate enough to describe all observable phenomena in planetary systems when they are supplemented by small relativistic corrections. A summary of the relevant effects of general relativity is given in Appendix J.

The simplest problem in celestial mechanics, solved by Newton but known as the **two-body problem** or the **Kepler problem**, is to determine

[1] For values of this and other constants, see Appendix A.

the orbits of two point masses ("particles") under the influence of their mutual gravitational attraction. This is the subject of the current chapter.[2]

The equations of motion for the particles labeled 0 and 1 are found by combining (1.1) and (1.2),

$$\frac{\mathrm{d}^2\mathbf{r}_0}{\mathrm{d}t^2} = \frac{\mathbb{G}m_1(\mathbf{r}_1 - \mathbf{r}_0)}{|\mathbf{r}_1 - \mathbf{r}_0|^3}, \qquad \frac{\mathrm{d}^2\mathbf{r}_1}{\mathrm{d}t^2} = \frac{\mathbb{G}m_0(\mathbf{r}_0 - \mathbf{r}_1)}{|\mathbf{r}_0 - \mathbf{r}_1|^3}. \tag{1.3}$$

The total energy and angular momentum of the particles are

$$E_{\mathrm{tot}} = \tfrac{1}{2}m_0|\dot{\mathbf{r}}_0|^2 + \tfrac{1}{2}m_1|\dot{\mathbf{r}}_1|^2 - \frac{\mathbb{G}m_0m_1}{|\mathbf{r}_1 - \mathbf{r}_0|},$$
$$\mathbf{L}_{\mathrm{tot}} = m_0\mathbf{r}_0 \times \dot{\mathbf{r}}_0 + m_1\mathbf{r}_1 \times \dot{\mathbf{r}}_1, \tag{1.4}$$

in which we have introduced the notation $\dot{\mathbf{r}} = \mathrm{d}\mathbf{r}/\mathrm{d}t$. Using equations (1.3) it is straightforward to show that the total energy and angular momentum are conserved, that is, $\mathrm{d}E_{\mathrm{tot}}/\mathrm{d}t = 0$ and $\mathrm{d}\mathbf{L}_{\mathrm{tot}}/\mathrm{d}t = 0$.

We now change variables from $\mathbf{r}_0$ and $\mathbf{r}_1$ to

$$\mathbf{r}_{\mathrm{cm}} \equiv \frac{m_0\mathbf{r}_0 + m_1\mathbf{r}_1}{m_0 + m_1}, \qquad \mathbf{r} \equiv \mathbf{r}_1 - \mathbf{r}_0; \tag{1.5}$$

here $\mathbf{r}_{\mathrm{cm}}$ is the **center of mass** or **barycenter** of the two particles and $\mathbf{r}$ is the **relative position**. These equations can be solved for $\mathbf{r}_0$ and $\mathbf{r}_1$ to yield

$$\mathbf{r}_0 = \mathbf{r}_{\mathrm{cm}} - \frac{m_1}{m_0 + m_1}\mathbf{r}, \qquad \mathbf{r}_1 = \mathbf{r}_{\mathrm{cm}} + \frac{m_0}{m_0 + m_1}\mathbf{r}. \tag{1.6}$$

Taking two time derivatives of the first of equations (1.5) and using equations (1.3), we obtain

$$\frac{\mathrm{d}^2\mathbf{r}_{\mathrm{cm}}}{\mathrm{d}t^2} = \mathbf{0}; \tag{1.7}$$

[2] Most of the basic material in the first part of this chapter can be found in textbooks on classical mechanics. The more advanced material in later sections and chapters has been treated in many books over more than two centuries. The most influential of these include Laplace (1799–1825), Tisserand (1889–1896), Poincaré (1892–1897), Plummer (1918), Brouwer & Clemence (1961) and Murray & Dermott (1999).

thus the center of mass travels at uniform velocity, a consequence of the absence of any external forces.

In these variables the total energy and angular momentum can be written

$$E_{\rm tot} = E_{\rm cm} + E_{\rm rel}, \quad \mathbf{L}_{\rm tot} = \mathbf{L}_{\rm cm} + \mathbf{L}_{\rm rel}, \tag{1.8}$$

where

$$\begin{aligned} E_{\rm cm} &= \tfrac{1}{2}M|\dot{\mathbf{r}}_{\rm cm}|^2, & \mathbf{L}_{\rm cm} &= M\mathbf{r}_{\rm cm}\times\dot{\mathbf{r}}_{\rm cm}, \\ E_{\rm rel} &= \tfrac{1}{2}\mu|\dot{\mathbf{r}}|^2 - \frac{\mathbb{G}\mu M}{|\mathbf{r}|}, & \mathbf{L}_{\rm rel} &= \mu\,\mathbf{r}\times\dot{\mathbf{r}}; \end{aligned} \tag{1.9}$$

here we have introduced the **reduced mass** and **total mass**

$$\mu \equiv \frac{m_0 m_1}{m_0 + m_1}, \quad M \equiv m_0 + m_1. \tag{1.10}$$

The terms $E_{\rm cm}$ and $\mathbf{L}_{\rm cm}$ are the kinetic energy and angular momentum associated with the motion of the center of mass. These are zero if we choose a reference frame in which the velocity of the center of mass $\dot{\mathbf{r}}_{\rm cm} = \mathbf{0}$. The terms $E_{\rm rel}$ and $\mathbf{L}_{\rm rel}$ are the energy and angular momentum associated with the relative motion of the two particles around the center of mass. These are the same as the energy and angular momentum of a particle of mass μ orbiting around a mass M (the "central body") that is fixed at the origin of the vector $\mathbf{r}$.

Taking two time derivatives of the second of equations (1.5) yields

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\frac{\mathbb{G}M}{r^3}\mathbf{r} = -\frac{\mathbb{G}M}{r^2}\hat{\mathbf{r}}, \tag{1.11}$$

where $r = |\mathbf{r}|$ and the unit vector $\hat{\mathbf{r}} = \mathbf{r}/r$. Equation (1.11) describes any one of the following:

(i) the motion of a particle of arbitrary mass subject to the gravitational attraction of a central body of mass M that is fixed at the origin;

(ii) the motion of a particle of negligible mass (a **test particle**) under the influence of a freely moving central body of mass M;

(iii) the motion of a particle with mass equal to the reduced mass μ around a fixed central body that attracts it with the force $\mathbf{F}$ of equation (1.2).

Whatever the interpretation, the two-body problem has been reduced to a one-body problem.

Equation (1.11) can be derived from a Hamiltonian, as described in §1.4. It can also be written

$$\ddot{\mathbf{r}} = -\nabla\Phi_{\mathrm{K}}, \tag{1.12}$$

where we have introduced the notation $\nabla f(\mathbf{r})$ for the gradient of the scalar function $f(\mathbf{r})$ (see §B.3 for a review of vector calculus). The function $\Phi_{\mathrm{K}}(r) = -\mathbb{G}M/r$ is the **Kepler potential**. The solution of equations (1.11) or (1.12) is known as the **Kepler orbit**.

We begin the solution of equation (1.11) by evaluating the rate of change of the relative angular momentum $\mathbf{L}_{\mathrm{rel}}$ from equation (1.9):

$$\frac{1}{\mu}\frac{\mathrm{d}\mathbf{L}_{\mathrm{rel}}}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + \mathbf{r} \times \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\frac{\mathbb{G}M}{r^2}\mathbf{r} \times \hat{\mathbf{r}} = \mathbf{0}. \tag{1.13}$$

Thus the relative angular momentum is conserved. Moreover, since $\mathbf{L}_{\mathrm{rel}}$ is normal to the plane containing the test particle's position and velocity vectors, the position vector must remain in a fixed plane, the **orbital plane**. The plane of the Earth's orbit around the Sun is called the **ecliptic**, and the directions perpendicular to this plane are called the north and south ecliptic poles.

We now simplify our notation. Since we can always choose an inertial reference frame in which the center-of-mass angular momentum $\mathbf{L}_{\mathrm{cm}} = \mathbf{0}$ for all time, we usually shorten "relative angular momentum" to "angular momentum." Similarly the "relative energy" E_{rel} is shortened to "energy." We usually work with the angular momentum per unit mass $\mathbf{L}_{\mathrm{rel}}/\mu = \mathbf{r} \times \dot{\mathbf{r}}$ and the energy per unit mass $\frac{1}{2}|\dot{\mathbf{r}}|^2 - \mathbb{G}M/|\mathbf{r}|$. These may be called "specific angular momentum" and "specific energy," but we shall just write "angular momentum" or "energy" when the intended meaning is clear. Moreover we typically use the same symbol—$\mathbf{L}$ for angular momentum and E for energy—whether we are referring to the total quantity or the quantity per unit mass. This casual use of the same notation for two different physical quantities is less dangerous than it may seem, because the intended meaning can always be deduced from the units of the equations.

