

10

## Post-Newtonian celestial mechanics, astrometry and navigation

In November 1915, Einstein completed a calculation whose result so agitated him that he worried that he might be having a heart attack. He later wrote to a friend that “for several days I was beside myself in joyous excitement.” What Einstein calculated was the contribution to the advance of the perihelion of Mercury from the first post-Newtonian corrections to Newtonian gravity provided by his newly completed theory of general relativity. This had been a notorious and unsolved problem in astronomy, ever since Le Verrier pointed out in 1859 that there was a discrepancy of approximately 43 arcseconds per century in the rate of advance between what was observed and what could be accounted for in Newtonian theory from planetary perturbations, refer to Secs. 3.1 and 3.4. Many earlier attempts to devise relativistic theories of gravity, including Einstein’s own “Entwurf” (outline) theory of 1913 with Marcel Grossmann, had failed to give the correct answer. Now armed with the correct field equations, Einstein found an approximate vacuum solution that could be applied to the geodesic motion of Mercury around the Sun. He found that the orbit was almost Keplerian, but with a periastron that advances at a rate that matched Le Verrier’s observations.

For Einstein, this success with Mercury was the first concrete evidence that his theory, over which he had struggled so mightily for the past four years, might actually be correct. His prediction for the deflection of light by the Sun, completed that same month, and which doubled the value that he had derived in 1907 using just the principle of equivalence, would not be confirmed until 1919.

For the next 60 years, until the discovery of the first binary pulsar by Hulse and Taylor, the main testing-ground for general relativity would be the solar system, where gravitational fields are weak and motions are slow, so that the conditions for the post-Newtonian approximation are valid. In this chapter we take the formal development of post-Newtonian theory, presented in Chapters 8 and 9, and apply it to the real world of solar-system dynamics, solar-system experiments, and the dynamics of binary star (and binary black-hole) systems. We will encounter some of the famous experimental tests of general relativity, and will see how the precision of modern tools such as atomic clocks, satellite navigation systems, radio interferometry, and laser ranging make it necessary to take relativistic effects into account in a number of practical applications.

We begin in Sec. 10.1 with a description of the post-Newtonian motion of two self-gravitating bodies; it is in this section that we emulate Einstein with his calculation of the perihelion advance of Mercury. In Section 10.2 we examine the motion of light in weak gravitational fields, and describe such phenomena as the deflection of a light ray by a massive body, gravitational lenses, and the Shapiro time delay. The practically important

issue of clock synchronization in the presence of gravitation is the topic of Sec. 10.3. We conclude the chapter in Sec. 10.4 with a discussion of the motion of a binary system of spinning bodies.

## 10.1 Post-Newtonian two-body problem

We begin our exploration of post-Newtonian dynamics with an examination of the two-body problem, the relativistic generalization of the Kepler problem reviewed in Sec. 3.2. The foundations of the post-Newtonian problem were presented in Chapter 9, and the equations that govern the motion of binary systems were derived in Sec. 9.3.7. While the method of derivation employed in most of Chapter 9 was restricted to bodies with weak internal gravity, we saw back in Sec. 9.4 that in fact the equations of motion apply just as well to compact bodies such as neutron stars and black holes; the internal gravity of each body can be arbitrarily strong, but if the mutual gravity is weak, the equations apply.

The equations of motion can be applied to any binary-star system, but they can also be applied to the solar system, in spite of the fact that the number of bodies exceeds two. The reason is that the Sun dominates the mass of the solar system by a factor of 1000, so that from the point of view of relativistic effects, which scale as  $GM/(c^2r)$ , the Sun is the elephant in the room, making the other planets largely irrelevant. Consider, for example, the perihelion advance of Mercury, which is produced in part by relativistic effects, and in part by perturbations generated by other planets (mostly Jupiter). As we shall see below, the relativistic part of the advance per orbit is of order  $GM_{\odot}/(c^2a) \sim 3 \times 10^{-8}$  radians, where  $a$  is Mercury's semi-major axis. On the other hand, we recall from Sec. 3.4.1, refer to Eq. (3.83), that the advance produced by a perturbing planet scales as  $(m_p/M_{\odot})(a/R)^3$ , where  $m_p$  and  $R$  are the mass and semi-major axis of the perturbing planet; this is of order  $4 \times 10^{-7}$  radians per orbit when Jupiter is the perturbing planet, and is therefore comparable to the relativistic contribution. We can expect the post-Newtonian corrections to Newtonian third-body perturbations, as well as third-body contributions to the post-Newtonian perturbation, to be of the same order of magnitude as the product of these two small factors, and thus to be unmeasurably small. It follows that for most solar-system applications, we can safely calculate post-Newtonian two-body effects and Newtonian  $N$ -body effects separately, and simply add them together. The conclusion of this discussion is that the relativistic two-body problem provides an adequate foundation to describe most relativistic effects in the solar system.

An important exception to this rule occurs in the Earth–Moon system, for which there is a measurable post-Newtonian three-body effect produced by the Sun, a contribution to the precession of the lunar orbital plane. This effect is often called *de Sitter precession*, because it was predicted in 1916 by the Dutch astronomer Willem de Sitter, on the basis of his post-Newtonian equations of motion. We shall conclude this section with a description of the de Sitter precession.

### 10.1.1 Equations of motion

The equations that govern the motion of a two-body system were obtained back in Sec. 9.3.7, where they were cast in the form of an effective one-body problem. We work in terms of the separation  $\mathbf{r} := \mathbf{r}_1 - \mathbf{r}_2$ , relative velocity  $\mathbf{v} := \mathbf{v}_1 - \mathbf{v}_2$ , relative acceleration  $\mathbf{a} := \mathbf{a}_1 - \mathbf{a}_2$ , and involve the mass parameters  $m := M_1 + M_2$ ,  $\eta := M_1 M_2 / (M_1 + M_2)^2$ , and  $\Delta := (M_1 - M_2) / (M_1 + M_2)$ . The relative acceleration is given by Eq. (9.142),

$$\mathbf{a} = -\frac{Gm}{r^2} \mathbf{n} - \frac{Gm}{c^2 r^2} \left\{ \left[ (1 + 3\eta)v^2 - \frac{3}{2}\eta\dot{r}^2 - 2(2 + \eta)\frac{Gm}{r} \right] \mathbf{n} - 2(2 - \eta)\dot{r}\mathbf{v} \right\} + O(c^{-4}), \quad (10.1)$$

where  $r := |\mathbf{r}|$  is the inter-body distance,  $\mathbf{n} := \mathbf{r}/r$  a unit vector that points from body 2 to body 1, and  $\dot{r} := \mathbf{v} \cdot \mathbf{n}$  the radial component of the velocity vector. Equation (10.1) is a second-order differential equation for  $\mathbf{r}(t)$ , and its solution determines the position of each body. This information is provided by Eq. (9.141),

$$\mathbf{r}_1 = \frac{M_2}{m} \mathbf{r} + \frac{\eta\Delta}{2c^2} \left( v^2 - \frac{Gm}{r} \right) \mathbf{r} + O(c^{-4}), \quad (10.2a)$$

$$\mathbf{r}_2 = -\frac{M_1}{m} \mathbf{r} + \frac{\eta\Delta}{2c^2} \left( v^2 - \frac{Gm}{r} \right) \mathbf{r} + O(c^{-4}). \quad (10.2b)$$

These equations imply that  $\mathbf{v}_1 = (M_2/m)\mathbf{v} + O(c^{-2})$  and  $\mathbf{v}_2 = -(M_1/m)\mathbf{v} + O(c^{-2})$ .

We saw back in Sec. 3.2 that the Newtonian two-body problem admits constants of the motion, the orbital energy  $E = \mu\varepsilon$  and the orbital angular momentum  $\mathbf{L} = \mu\mathbf{h}$ , in which  $\mu := M_1 M_2 / (M_1 + M_2) = \eta m$  is the system's reduced mass. The post-Newtonian problem also admits a conserved energy and a conserved angular-momentum vector, and it is not difficult to deduce their expressions. We know that at the Newtonian level, the conserved energy is given by  $\varepsilon = \frac{1}{2}v^2 - Gm/r$ , and we also know that post-Newtonian corrections come with a multiplicative factor of  $c^{-2}$ . Possible contributions must then be proportional to  $v^4$ ,  $\dot{r}^2 v^2$ ,  $\dot{r}^4$ ,  $v^2 Gm/r$ ,  $\dot{r}^2 Gm/r$ , and  $(Gm/r)^2$ , and the correct combination of such terms can be identified by including them all (with unknown coefficients) in a trial expression for  $\varepsilon$ , and demanding that  $d\varepsilon/dt = 0$  by virtue of the post-Newtonian dynamics of Eq. (10.1). The end result of this simple exercise is the expression