1.2 The shape of the Kepler orbit

We let (r, ψ) denote polar coordinates in the orbital plane, with ψ increasing in the direction of motion of the orbit. If $\mathbf{r}$ is a vector in the orbital plane, then $\mathbf{r} = r\hat{\mathbf{r}}$ where $(\hat{\mathbf{r}}, \hat{\boldsymbol{\psi}})$ are unit vectors in the radial and azimuthal directions. The acceleration vector lies in the orbital plane and is given by equation (B.18),

$$\ddot{\mathbf{r}} = (\ddot{r} - r\dot{\psi}^2)\hat{\mathbf{r}} + (2\dot{r}\dot{\psi} + r\ddot{\psi})\hat{\boldsymbol{\psi}}, \tag{1.14}$$

so the equations of motion (1.12) become

$$\ddot{r} - r\dot{\psi}^2 = -\frac{\mathrm{d}\Phi_\mathrm{K}(r)}{\mathrm{d}r}, \quad 2\dot{r}\dot{\psi} + r\ddot{\psi} = 0. \tag{1.15}$$

The second equation may be multiplied by r and integrated to yield

$$r^2\dot{\psi} = \text{constant} = L, \tag{1.16}$$

where $L = |\mathbf{L}|$. This is just a restatement of the conservation of angular momentum, equation (1.13).

We may use equation (1.16) to eliminate $\dot{\psi}$ from the first of equations (1.15),

$$\ddot{r} - \frac{L^2}{r^3} = -\frac{\mathrm{d}\Phi_\mathrm{K}}{\mathrm{d}r}. \tag{1.17}$$

Multiplying by $\dot{r}$ and integrating yields

$$\tfrac{1}{2}\dot{r}^2 + \frac{L^2}{2r^2} + \Phi_\mathrm{K}(r) = E, \tag{1.18}$$

where E is a constant that is equal to the energy per unit mass of the test particle. Equation (1.18) can be rewritten as

$$\tfrac{1}{2}v^2 - \frac{\mathbb{G}M}{r} = E, \tag{1.19}$$

where $v = (\dot{r}^2 + r^2\dot{\psi}^2)^{1/2}$ is the speed of the test particle.

Equation (1.18) implies that

$$\dot{r}^2 = 2E + \frac{2\mathbb{G}M}{r} - \frac{L^2}{r^2}. \tag{1.20}$$

As $r \to 0$, the right side approaches $-L^2/r^2$, which is negative, while the left side is positive. Thus there must be a point of closest approach of the test particle to the central body, known as the **periapsis** or **pericenter**.[3] In the opposite limit, $r \to \infty$, the right side of equation (1.20) approaches $2E$. Since the left side is positive, when $E < 0$ there is a maximum distance that the particle can achieve, known as the **apoapsis** or **apocenter**. Orbits with $E < 0$ are referred to as **bound** orbits since there is an upper limit to their distance from the central body. Orbits with $E > 0$ are **unbound** or **escape** orbits; they have no apoapsis, and particles on such orbits eventually travel arbitrarily far from the central body, never to return.[4]

The periapsis distance q and apoapsis distance Q of an orbit are determined by setting $\dot{r} = 0$ in equation (1.20), which yields the quadratic equation

$$2Er^2 + 2\mathbb{G}Mr - L^2 = 0. \tag{1.22}$$

For bound orbits, $E < 0$, there are two roots on the positive real axis,

$$q = \frac{\mathbb{G}M - \left[(\mathbb{G}M)^2 + 2EL^2\right]^{1/2}}{2|E|}, \quad Q = \frac{\mathbb{G}M + \left[(\mathbb{G}M)^2 + 2EL^2\right]^{1/2}}{2|E|}. \tag{1.23}$$

For unbound orbits, $E > 0$, there is only one root on the positive real axis,

$$q = \frac{\left[(\mathbb{G}M)^2 + 2EL^2\right]^{1/2} - \mathbb{G}M}{2E}. \tag{1.24}$$

[3] For specific central bodies other names are used, such as perihelion (Sun), perigee (Earth), periastron (a star), and so forth. "Periapse" is incorrect—an apse is not an apsis.

[4] The **escape speed** $v_{\rm esc}$ from an object is the minimum speed needed for a test particle to escape from its surface; if the object is spherical, with mass M and radius R, equation (1.19) implies that

$$v_{\rm esc} = \left(\frac{2\mathbb{G}M}{R}\right)^{1/2}. \tag{1.21}$$

To solve the differential equation (1.17) we introduce the variable $u \equiv 1/r$, and change the independent variable from t to ψ using the relation

$$\frac{\mathrm{d}}{\mathrm{d}t} = \dot{\psi}\frac{\mathrm{d}}{\mathrm{d}\psi} = Lu^2\frac{\mathrm{d}}{\mathrm{d}\psi}. \qquad (1.25)$$

With these substitutions, $\dot{r} = -L\mathrm{d}u/\mathrm{d}\psi$ and $\ddot{r} = -L^2u^2\mathrm{d}^2u/\mathrm{d}\psi^2$, so equation (1.17) becomes

$$\frac{\mathrm{d}^2u}{\mathrm{d}\psi^2} + u = -\frac{1}{L^2}\frac{\mathrm{d}\Phi_\mathrm{K}}{\mathrm{d}u}. \qquad (1.26)$$

Since $\Phi_\mathrm{K}(r) = -\mathbb{G}M/r = -\mathbb{G}Mu$ the right side is equal to a constant, $\mathbb{G}M/L^2$, and the equation is easily solved to yield

$$u = \frac{1}{r} = \frac{\mathbb{G}M}{L^2}[1 + e\cos(\psi - \varpi)], \qquad (1.27)$$

where $e \geq 0$ and ϖ are constants of integration.[5] We replace the angular momentum L by another constant of integration, a, defined by the relation

$$L^2 = \mathbb{G}Ma(1 - e^2), \qquad (1.28)$$

so the shape of the orbit is given by

$$r = \frac{a(1 - e^2)}{1 + e\cos f}, \qquad (1.29)$$

where $f = \psi - \varpi$ is known as the **true anomaly**.[6]

The closest approach of the two bodies occurs at $f = 0$ or azimuth $\psi = \varpi$ and hence ϖ is known as the **longitude of periapsis**. The periapsis distance is $r(f = 0)$ or

$$q = a(1 - e). \qquad (1.30)$$

---

[5] The symbol ϖ is a variant of the symbol for the Greek letter π, even though it looks more like the symbol for the letter ω; hence it is sometimes informally called "pomega."

[6] In a subject as old as this, there is a rich specialized vocabulary. The term "anomaly" refers to any angular variable that is zero at periapsis and increases by 2π as the particle travels from periapsis to apoapsis and back. There are also several old terms we shall not use: "semilatus rectum" for the combination $a(1 - e^2)$, "vis viva" for kinetic energy, and so on.

When the eccentricity is zero, the longitude of periapsis ϖ is undefined. This indeterminacy can drastically slow or halt numerical calculations that follow the evolution of the orbital elements, and can be avoided by replacing e and ϖ by two new elements, the **eccentricity components** or h **and** k **variables**

$$k \equiv e \cos \varpi, \quad h \equiv e \sin \varpi, \tag{1.31}$$

which are well defined even for $e = 0$. The generalization to nonzero inclination is given in equations (1.71).

Substituting q for r in equation (1.22) and replacing L^2 using equation (1.28) reveals that the energy per unit mass is simply related to the constant a:

$$E = -\frac{\mathbb{G}M}{2a}. \tag{1.32}$$

First consider bound orbits, which have $E < 0$. Then $a > 0$ by equation (1.32) and hence $e < 1$ by equation (1.28). A circular orbit has $e = 0$ and angular momentum per unit mass $L = (\mathbb{G}Ma)^{1/2}$. The circular orbit has the largest possible angular momentum for a given semimajor axis or energy, so we sometimes write

$$\mathbf{j} \equiv \frac{\mathbf{L}}{(\mathbb{G}Ma)^{1/2}}, \quad \text{where} \quad j = |\mathbf{j}| = (1 - e^2)^{1/2} \tag{1.33}$$

ranges from 0 to 1 and represents a dimensionless angular momentum at a given semimajor axis.