$$\varepsilon := \frac{1}{2}v^2 - \frac{Gm}{r} + \frac{1}{c^2} \left\{ \frac{3}{8}(1 - 3\eta)v^4 + \frac{Gm}{2r} \left[ (3 + \eta)v^2 + \eta\dot{r}^2 + \frac{Gm}{r} \right] \right\} + O(c^{-4}) \quad (10.3)$$

for the post-Newtonian energy of the two-body system. In a similar way we can show that

$$\mathbf{h} := \left\{ 1 + \frac{1}{c^2} \left[ \frac{1}{2}(1 - 3\eta)v^2 + (3 + \eta)\frac{Gm}{r} \right] \right\} (\mathbf{r} \times \mathbf{v}) + O(c^{-4}) \quad (10.4)$$

is the post-Newtonian angular momentum.

Box 10.1

Ambiguities in energy and angular momentum

The expressions of Eqs. (10.3) and (10.4) are actually not unique. For example, six arbitrary coefficients are needed when constructing a post-Newtonian trial expression for  $\varepsilon$ , and six conditions are found when  $d\varepsilon/dt$  is required to vanish after involving the post-Newtonian equations of motion. Two of the conditions, however, turn out to be redundant, and as a result, one coefficient cannot be determined. It is easy to show that the free coefficient represents the freedom to add to the expression for  $\varepsilon$  an arbitrary amount of a post-Newtonian contribution  $c^{-2}(v^2 - 2Gm/r)^2$ . This is constant by virtue of the Newtonian equations of motion, and so the constancy of  $\varepsilon$  through  $O(c^{-2})$  still holds. This freedom reflects the fact that the zero of energy is not fixed in classical mechanics. To arrive at Eq. (10.3) we fixed the free coefficient so that the  $v^4$  term in  $\varepsilon$  has the factor shown. With this choice, the limit  $Gm/r \rightarrow 0$  of  $\varepsilon$  matches the Newtonian and post-Newtonian terms in the expansion of the special relativistic energy of two bodies, given by  $\gamma_1 m_1 c^2 + \gamma_2 m_2 c^2$ , where  $\gamma_A = (1 - v_A^2/c^2)^{-1/2}$ . Similar considerations apply to the conserved angular momentum  $\mathbf{h}$ .

It is interesting to note that although  $\mathbf{h}$  is conserved in the post-Newtonian dynamics, the vector  $\mathbf{r} \times \mathbf{v}$  is not; this is a point of departure from the Newtonian situation. But while  $\mathbf{r} \times \mathbf{v}$  is no longer constant in magnitude, it is still constant *in direction*, and this is sufficient to establish that *the orbital motion proceeds within a fixed orbital plane*, just as in the Newtonian situation. As in Sec. 3.2, we may simplify the mathematical description of the post-Newtonian motion by taking the orbital plane to coincide with the  $x$ - $y$  plane of the coordinate system, and by introducing the vectorial basis  $\mathbf{n} := [\cos \phi, \sin \phi, 0]$ ,  $\boldsymbol{\lambda} := [-\sin \phi, \cos \phi, 0]$ , and  $\mathbf{e}_z := [0, 0, 1]$ , in which  $\phi$  is the orbital angle. In terms of the orbital basis we have  $\mathbf{r} = r \mathbf{n}$ ,  $\mathbf{v} = \dot{r} \mathbf{n} + r \dot{\phi} \boldsymbol{\lambda}$ ,

$$\mathbf{a} = (\ddot{r} - r \dot{\phi}^2) \mathbf{n} + \frac{1}{r} \frac{d}{dt}(r^2 \dot{\phi}) \boldsymbol{\lambda}, \quad (10.5)$$

and  $\mathbf{r} \times \mathbf{v} = (r^2 \dot{\phi}) \mathbf{e}_z$ . The equations of motion become

$$\ddot{r} = r \dot{\phi}^2 - \frac{Gm}{r^2} + \frac{Gm}{c^2 r^2} \left[ \frac{1}{2} (6 - 7\eta) \dot{r}^2 - (1 + 3\eta) (r \dot{\phi})^2 + 2(2 + \eta) \frac{Gm}{r} \right] + O(c^{-4}), \quad (10.6a)$$

$$\frac{d}{dt}(r^2 \dot{\phi}) = 2(2 - \eta) \frac{Gm}{c^2} \dot{r} \dot{\phi} + O(c^{-4}), \quad (10.6b)$$

when they are expressed in terms of the dynamical variables  $r(t)$  and  $\phi(t)$ .