The apoapsis distance, obtained from equation (1.29) with $f = \pi$, is

$$Q = a(1 + e). \tag{1.34}$$

The periapsis and the apoapsis are joined by a straight line known as the **line of apsides**. Equation (1.29) describes an ellipse with one focus at the origin (**Kepler's first law**). Its major axis is the line of apsides and has length $q + Q = 2a$; thus the constant a is known as the **semimajor axis**. The **semiminor axis** of the ellipse is the maximum perpendicular distance of the orbit from the line of apsides, $b = \max_f[a(1 - e^2)\sin f/(1 + e\cos f)] = a(1 - e^2)^{1/2}$. The **eccentricity** of the ellipse, $(1 - b^2/a^2)^{1/2}$, is therefore equal to the constant e.

Box 1.1: The eccentricity vector

The **eccentricity vector** offers a more elegant but less transparent derivation of the equation for the shape of a Kepler orbit. Take the cross product of $\mathbf{L}$ with equation (1.11),

$$\mathbf{L}\times\ddot{\mathbf{r}} = -\frac{\mathbb{G}M}{r^3}\mathbf{L}\times\mathbf{r}. \tag{a}$$

Using the vector identity (B.9b), $\mathbf{L}\times\mathbf{r} = -\mathbf{r}\times\mathbf{L} = -\mathbf{r}\times(\mathbf{r}\times\dot{\mathbf{r}}) = r^2\dot{\mathbf{r}} - (\mathbf{r}\cdot\dot{\mathbf{r}})\mathbf{r}$, which is equal to $r^3\mathrm{d}\hat{\mathbf{r}}/\mathrm{d}t$. Thus

$$\mathbf{L}\times\ddot{\mathbf{r}} = -\mathbb{G}M\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t}. \tag{b}$$

Since $\mathbf{L}$ is constant, we may integrate to obtain

$$\mathbf{L}\times\dot{\mathbf{r}} = -\mathbb{G}M(\hat{\mathbf{r}}+\mathbf{e}), \tag{c}$$

where $\mathbf{e}$ is a vector constant of motion, the **eccentricity vector**. Rearranging equation (c), we have

$$\mathbf{e} = \frac{\dot{\mathbf{r}}\times(\mathbf{r}\times\dot{\mathbf{r}})}{\mathbb{G}M} - \frac{\mathbf{r}}{r}. \tag{d}$$

To derive the shape of the orbit, we take the dot product of (c) with $\hat{\mathbf{r}}$ and use the vector identity (B.9a) to show that $\hat{\mathbf{r}}\cdot(\mathbf{L}\times\dot{\mathbf{r}}) = -L^2/r$. The resulting formula is

$$r = \frac{L^2}{\mathbb{G}M}\frac{1}{1+\mathbf{e}\cdot\hat{\mathbf{r}}} = \frac{a(1-e^2)}{1+\mathbf{e}\cdot\hat{\mathbf{r}}}; \tag{e}$$

in the last equation we have eliminated L^2 using equation (1.28). This result is the same as equation (1.29) if the magnitude of the eccentricity vector equals the eccentricity, $|\mathbf{e}| = e$, and the eccentricity vector points toward periapsis.

The eccentricity vector is often called the **Runge–Lenz vector**, although its history can be traced back at least to Laplace (Goldstein 1975–1976). Runge and Lenz appear to have taken their derivation from Gibbs & Wilson (1901), the classic text that introduced modern vector notation.

Unbound orbits have $E > 0$, $a < 0$ and $e > 1$. In this case equation (1.29) describes a hyperbola with focus at the origin and asymptotes at azimuth

$$\psi = \varpi \pm f_\infty, \quad \text{where} \quad f_\infty \equiv \cos^{-1}(-1/e) \tag{1.35}$$

is the **asymptotic true anomaly**, which varies between π (for $e = 1$) and $\frac{1}{2}\pi$ (for $e \to \infty$). The constants a and e are still commonly referred to as semimajor axis and eccentricity even though these terms have no direct geometric interpretation.

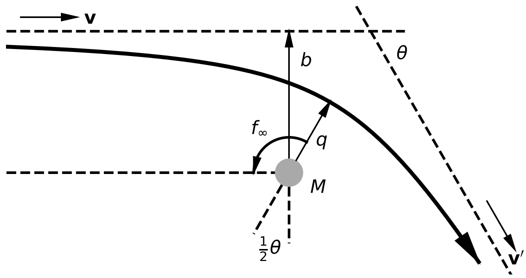

Figure 1.1: The geometry of an unbound or hyperbolic orbit around mass M. The impact parameter is b, the deflection angle is θ, the asymptotic true anomaly is f_∞, and the periapsis is located at the tip of the vector $\mathbf{q}$.

Suppose that a particle is on an unbound orbit around a mass M. Long before the particle approaches M, it travels at a constant velocity which we denote by $\mathbf{v}$ (Figure 1.1). If there were no gravitational forces, the particle would continue to travel in a straight line that makes its closest approach to M at a point $\mathbf{b}$ called the **impact parameter vector**. Long after the particle passes M, it again travels at a constant velocity $\mathbf{v}'$, where $v \equiv |\mathbf{v}| = |\mathbf{v}'|$ because of energy conservation. The deflection angle θ is the angle between $\mathbf{v}$ and $\mathbf{v}'$, given by $\cos\theta = \mathbf{v} \cdot \mathbf{v}'/v^2$. The deflection angle is related to the asymptotic true anomaly f_∞ by $\theta = 2f_\infty - \pi$; then

$$\tan\tfrac{1}{2}\theta = -\frac{\cos f_\infty}{\sin f_\infty} = \frac{1}{(e^2-1)^{1/2}}. \tag{1.36}$$

The relation between the pre- and post-encounter velocities can be written

$$\mathbf{v}' = \mathbf{v}\cos\theta - \hat{\mathbf{b}}v\sin\theta. \tag{1.37}$$

In many cases the properties of unbound orbits are best described by the asymptotic speed v and the impact parameter $b = |\mathbf{b}|$, rather than by orbital elements such as a and e. It is straightforward to show that the angular momentum and energy of the orbit per unit mass are $L = bv$ and $E = \frac{1}{2}v^2$. From equations (1.28) and (1.32) it follows that

$$a = -\frac{\mathbb{G}M}{v^2}, \quad e^2 = 1 + \frac{b^2v^4}{(\mathbb{G}M)^2}. \tag{1.38}$$

Then from equation (1.36),

$$\tan\tfrac{1}{2}\theta = \frac{\mathbb{G}M}{bv^2}. \tag{1.39}$$

The periapsis distance $q = a(1 - e)$ is related to the impact parameter b by

$$q = \frac{\mathbb{G}M}{v^2}\left[\left(1 + \frac{b^2v^4}{\mathbb{G}^2M^2}\right)^{1/2} - 1\right] \quad \text{or} \quad b^2 = q^2 + \frac{2\,\mathbb{G}Mq}{v^2}. \tag{1.40}$$

Thus, for example, if the central body has radius R, the particle will collide with it if

$$b^2 \le b_{\rm coll}^2 \equiv R^2 + \frac{2\,\mathbb{G}MR}{v^2}. \tag{1.41}$$

The corresponding cross section is $\pi b_{\rm coll}^2$. If the central body has zero mass the cross section is just πR^2; the enhancement arising from the second term in equation (1.41) is said to be due to **gravitational focusing**.

In the special case $E = 0$, a is infinite and $e = 1$, so equation (1.29) is undefined; however, in this case equation (1.22) implies that the periapsis distance $q = L^2/(2\,\mathbb{G}M)$, so equation (1.27) implies

$$r = \frac{2q}{1 + \cos f}, \tag{1.42}$$

which describes a parabola. This result can also be derived from equation (1.29) by replacing $a(1 - e^2)$ by $q(1 + e)$ and letting $e \to 1$.

1.3 Motion in the Kepler orbit

The **period** P of a bound orbit is the time taken to travel from periapsis to apoapsis and back. Since $\mathrm{d}\psi/\mathrm{d}t = L/r^2$, we have $\int_{t_1}^{t_2} \mathrm{d}t = L^{-1} \int_{\psi_1}^{\psi_2} r^2 \mathrm{d}\psi$; the integral on the right side is twice the area contained in the ellipse between azimuths ψ_1 and ψ_2, so the radius vector to the particle sweeps out equal areas in equal times (**Kepler's second law**). Thus $P = 2A/L$, where the area of the ellipse is $A = \pi ab$ with a and $b = a(1-e^2)^{1/2}$ the semimajor and semiminor axes of the ellipse. Combining these results with equation (1.28), we find

$$P = 2\pi \left(\frac{a^3}{\mathbb{G}M} \right)^{1/2}. \tag{1.43}$$

The period, like the energy, depends only on the semimajor axis. The **mean motion** or mean rate of change of azimuth, usually written n and equal to $2\pi/P$, thus satisfies[7]

$$n^2 a^3 = \mathbb{G}M, \tag{1.44}$$

which is **Kepler's third law** or simply **Kepler's law**. If the particle passes through periapsis at $t = t_0$, the dimensionless variable

$$\ell = 2\pi \frac{t - t_0}{P} = n(t - t_0) \tag{1.45}$$

is called the **mean anomaly**. Notice that the mean anomaly equals the true anomaly f when $f = 0, \pi, 2\pi, \ldots$ but not at other phases unless the orbit is circular; similarly, ℓ and f always lie in the same semicircle (0 to π, π to 2π, and so on).

---

[7] The relation $n = 2\pi/P$ holds because Kepler orbits are closed—that is, they return to the same point once per orbit. In more general spherical potentials we must distinguish the **radial period**, the time between successive periapsis passages, from the **azimuthal period** $2\pi/n$. For example, in a harmonic potential $\Phi(r) = \frac{1}{2}\omega^2 r^2$ the radial period is π/ω but the azimuthal period is $2\pi/\omega$. Smaller differences between the radial and azimuthal period arise in perturbed Kepler systems such as multi-planet systems or satellites orbiting a flattened planet (§1.8.3). For the Earth the radial period is called the **anomalistic year**, while the azimuthal period of 365.256 363 d is the **sidereal year**. The anomalistic year is longer than the sidereal year by 0.003 27 d. When we use the term "year" in this book, we refer to the Julian year of exactly 365.25 d (§1.5).

The position and velocity of a particle in the orbital plane at a given time are determined by four **orbital elements**: two specify the size and shape of the orbit, which we can take to be e and a (or e and n, q and Q, L and E, and so forth); one specifies the orientation of the line of apsides (ϖ); and one specifies the location or phase of the particle in its orbit (f, ℓ, or t_0).

The trajectory $[r(t), \psi(t)]$ can be derived by solving the differential equation (1.20) for $r(t)$, then (1.16) for $\psi(t)$. However, there is a simpler method.

First consider bound orbits. Since the radius of a bound orbit oscillates between $a(1-e)$ and $a(1+e)$, it is natural to define a variable $u(t)$, the **eccentric anomaly**, by

$$r = a(1 - e\cos u); \tag{1.46}$$

since the cosine is multivalued, we must add the supplemental condition that u and f always lie in the same semicircle (0 to π, π to 2π, and so on). Thus u increases from 0 to 2π as the particle travels from periapsis to apoapsis and back. The true, eccentric and mean anomalies f, u and ℓ are all equal for circular orbits, and for any bound orbit $f = u = \ell = 0$ at periapsis and π at apoapsis.

Substituting equation (1.46) into the energy equation (1.20) and using equations (1.28) and (1.32) for L^2 and E, we obtain

$$\dot{r}^2 = a^2e^2\sin^2 u\,\dot{u}^2 = -\frac{\mathbb{G}M}{a} + \frac{2\,\mathbb{G}M}{a(1-e\cos u)} - \frac{\mathbb{G}M(1-e^2)}{a(1-e\cos u)^2}, \tag{1.47}$$

which simplifies to

$$(1-e\cos u)^2\dot{u}^2 = \frac{\mathbb{G}M}{a^3} = n^2 = \dot{\ell}^2. \tag{1.48}$$

Since $\dot{u}, \dot{\ell} > 0$ and $u = \ell = 0$ at periapsis, we may take the square root of this equation and then integrate to obtain **Kepler's equation**

$$\ell = u - e\sin u. \tag{1.49}$$

Kepler's equation cannot be solved analytically for u, but many efficient numerical methods of solution are available.

The relation between the true and eccentric anomalies is found by eliminating r from equations (1.29) and (1.46):

$$\cos f = \frac{\cos u - e}{1 - e\cos u}, \qquad \cos u = \frac{\cos f + e}{1 + e\cos f}, \tag{1.50}$$

with the understanding that f and u always lie in the same semicircle. Similarly,

$$\sin f = \frac{(1-e^2)^{1/2}\sin u}{1 - e\cos u}, \quad \sin u = \frac{(1-e^2)^{1/2}\sin f}{1 + e\cos f}, \tag{1.51a}$$

$$\tan \tfrac{1}{2} f = \left(\frac{1+e}{1-e}\right)^{1/2} \tan \tfrac{1}{2} u, \tag{1.51b}$$

$$\exp(\mathrm{i}f) = \frac{\exp(\mathrm{i}u) - \beta}{1 - \beta\exp(\mathrm{i}u)}, \quad \exp(\mathrm{i}u) = \frac{\exp(\mathrm{i}f) + \beta}{1 + \beta\exp(\mathrm{i}f)}, \tag{1.51c}$$

where

$$\beta \equiv \frac{1 - (1-e^2)^{1/2}}{e}. \tag{1.52}$$

If we assume that the periapsis lies on the x-axis of a rectangular coordinate system in the orbital plane, the coordinates of the particle are

$$x = r\cos f = a(\cos u - e), \quad y = r\sin f = a(1-e^2)^{1/2}\sin u. \tag{1.53}$$

The position and velocity of a bound particle at a given time t can be determined from the orbital elements a, e, ϖ and t_0 by the following steps. First compute the mean motion n from Kepler's third law (1.44), then find the mean anomaly ℓ from (1.45). Solve Kepler's equation for the eccentric anomaly u. The radius r is then given by equation (1.46); the true anomaly f is given by equation (1.50); and the azimuth $\psi = f + \varpi$. The radial velocity is

$$v_r = \dot{r} = n\frac{\mathrm{d}r}{\mathrm{d}\ell} = n\frac{\mathrm{d}r/\mathrm{d}u}{\mathrm{d}\ell/\mathrm{d}u} = \frac{nae\sin u}{1 - e\cos u} = \frac{nae\sin f}{(1-e^2)^{1/2}}, \tag{1.54}$$

and the azimuthal velocity is

$$v_\psi = r\dot{\psi} = \frac{L}{r} = na\frac{(1-e^2)^{1/2}}{1 - e\cos u} = na\frac{1 + e\cos f}{(1-e^2)^{1/2}}, \tag{1.55}$$

in which we have used equation (1.28).

For unbound particles, recall that $a < 0$, $e > 1$, and the period is undefined since the particle escapes to infinity. The physical interpretations of the mean anomaly ℓ and mean motion n that led to equations (1.44) and (1.45) no longer apply, but we may *define* these quantities by the relations

$$\ell = n(t - t_0), \quad n^2|a|^3 = \mathbb{G}M. \tag{1.56}$$

Similarly, we define the eccentric anomaly u by

$$r = |a|(e \cosh u - 1). \tag{1.57}$$

The eccentric and mean anomalies increase from 0 to ∞ as the true anomaly increases from 0 to $\cos^{-1}(-1/e)$ (eq. 1.35).

By following the chain of argument in equations (1.47)–(1.49), we may derive the analog of Kepler's equation for unbound orbits,

$$\ell = e \sinh u - u. \tag{1.58}$$

The relation between the true and eccentric anomalies is

$$\cos f = \frac{e - \cosh u}{e \cosh u - 1}, \quad \cosh u = \frac{e + \cos f}{1 + e \cos f}, \tag{1.59a}$$

$$\sin f = \frac{(e^2 - 1)^{1/2} \sinh u}{e \cosh u - 1}, \quad \sinh u = \frac{(e^2 - 1)^{1/2} \sin f}{1 + e \cos f}, \tag{1.59b}$$

$$\tan \tfrac{1}{2} f = \left(\frac{e + 1}{e - 1}\right)^{1/2} \tanh \tfrac{1}{2} u. \tag{1.59c}$$

A more direct but less physical approach to deriving these results is to substitute $u \to \mathrm{i}u$, $\ell \to -\mathrm{i}\ell$ in the analogous expressions for bound orbits.

For parabolic orbits we do not need the eccentric anomaly since the relation between time from periapsis and true anomaly can be determined analytically. Since $\dot{f} = L/r^2$, we can use equation (1.42) to write

$$t - t_0 = \int_0^f \frac{\mathrm{d}f\, r^2}{L} = \left(\frac{8q^3}{\mathbb{G}M}\right)^{1/2} \int_0^f \frac{\mathrm{d}f}{(1 + \cos f)^2}. \tag{1.60}$$

In the last equation we have used the relation $L^2 = 2\mathbb{G}Mq$ for parabolic orbits. Evaluating the integral, we obtain

$$\left(\frac{\mathbb{G}M}{2q^3}\right)^{1/2}(t - t_0) = \tan\tfrac{1}{2}f + \tfrac{1}{3}\tan^3\tfrac{1}{2}f. \tag{1.61}$$

This is a cubic equation for $\tan\frac{1}{2}f$ that can be solved analytically.