The integration of Eqs. (10.6) is greatly facilitated by the existence of  $\varepsilon$  and  $\mathbf{h} := |\mathbf{h}|$  as constants of the motion. In fact, the conserved quantities can be used to express  $\dot{r}^2$  and  $\dot{\phi}$  as simple polynomials in  $1/r$ . A simple computation reveals that

$$\begin{aligned} \dot{r}^2 = & 2\varepsilon \left[ 1 - \frac{3}{2}(1 - 3\eta) \frac{\varepsilon}{c^2} \right] + 2 \frac{Gm}{r} \left[ 1 - (6 - 7\eta) \frac{\varepsilon}{c^2} \right] - \frac{h^2}{r^2} \left[ 1 - 2(1 - 3\eta) \frac{\varepsilon}{c^2} \right] \\ & - 5(2 - \eta) \frac{(Gm)^2}{c^2 r^2} + (8 - 3\eta) \frac{Gm h^2}{c^2 r^3} + O(c^{-4}) \end{aligned} \quad (10.7)$$

and

$$\dot{\phi} = \frac{h}{r^2} \left[ 1 - (1 - 3\eta) \frac{\varepsilon}{c^2} \right] - 2(2 - \eta) \frac{Gmh}{c^2 r^3}. \quad (10.8)$$

These equations can be compared with their Keplerian version displayed in Eqs. (3.10) and (3.13).

### 10.1.2 Circular orbits

Our next task is integration of the post-Newtonian equations of motion. We begin with the simple case of a circular orbit of radius  $r$ . Setting  $\dot{r} = 0$  in Eq. (10.6b) reveals that the orbital angular velocity  $\dot{\phi}$  is a constant that we denote  $\Omega$ . Setting  $\ddot{r} = 0$  in Eq. (10.6a) allows us to relate  $\Omega$  to  $r$ , and we obtain

$$\Omega^2 = \frac{Gm}{r^3} \left[ 1 - (3 - \eta) \frac{Gm}{c^2 r} + O(c^{-4}) \right], \quad (10.9)$$

the relativistic version of the familiar Keplerian relation  $\Omega^2 = Gm/r^3$ . The orbital velocity is then given by

$$v^2 = (r\Omega)^2 = \frac{Gm}{r} \left[ 1 - (3 - \eta) \frac{Gm}{c^2 r} + O(c^{-4}) \right], \quad (10.10)$$

and the constants of the motion reduce to

$$\varepsilon = -\frac{Gm}{2r} \left[ 1 - \frac{1}{4}(7 - \eta) \frac{Gm}{c^2 r} + O(c^{-4}) \right] \quad (10.11)$$

and

$$h = \sqrt{Gmr} \left[ 1 + 2 \frac{Gm}{c^2 r} + O(c^{-4}) \right]. \quad (10.12)$$

### 10.1.3 Perturbed Keplerian orbits

A possible way of integrating the equations of motion is to treat the post-Newtonian terms in Eq. (10.1) as a perturbing force  $\mathbf{f}$  in the perturbed Kepler problem described in Sec. 3.3. (Recall our admonition from Chapter 3, that even though  $\mathbf{f}$  is an acceleration, we nevertheless use the conventional term “force” to describe it.) We exploit the method of osculating orbital elements, in which the perturbed motion is represented by a Keplerian orbit with time-dependent orbital elements, which vary in response to the perturbing force. The components of the perturbing force in the orbital basis  $(\mathbf{n}, \boldsymbol{\lambda}, \mathbf{e}_z)$  are

$$\mathcal{R} = \frac{Gm}{c^2 r^2} \left[ -(1 + 3\eta)v^2 + \frac{1}{2}(8 - \eta)\dot{r}^2 + 2(2 + \eta) \frac{Gm}{r} \right], \quad (10.13a)$$

$$\mathcal{S} = \frac{Gm}{c^2 r^2} \left[ 2(2 - \eta)\dot{r}(r\dot{\phi}) \right], \quad (10.13b)$$