1.3.1 Orbit averages

Many applications require the time average of some quantity $X(\mathbf{r},\mathbf{v})$ over one period of a bound Kepler orbit of semimajor axis a and eccentricity e. We call this the **orbit average** of X and use the notation

$$\langle X\rangle = \int_0^{2\pi}\frac{\mathrm{d}\ell}{2\pi}X = \int_0^{2\pi}\frac{\mathrm{d}u}{2\pi}(1 - e\cos u)X, \tag{1.62}$$

in which we have used Kepler's equation (1.49) to derive the second integral. An alternative is to write

$$\langle X\rangle = \int_0^{P}\frac{\mathrm{d}t}{P}X = \int_0^{2\pi}\frac{\mathrm{d}f}{P\dot{f}}X = \frac{1}{PL}\int_0^{2\pi}\mathrm{d}f\, r^2 X; \tag{1.63}$$

here P and $L = r^2\dot{f}$ are the orbital period and angular momentum. Substituting equations (1.28), (1.29) and (1.43) for the angular momentum, orbit shape and period, the last equation can be rewritten as

$$\langle X\rangle = (1 - e^2)^{3/2}\int_0^{2\pi}\frac{\mathrm{d}f}{2\pi}\frac{X}{(1 + e\cos f)^2}. \tag{1.64}$$

Equation (1.62) provides the simplest route to derive such results as

$$\langle a/r\rangle = 1, \tag{1.65a}$$
$$\langle r/a\rangle = 1 + \tfrac{1}{2}e^2, \tag{1.65b}$$
$$\langle (r/a)^2\rangle = 1 + \tfrac{3}{2}e^2, \tag{1.65c}$$
$$\langle (r/a)^2\cos^2 f\rangle = \tfrac{1}{2} + 2e^2, \tag{1.65d}$$

$$\langle (r/a)^2 \sin^2 f \rangle = \tfrac{1}{2} - \tfrac{1}{2}e^2, \tag{1.65e}$$
$$\langle (r/a)^2 \cos f \sin f \rangle = 0. \tag{1.65f}$$

Equation (1.64) gives

$$\langle (a/r)^2 \rangle = (1 - e^2)^{-1/2}, \tag{1.66a}$$
$$\langle (a/r)^3 \rangle = (1 - e^2)^{-3/2}, \tag{1.66b}$$
$$\langle (a/r)^3 \cos^2 f \rangle = \tfrac{1}{2}(1 - e^2)^{-3/2}, \tag{1.66c}$$
$$\langle (a/r)^3 \sin^2 f \rangle = \tfrac{1}{2}(1 - e^2)^{-3/2}, \tag{1.66d}$$
$$\langle (a/r)^3 \sin f \cos f \rangle = 0. \tag{1.66e}$$

Additional orbit averages are given in Problems 1.2 and 1.3.

1.3.2 Motion in three dimensions

So far we have described the motion of a particle in its orbital plane. To characterize the orbit fully we must also specify the spatial orientation of the orbital plane, as shown in Figure 1.2.

We work with the usual Cartesian coordinates (x, y, z) and spherical coordinates (r, θ, ϕ) (see Appendix B.2). We call the plane $z = 0$, corresponding to $\theta = \frac{1}{2}\pi$, the **reference plane**. The **inclination** of the orbital plane to the reference plane is denoted I, with $0 \le I \le \pi$; thus $\cos I = \hat{\mathbf{z}} \cdot \hat{\mathbf{L}}$, where $\hat{\mathbf{z}}$ and $\hat{\mathbf{L}}$ are unit vectors in the direction of the z-axis and the angular-momentum vector. Orbits with $0 \le I \le \frac{1}{2}\pi$ are **direct** or **prograde**; orbits with $\frac{1}{2}\pi < I < \pi$ are **retrograde**.

Any bound Kepler orbit pierces the reference plane at two points known as the **nodes** of the orbit. The particle travels upward ($\dot{z} > 0$) at the **ascending node** and downward at the **descending node**. The azimuthal angle ϕ of the ascending node is denoted Ω and is called the **longitude of the ascending node**. The angle from ascending node to periapsis, measured in the direction of motion of the particle in the orbital plane, is denoted ω and is called the **argument of periapsis**.

An unfortunate feature of these elements is that neither ω nor Ω is defined for orbits in the reference plane ($I = 0$). Partly for this reason, the

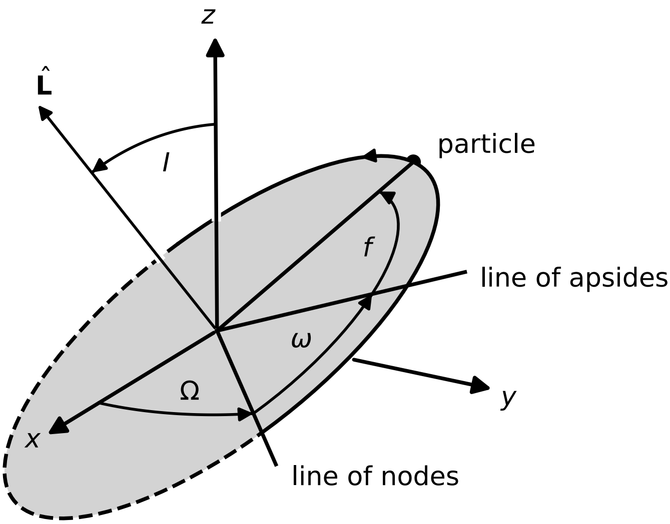

Figure 1.2: The angular elements of a Kepler orbit. The usual Cartesian coordinate axes are denoted by (x, y, z), the reference plane is $z = 0$, and the orbital plane is denoted by a solid curve above the equatorial plane ($z > 0$) and a dashed curve below. The plot shows the inclination I, the longitude of the ascending node Ω, the argument of periapsis ω and the true anomaly f.

argument of periapsis is often replaced by a variable called the **longitude of periapsis** which is defined as

$$\varpi \equiv \Omega + \omega. \tag{1.67}$$

For orbits with zero inclination, the longitude of periapsis has a simple interpretation—it is the azimuthal angle between the x-axis and the periapsis, consistent with our earlier definition of the same symbol following equation (1.29)—but if the inclination is nonzero, it is the sum of two angles measured in different planes (the reference plane and the orbital plane).[8] Despite this awkwardness, for most purposes the three elements (Ω, ϖ, I) provide the most convenient way to specify the orientation of a Kepler orbit.

The **mean longitude** is

$$\lambda \equiv \varpi + \ell = \Omega + \omega + \ell, \tag{1.68}$$

where ℓ is the mean anomaly; like the longitude of perihelion, the mean longitude is the sum of angles measured in the reference plane (Ω) and the orbital plane ($\omega + \ell$).

Some of these elements are closely related to the Euler angles that describe the rotation of one coordinate frame into another (Appendix B.6). Let (x', y', z') be Cartesian coordinates in the **orbital reference frame**, defined such that the z'-axis points along the angular-momentum vector **L** and the x'-axis points toward periapsis, along the eccentricity vector **e**. Then the rotation from the (x, y, z) reference frame to the orbital reference frame is described by the Euler angles

$$(\alpha, \beta, \gamma) = (\Omega, I, \omega). \tag{1.69}$$

The position and velocity of a particle in space at a given time t are specified by six orbital elements: two specify the size and shape of the orbit (e and a); three specify the orientation of the orbit (I, Ω and ω), and one specifies the location of the particle in the orbit (f, u, ℓ, λ, or t_0). Thus, for example, to find the Cartesian coordinates (x, y, z) in terms of the orbital elements, we write the position in the orbital reference frame as $(x', y', z') = r(\cos f, \sin f, 0)$ and use equation (1.69) and the rotation matrix for the transformation from primed to unprimed coordinates (eq. B.61):

$$\begin{aligned}
\frac{x}{r} &= \cos\Omega\cos(f+\omega) - \cos I \sin\Omega \sin(f+\omega),\\
\frac{y}{r} &= \sin\Omega\cos(f+\omega) + \cos I \cos\Omega \sin(f+\omega),\\
\frac{z}{r} &= \sin I \sin(f+\omega);
\end{aligned} \tag{1.70}$$

[8] Thus "longitude of periapsis" is a misnomer, since ϖ is *not* equal to the azimuthal angle of the eccentricity vector, except for orbits of zero inclination.

here r is given in terms of the orbital elements by equation (1.29).

When the eccentricity or inclination is small, the polar coordinate pairs e–ϖ and I–Ω are sometimes replaced by the eccentricity and inclination components[9]

$$k \equiv e\cos\varpi, \quad h \equiv e\sin\varpi, \quad q \equiv \tan I\cos\Omega, \quad p \equiv \tan I\sin\Omega. \tag{1.71}$$

The first two equations are the same as equations (1.31).

For some purposes the shape, size and orientation of an orbit can be described most efficiently using the angular-momentum and eccentricity vectors, $\mathbf{L}$ and $\mathbf{e}$. The two vectors describe five of the six orbital elements: the missing element is the one specifying the location of the particle in its orbit, f, u, ℓ, λ or t_0 (the six components of the two vectors determine only five elements, because $\mathbf{e}$ is restricted to the plane normal to $\mathbf{L}$).

Note that ω and Ω are undefined for orbits with zero inclination; ω and ϖ are undefined for circular orbits; and ϖ, Ω and I are undefined for radial orbits ($e \to 1$). In contrast the angular-momentum and eccentricity vectors are well defined for *all* orbits. The cost of avoiding indeterminacy is redundancy: instead of five orbital elements we need six vector components.

1.3.3 Gauss's *f* and *g* functions

A common task is to determine the position and velocity, $\mathbf{r}(t)$ and $\mathbf{v}(t)$, of a particle in a Kepler orbit given its position and velocity $\mathbf{r}_0$ and $\mathbf{v}_0$ at some initial time t_0. This can be done by converting $\mathbf{r}_0$ and $\mathbf{v}_0$ into the orbital elements $a, e, I, \omega, \Omega, \ell_0$, replacing ℓ_0 by $\ell = \ell_0 + n(t - t_0)$ and then reversing the conversion to determine the position and velocity from the new orbital elements. But there is a simpler method, due to Gauss.

Since the particle is confined to the orbital plane, and $\mathbf{r}_0, \mathbf{v}_0$ are vectors lying in this plane, we can write

$$\mathbf{r}(t) = f(t, t_0)\mathbf{r}_0 + g(t, t_0)\mathbf{v}_0, \tag{1.72}$$

[9] The function $\tan I$ in the elements q and p can be replaced by any function that is proportional to I as $I \to 0$. Various authors use I, $\sin\frac{1}{2}I$, and so forth. The function $\sin I$ is not used because it has the same value for I and $\pi - I$.

which defines **Gauss's f and g functions**. This expression also gives the velocity of the particle,

$$\mathbf{v}(t) = \frac{\partial f(t,t_0)}{\partial t}\mathbf{r}_0 + \frac{\partial g(t,t_0)}{\partial t}\mathbf{v}_0. \tag{1.73}$$

To evaluate f and g for bound orbits we use polar coordinates (r,ψ) and Cartesian coordinates (x,y) in the orbital plane, and assume that $\mathbf{r}_0$ lies along the positive x-axis ($\psi_0 = 0$). Then the components of equation (1.72) along the x- and y-axes are:

$$\begin{aligned} r(t)\cos\psi(t) &= f(t,t_0)r_0 + g(t,t_0)v_r(t_0), \\ r(t)\sin\psi(t) &= g(t,t_0)v_\psi(t_0), \end{aligned} \tag{1.74}$$

where v_r and v_ψ are the radial and azimuthal velocities. These equations can be solved for f and g:

$$\begin{aligned} f(t,t_0) &= \frac{r(t)}{r_0}\left[\cos\psi(t) - \frac{v_r(t_0)}{v_\psi(t_0)}\sin\psi(t)\right], \\ g(t,t_0) &= \frac{r(t)}{v_\psi(t_0)}\sin\psi(t). \end{aligned} \tag{1.75}$$

We use equations (1.16), (1.28), (1.29), (1.54) and the relation $\psi = f - f_0$ to replace the quantities on the right sides by orbital elements (unfortunately f is used to denote both true anomaly and one of Gauss's functions). The result is

$$\begin{aligned} f(t,t_0) &= \frac{\cos(f-f_0) + e\cos f}{1+e\cos f}, \\ g(t,t_0) &= \frac{(1-e^2)^{3/2}\sin(f-f_0)}{n(1+e\cos f)(1+e\cos f_0)}. \end{aligned} \tag{1.76}$$

Since these expressions contain only the orbital elements n, e and f, they are valid in any coordinate system, not just the one we used for the derivation. For deriving velocities from equation (1.73), we need

$$\frac{\partial f(t,t_0)}{\partial t} = n\frac{e\sin f_0 - e\sin f - \sin(f-f_0)}{(1-e^2)^{3/2}},$$

$$\frac{\partial g(t,t_0)}{\partial t} = \frac{e\cos f_0 + \cos(f - f_0)}{1 + e\cos f_0}. \tag{1.77}$$

The f and g functions can also be expressed in terms of the eccentric anomaly, using equations (1.50) and (1.51a):

$$\begin{aligned} f(t,t_0) &= \frac{\cos(u-u_0) - e\cos u_0}{1 - e\cos u_0}, \\ g(t,t_0) &= \frac{1}{n}[\sin(u-u_0) - e\sin u + e\sin u_0], \\ \frac{\partial f(t,t_0)}{\partial t} &= -\frac{n\sin(u-u_0)}{(1-e\cos u)(1-e\cos u_0)}, \\ \frac{\partial g(t,t_0)}{\partial t} &= \frac{\cos(u-u_0) - e\cos u}{1 - e\cos u}. \end{aligned} \tag{1.78}$$

To compute $\mathbf{r}(t), \mathbf{v}(t)$ from $\mathbf{r}_0 \equiv \mathbf{r}(t_0), \mathbf{v}_0 = \mathbf{v}(t_0)$ we use the following procedure. From equations (1.19) and (1.32) we have

$$\frac{1}{a} = \frac{2}{r} - \frac{v^2}{\mathbb{G}M}; \tag{1.79}$$

so we can compute the semimajor axis a from $r_0 = |\mathbf{r}_0|$ and $v_0 = |\mathbf{v}_0|$. Then Kepler's law (1.44) yields the mean motion n. The total angular momentum is $L = |\mathbf{r}_0 \times \mathbf{v}_0|$ and this yields the eccentricity e through equation (1.28). To determine the eccentric anomaly at t_0, we use equation (1.46) which determines $\cos u_0$, and then determine the quadrant of u_0 by observing that the radial velocity $\dot{r}$ is positive when $0 < u_0 < \pi$ and negative when $\pi < u_0 < 2\pi$. From Kepler's equation (1.49) we then find the mean anomaly ℓ_0 at $t = t_0$.

The mean anomaly at t is then $\ell = \ell_0 + n(t - t_0)$. By solving Kepler's equation numerically we can find the eccentric anomaly u. We may then evaluate the f and g functions using equations (1.78) and the position and velocity at t from equations (1.72) and (1.73).

1.4 Canonical orbital elements

The powerful tools of Lagrangian and Hamiltonian dynamics are essential for solving many of the problems addressed later in this book. A summary of the relevant aspects of this subject is given in Appendix D. In this section we show how Hamiltonian methods are applied to the two-body problem.

The Hamiltonian that describes the trajectory of a test particle around a point mass M at the origin is

$$H_{\rm K}(\mathbf{r},\mathbf{v}) = \tfrac{1}{2}\mathbf{v}^2 - \frac{\mathbb{G}M}{|\mathbf{r}|}. \tag{1.80}$$

Here $\mathbf{r}$ and $\mathbf{v}$ are the position and velocity, which together determine the position of the test particle in 6-dimensional phase space. The vectors $\mathbf{r}$ and $\mathbf{v}$ are a canonical coordinate-momentum pair.[10] Hamilton's equations read

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\partial H_{\rm K}}{\partial \mathbf{v}} = \mathbf{v}, \qquad \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{\partial H_{\rm K}}{\partial \mathbf{r}} = -\frac{\mathbb{G}M}{|\mathbf{r}|^3}\mathbf{r}. \tag{1.81}$$

These are equivalent to the usual equations of motion (1.11).

The advantage of Hamiltonian methods is that the equations of motion are the same in any set of phase-space coordinates $\mathbf{z} = (\mathbf{q},\mathbf{p})$ that are obtained from $(\mathbf{r},\mathbf{v})$ by a canonical transformation (Appendix D.6). For example, suppose that the test particle is also subject to an additional potential $\Phi(\mathbf{r},t)$ arising from some external mass distribution, such as another planet. Then the Hamiltonian and the equations of motion in the original variables are

$$H(\mathbf{r},\mathbf{v},t) = H_{\rm K}(\mathbf{r},\mathbf{v}) + \Phi(\mathbf{r},t), \qquad \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\partial H}{\partial \mathbf{v}}, \qquad \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{\partial H}{\partial \mathbf{r}}. \tag{1.82}$$

[10] We usually—but not always—adopt the convention that the canonical momentum $\mathbf{p}$ that is conjugate to the position $\mathbf{r}$ is velocity $\mathbf{v}$ rather than Newtonian momentum $m\mathbf{v}$. Velocity is often more convenient than Newtonian momentum in gravitational dynamics since the acceleration of a body in a gravitational potential is independent of mass. If necessary, the convention used in a particular set of equations can be verified by dimensional analysis.