$$\mathcal{W} = 0, \quad (10.13c)$$

in which we insert the Keplerian relations  $r = p/(1 + e \cos f)$ ,  $\dot{r} = \sqrt{Gm/p} e \sin f$ ,  $r\dot{\phi} = \sqrt{Gm/p}(1 + e \cos f)$ , where  $p$  is the Keplerian semi-latus rectum,  $e$  is the Keplerian eccentricity, and  $f := \phi - \omega$  is the true anomaly, with  $\omega$  denoting the Keplerian longitude of pericenter. Substituting the force components within Eqs. (3.69) produces

$$\frac{dp}{df} = 4(2 - \eta) \frac{Gm}{c^2} e \sin f, \quad (10.14a)$$

$$\frac{de}{df} = \frac{Gm}{c^2 p} \left\{ \left[ 3 - \eta + \frac{1}{8}(56 - 47\eta)e^2 \right] \sin f + (5 - 4\eta)e \sin 2f - \frac{3}{8}\eta e^2 \sin 3f \right\}, \quad (10.14b)$$

$$\frac{d\omega}{df} = \frac{1}{e} \frac{Gm}{c^2 p} \left\{ 3e - \left[ 3 - \eta - \frac{1}{8}(8 + 21\eta)e^2 \right] \cos f - (5 - 4\eta)e \cos 2f + \frac{3}{8}\eta e^2 \cos 3f \right\}. \quad (10.14c)$$

The vanishing of  $\mathcal{W} := \mathbf{f} \cdot \mathbf{e}_z$  implies that the inclination angle  $i$  and longitude of ascending node  $\Omega$  are not affected by the perturbing force; this is a consequence of the fact that the post-Newtonian motion proceeds in a fixed orbital plane.

The perturbing force  $\mathbf{f}$  is of first post-Newtonian order, and working consistently at this order, it is appropriate to integrate Eqs. (10.14) while keeping the orbital elements constant on the right-hand side of the equations. In this way we obtain

$$p(f) = p_0 \left[ 1 + 4(2 - \eta) \frac{Gm}{c^2 p_0} (1 - \cos f) \right] \quad (10.15)$$

and more complicated expressions for  $e(f)$  and  $\omega(f)$ ; here  $p_0 := p(f = 0)$ . These expressions can then be inserted within the original Keplerian relations to obtain the complete solutions for  $\mathbf{r}(f)$  and  $\mathbf{v}(f)$ . The post-Newtonian motion is thus parameterized by the true anomaly  $f$ , and its description in terms of time  $t$  can be obtained by integrating Eq. (3.70).

### 10.1.4 Pericenter advance

The description of the post-Newtonian motion in terms of osculating Keplerian orbits is perfectly adequate from a mathematical point of view, but it is fairly awkward to use and not well suited to practical implementations. An illustration of this is the curious fact that the circular orbit of radius  $r$  examined in Sec. 10.1.2 is one with Keplerian orbital parameters  $e = (3 - \eta)Gm/(c^2 r)$  and  $p = r[1 - (3 - \eta)Gm/(c^2 r)]$ , and also this orbit has a constant true anomaly given by  $f = \pi$ , so that the pericenter advances at the same rate as the body itself. We shall give the motion a much better description in the next subsection, but let's not give up just yet on the osculating formulation.

In the applications of the method of osculating orbital elements examined in Sec. 3.4, we saw that the method's most powerful insights are delivered when it is asked to reveal the *secular changes* in the orbital elements, and allowed to discard any information about the periodic changes that average out after each orbital cycle. We shall adopt this wisdom here, and calculate the secular changes in  $p$ ,  $e$ , and  $\omega$  produced by the post-Newtonian perturbation of Eqs. (10.13). They are obtained by integrating Eqs. (10.14) over a complete orbital period (from  $f = 0$  to  $f = 2\pi$ ), and the equations produce  $\Delta p = 0$ ,  $\Delta e = 0$ , as

well as

$$\Delta\omega = 6\pi \frac{Gm}{c^2 p}. \quad (10.16)$$

The fact that  $p$  undergoes no secular change is a consequence of angular-momentum conservation, and the absence of a secular change in  $e$  (and therefore in the Keplerian semi-major axis  $a$ ) is a consequence of energy conservation. The only parameter that undergoes a secular evolution is the longitude of periastron  $\omega$ , and Eq. (10.16) describes the pericenter advance that was so famously calculated by Einstein.