In the new canonical variables,[11]

$$H(\mathbf{z},t) = H_{\mathrm{K}}(\mathbf{z}) + \Phi(\mathbf{z},t), \quad \frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \frac{\partial H}{\partial \mathbf{p}}, \quad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\partial H}{\partial \mathbf{q}}. \tag{1.83}$$

If the additional potential is small compared to the Kepler potential, $|\phi(\mathbf{r},t)| \ll \mathbb{G}M/r$, then the trajectory will be close to a Kepler ellipse. Therefore the analysis can be much easier if we use new coordinates and momenta $\mathbf{z}$ in which Kepler motion is simple.[12] The six orbital elements—semimajor axis a, eccentricity e, inclination I, longitude of the ascending node Ω, argument of periapsis ω and mean anomaly ℓ—satisfy this requirement as all of the elements are constant except for ℓ, which increases linearly with time. This set of orbital elements is not canonical, but they can be rearranged to form a canonical set called the **Delaunay variables**, in which the coordinate-momentum pairs are:

$$\begin{aligned} &\ell, \qquad && \Lambda \equiv (\mathbb{G}Ma)^{1/2}, \\ &\omega, && L = [\mathbb{G}Ma(1-e^2)]^{1/2}, \\ &\Omega, && L_z = L\cos I. \end{aligned} \tag{1.84}$$

Here L_z is the z-component of the angular-momentum vector $\mathbf{L}$ (see Figure 1.2); $L = |\mathbf{L}|$ (eq. 1.28); and Λ is sometimes called the **circular angular momentum** since it equals the angular momentum for a circular orbit. The proof that the Delaunay variables are canonical is given in Appendix E.

The Kepler Hamiltonian (1.80) is equal to the energy per unit mass, which is related to the semimajor axis by equation (1.32); thus

$$H_{\mathrm{K}} = -\frac{\mathbb{G}M}{2a} = -\frac{(\mathbb{G}M)^2}{2\Lambda^2}. \tag{1.85}$$

[11] For notational simplicity, we usually adopt the convention that the Hamiltonian and the potential are functions of position, velocity, or position in phase space rather than functions of the coordinates. Thus $H(\mathbf{r},\mathbf{v},t)$ and $H(\mathbf{z},t)$ have the same value if $(\mathbf{r},\mathbf{v})$ and $\mathbf{z}$ are coordinates of the same phase-space point in different coordinate systems.

[12] However, the additional potential $\Phi(\mathbf{z},t)$ is often much more complicated in the new variables; for a start, it generally depends on all six phase-space coordinates rather than just the three components of $\mathbf{r}$. Since dynamics is more difficult than potential theory, the tradeoff—simpler dynamics at the cost of more complicated potential theory—is generally worthwhile.

Since the Kepler Hamiltonian is independent of the coordinates, the momenta Λ, L and L_z are all constants along a trajectory in the absence of additional forces; such variables are called **integrals of motion**. Because the Hamiltonian is independent of the momenta L and L_z their conjugate coordinates ω and Ω are also constant, and $\mathrm{d}\ell/\mathrm{d}t = \partial H_\mathrm{K}/\partial\Lambda = (\mathbb{G}M)^2\Lambda^{-3} = (\mathbb{G}M/a^3)^{1/2} = n$, where n is the mean motion defined by Kepler's law (1.44). Of course, all of these conclusions are consistent with what we already know about Kepler orbits.

Because the momenta are integrals of motion in the Kepler Hamiltonian and the coordinates are angular variables that range from 0 to 2π, the Delaunay variables are also angle-action variables for the Kepler Hamiltonian (Appendix D.7). For an application of this property, see Box 1.2.

One shortcoming of the Delaunay variables is that they have coordinate singularities at zero eccentricity, where ω is ill-defined, and zero inclination, where Ω and ω are ill-defined. Even if the eccentricity or inclination of an orbit is small but nonzero, these elements can vary rapidly in the presence of small perturbing forces, so numerical integrations that follow the evolution of the Delaunay variables can grind to a near-halt.

To address this problem we introduce other sets of canonical variables derived from the Delaunay variables. We write $\mathbf{q} = (\ell, \omega, \Omega)$, $\mathbf{p} = (\Lambda, L, L_z)$ and introduce a generating function $S_2(\mathbf{q}, \mathbf{P})$ as described in Appendix D.6.1. From equations (D.63)

$$\mathbf{p} = \frac{\partial S_2}{\partial \mathbf{q}}, \quad \mathbf{Q} = \frac{\partial S_2}{\partial \mathbf{P}}, \tag{1.86}$$

and these equations can be solved for the new variables $\mathbf{Q}$ and $\mathbf{P}$. For example, if $S_2(\mathbf{q}, \mathbf{P}) = (\ell + \omega + \Omega)P_1 + (\omega + \Omega)P_2 + \Omega P_3$ then the new coordinate-momentum pairs are

$$\begin{aligned} \lambda = \ell + \omega + \Omega, &\quad \Lambda, \\ \varpi = \omega + \Omega, &\quad L - \Lambda = (\mathbb{G}Ma)^{1/2}\left[(1-e^2)^{1/2} - 1\right], \\ \Omega, &\quad L_z - L = (\mathbb{G}Ma)^{1/2}(1-e^2)^{1/2}(\cos I - 1). \end{aligned} \tag{1.87}$$

Here we have reintroduced the mean longitude λ (eq. 1.68) and the longitude of periapsis ϖ (eq. 1.67). Since λ and ϖ are well defined for orbits of zero inclination, these variables are better suited for describing nearly equatorial prograde orbits. The longitude of the node Ω is still ill-defined when the inclination is zero, although if the motion is known or assumed to be restricted to the equatorial plane the first two coordinate-momentum pairs are sufficient to describe the motion completely.

With the variables (1.87) two of the momenta $L - \Lambda$ and $L_z - L$ are always negative. For this reason some authors prefer to use the generating function $S_2(\mathbf{q}, \mathbf{P}) = (\ell + \omega + \Omega)P_1 - (\omega + \Omega)P_2 - \Omega P_3$, which yields new coordinates and momenta

$$\begin{aligned} \lambda = \ell + \omega + \Omega, &\quad \Lambda, \\ -\varpi = -\omega - \Omega, &\quad \Lambda - L = (\mathbb{G}Ma)^{1/2}\big[1 - (1 - e^2)^{1/2}\big], \\ -\Omega, &\quad L - L_z = (\mathbb{G}Ma)^{1/2}(1 - e^2)^{1/2}(1 - \cos I). \end{aligned} \tag{1.88}$$

Another set is given by the generating function $S_2(\mathbf{q}, \mathbf{P}) = \ell P_1 + (\ell + \omega)P_2 + \Omega P_3$, which yields coordinates and momenta

$$\begin{aligned} \ell, &\quad \Lambda - L = (\mathbb{G}Ma)^{1/2}\big[1 - (1 - e^2)^{1/2}\big], \\ \ell + \omega, &\quad L = (\mathbb{G}Ma)^{1/2}(1 - e^2)^{1/2}, \\ \Omega, &\quad L_z = (\mathbb{G}Ma)^{1/2}(1 - e^2)^{1/2}\cos I. \end{aligned} \tag{1.89}$$

The action $\Lambda - L$ that appears in (1.88) and (1.89) has a simple physical interpretation. At a given angular momentum L, the radial motion in the Kepler orbit is governed by the Hamiltonian $H(r, p_r) = \frac{1}{2}p_r^2 + \frac{1}{2}L^2/r^2 - \mathbb{G}M/r$ (cf. eq. 1.18). The corresponding action is $J_r = \oint \mathrm{d}r\, p_r/(2\pi)$ (eq. D.72). The radial momentum $p_r = \dot{r}$ by Hamilton's equations; writing r and $\dot{r}$ in terms of the eccentric anomaly u using equations (1.46) and (1.54) gives

$$J_r = \frac{na^2e^2}{2\pi}\int_0^{2\pi} \mathrm{d}u\, \frac{\sin^2 u}{1 - e\cos u} = na^2\big[1 - (1 - e^2)^{1/2}\big] = \Lambda - L. \tag{1.90}$$

Thus $\Lambda - L$ is the action associated with the radial coordinate, sometimes called the **radial action**. The radial action is zero for circular orbits and equal to $\frac{1}{2}(\mathbb{G}Ma)^{1/2}e^2$ when $e \ll 1$.

Box 1.2: The effect of slow mass loss on a Kepler orbit

If the mass of the central object is changing, the constant M in equations like (1.11) must be replaced by a variable $M(t)$. We assume that the evolution of the mass is (i) due to some spherically symmetric process (e.g., a spherical wind from the surface of a star), so there is no recoil force on the central object; (ii) slow, in the sense that $|\mathrm{d}M/\mathrm{d}t| \ll M/P$, where P is the orbital period of a planet.

Since the gravitational potential remains spherically symmetric, the angular momentum $L = (\mathbb{G}Ma)^{1/2}(1-e^2)^{1/2}$ (eq. 1.28) is conserved.

Moreover, actions are adiabatic invariants (Appendix D.10), so during slow mass loss the actions remain almost constant. The Delaunay variable $\Lambda = (\mathbb{G}Ma)^{1/2}$ (eq. 1.84) is an action. Since Λ and L are distinct functions of Ma and e, and both are conserved—one adiabatically and one exactly—then both Ma and e are also conserved. In words, during slow mass loss the orbit expands, with $a(t) \propto 1/M(t)$, but its eccentricity remains constant. The accuracy of this approximate conservation law is explored in Problem 2.8.

At present the Sun is losing mass at a rate $\dot{M}/M = -(1.1\pm0.3)\times10^{-13}\,\mathrm{yr}^{-1}$ (Pitjeva et al. 2021). Near the end of its life, the Sun will become a red-giant star and expand dramatically in radius and luminosity. At the tip of the red-giant branch, about 7.6 Gyr from now, the solar radius will be about 250 times its present value or 1.2 au and its luminosity will be 2 700 times its current value (Schröder & Connon Smith 2008). During its evolution up the red-giant branch the Sun will lose about 30% of its mass, and according to the arguments above the Earth's orbit will expand by the same fraction. Whether or not the Earth escapes being engulfed by the Sun depends on the uncertain relative rates of the Sun's future expansion and its mass loss.

Finally, consider the generating function $S_2(\mathbf{q},\mathbf{P}) = P_1(\ell+\omega+\Omega) + \frac{1}{2}{P_2}^2\cot(\omega+\Omega) + \frac{1}{2}{P_3}^2\cot\Omega$, which yields the **Poincaré variables**

$$
\begin{aligned}
&\lambda = \ell+\omega+\Omega, && \Lambda,\\
&[2(\Lambda-L)]^{1/2}\cos\varpi, && [2(\Lambda-L)]^{1/2}\sin\varpi,\\
&[2(L-L_z)]^{1/2}\cos\Omega, && [2(L-L_z)]^{1/2}\sin\Omega.
\end{aligned}
\tag{1.91}
$$

These are well defined even when $e = 0$ or $I = 0$. In particular, in the limit of small eccentricity and inclination the Poincaré variables simplify to

$$\begin{aligned} &\lambda, && \Lambda, \\ &(\mathbb{G}Ma)^{1/4} e \cos\varpi, && (\mathbb{G}Ma)^{1/4} e \sin\varpi, \\ &(\mathbb{G}Ma)^{1/4} I \cos\Omega, && (\mathbb{G}Ma)^{1/4} I \sin\Omega. \end{aligned} \tag{1.92}$$

Apart from the constant of proportionality $(\mathbb{G}Ma)^{1/4}$ these are just the Cartesian elements defined in equations (1.71).

All of these sets of orbital elements remain ill-defined when the inclination $I = \pi$ (retrograde orbits in the reference plane) or $e = 1$ (orbits with zero angular momentum); however, such orbits are relatively rare in planetary systems.[13]

1.5 Units and reference frames

Measurements of the trajectories of solar-system bodies are some of the most accurate in any science, and provide exquisitely precise tests of physical theories such as general relativity. Precision of this kind demands careful definitions of units and reference frames. These will only be treated briefly in this book, since our focus is on understanding rather than measuring the behavior of celestial bodies.

Tables of physical, astronomical and solar-system constants are given in Appendix A.

1.5.1 Time

The unit of time is the Système Internationale or SI second (s), which is defined by a fixed value for the frequency of a particular transition of cesium atoms. Measurements from several cesium frequency standards are combined to form a timescale known as **International Atomic Time** (TAI).

[13] A set of canonical coordinates and momenta that is well defined for orbits with zero angular momentum is described by Tremaine (2001). Alternatively, the orbit can be described using the angular-momentum and eccentricity vectors, which are well defined for any Kepler orbit; see §5.3 or Allan & Ward (1963).

In contrast, **Universal Time** (UT) employs the Earth's rotation on its axis as a clock. UT is not tied precisely to this clock because the Earth's angular speed is not constant. The most important nonuniformity is that the length of the day increases by about 2 milliseconds per century because of the combined effects of tidal friction and post-glacial rebound. There are also annual and semiannual variations of a few tenths of a millisecond. Despite these irregularities, a timescale based approximately on the Earth's rotation is essential for everyday life: for example, we would like noon to occur close to the middle of the day. Therefore all civil timekeeping is based on **Coordinated Universal Time** (UTC), which is an atomic timescale that is kept in close agreement with UT by adding extra seconds ("leap seconds") at regular intervals.[14] Thus UTC is a discontinuous timescale composed of segments that follow TAI apart from a constant offset.

An inconvenient feature of TAI for high-precision work is that it measures the rate of clocks at sea level on the Earth; general relativity implies that the clock rate depends on the gravitational potential and hence the rate of TAI is different from the rate measured by the same clock outside the solar system. For example, the rate of TAI varies with a period of one year and an amplitude of 1.7 milliseconds because of the eccentricity of the Earth's orbit. **Barycentric Coordinate Time** (TCB) measures the proper time experienced by a clock that co-moves with the center of mass of the solar system but is far outside it. TCB ticks faster than TAI by 0.49 seconds per year, corresponding to a fractional speedup of 1.55×10^{-8}.

The times of astronomical events are usually measured by the **Julian date**, denoted by the prefix JD. The Julian date is expressed in days and decimals of a day. Each day has $86\,400$ seconds. The Julian year consists of exactly 365.25 days and is denoted by the prefix J. For example, the initial conditions of orbits are often specified at a standard epoch, such as

$$\text{J2000.0} = \text{JD}\ 2\,451\,545.0, \tag{1.93}$$

which corresponds roughly to noon in England on January 1, 2000. The modified Julian day is defined as

$$\text{MJD} = \text{JD} - 2\,400\,000.5; \tag{1.94}$$

[14] The utility of leap seconds is controversial, and their future is uncertain.

the integer offset reduces the length of the number specifying relatively recent dates, and the half-integer offset ensures that the MJD begins at midnight rather than noon.

In contrast to SI seconds (s) and days (1 d = 86 400 s) there is no unique definition of "year": most astronomers use the Julian year but there is also the anomalistic year, sidereal year, and the like (see footnote 7). For this reason the use of "year" as a precise unit of time is deprecated. However, we shall occasionally use years, megayears and gigayears (abbreviated yr, Myr, Gyr) to denote 1, 10^6 and 10^9 Julian years. The age of the solar system is 4.567 Gyr and the age of the Universe is 13.79 Gyr. The future lifetime of the solar system as we know it is about 7.6 Gyr (see Box 1.2).

The SI unit of length is defined in terms of the second, such that the speed of light is exactly

$$c \equiv 299\,792\,458\,\mathrm{m\,s^{-1}}. \tag{1.95}$$

1.5.2 Units for the solar system

The history of the determination of the scale of the solar system and the mass of the Sun is worth a brief description. Until the mid-twentieth century virtually all of our data on the orbits of the Sun and planets came from tracking their positions on the sky as functions of time. This information could be combined with the theory of Kepler orbits developed earlier in this chapter (plus small corrections arising from mutual interactions between the planets, which are handled by the methods of Chapter 4) to determine all of the orbital elements of the planets including the Earth, except for the overall scale of the system. Thus, for example, the ratio of semimajor axes of any two planets was known to high accuracy, but the values of the semimajor axes in meters were not.[15] To reflect this uncertainty, astronomers introduced the concept of the **astronomical unit** (abbreviated au), which was originally defined to be the semimajor axis of the Earth's orbit. Thus the semimajor axes of the planets were known in astronomical units long before the value of the astronomical unit was known to comparable accuracy.

[15] This indeterminacy follows from dimensional analysis: measurements of angles and times cannot be combined to find a quantity with dimensions of length.

(continued...)

Index

Page numbers ending in "p" refer to problems.