Einstein's method of derivation was very different. He did not have access to the post-Newtonian metric of an  $N$ -body system, and he did not have access to the  $N$ -body equations of motion. What he did was to obtain the post-Newtonian metric of a single body (the Sun), and to calculate the motion of a second body (Mercury) under the assumption that it is a test mass that moves on a geodesic of the spacetime. This is a sensible assumption, given that the Mercury–Sun mass ratio is approximately one to 6 million. Einstein obtained the result of Eq. (10.16), but with the total mass  $m := M_1 + M_2$  well approximated by  $M_1$ , the mass of the Sun. Our result is more general, and it applies to a much broader range of situations.

The advance per orbit can be converted to a rate of advance by dividing by the orbital period. We can also eliminate the semi-major axis appearing in  $p = a(1 - e^2)$  by using Kepler's third law,  $a = (GM)^{1/3}(P/2\pi)^{2/3}$ , where  $P$  is the orbital period; in principle the relation should be modified by a post-Newtonian correction, but the modification is irrelevant because  $\Delta\omega$  is already of first post-Newtonian order. The result is

$$\begin{aligned} \left(\frac{d\omega}{dt}\right)_{\text{sec}} &= \frac{3}{1 - e^2} \frac{(GM/c^3)^{2/3}}{(P/2\pi)^{5/3}} \\ &= 716.25 \frac{1}{1 - e^2} \left(\frac{M}{M_\odot}\right)^{2/3} \left(\frac{P}{1 \text{ day}}\right)^{-5/3} \text{ as/yr}. \end{aligned} \quad (10.17)$$

Substituting the values for Mercury,  $e = 0.2056$  and  $P = 87.97$  days, we obtain 42.98 arcseconds per century. As we saw back in Table 3.1, the modern difference between the measured advance and the one predicted by Newtonian  $N$ -body perturbations is  $42.98 \pm 0.04$  arcseconds per century, in 0.1% agreement with the relativistic prediction.

The discovery of binary-pulsar systems with total masses of 2 to 3 solar masses, and with orbital periods as small as fractions of a day, resulted in the observation of periastron advances of several degrees per year. The relativistic periastron advance plays an interesting role in these systems. In the solar system,  $GM$  for the Sun is known to high precision from the measured orbital period and orbital radius of the Earth, combined with Kepler's third law. By contrast, the masses of the neutron stars are not known, apart from the general expectation that they should be around the Chandrasekhar limit of  $1.4 M_\odot$ , based on models of how such systems might have formed. In the famous Hulse–Taylor binary pulsar, the first such system to be discovered, it was possible to measure the orbital eccentricity, the orbital period, and the pericenter advance very accurately; the current values are  $e = 0.6171338(4)$ ,  $P = 0.322997448930(4)$  day, and  $\dot{\omega} = 4.226595(5)$  deg/yr, in which the number in parentheses denotes the error in the final digit. Assuming that there

is no other source of periastron advance, we can turn Eq. (10.17) around and use it to measure the total mass of the system. The result is  $M = 2.828296(5) M_\odot$ . The recently discovered double pulsar J0737-3039A/B, in which both stars are observed as pulsars, has  $e = 0.0877775(9)$ ,  $P = 0.10225156248(5)$  day, and  $\dot{\omega} = 16.8995(7)$  deg/yr, giving a total mass of  $2.5871(2) M_\odot$ . These are remarkably accurate measurements of an astrophysical quantity, and as we can see, general relativity plays a central role in the analysis.

### 10.1.5 Integration of the equations of motion

To integrate the post-Newtonian equations of motion listed in Sec. 10.1.1, we adopt the approach followed by Damour and Deruelle in their seminal 1985 paper. Our starting point is the observation that the transformation

$$r = \bar{r} - \frac{1}{2}(8 - 3\eta)\frac{Gm}{c^2} + O(c^{-4}) \quad (10.18)$$

turns Eq. (10.7) into the simpler polynomial

$$\dot{\bar{r}}^2 = 2\varepsilon_K + 2\frac{Gm_K}{\bar{r}} - \frac{h_K^2}{\bar{r}^2}, \quad (10.19)$$

in which

$$\varepsilon_K := \varepsilon \left[ 1 - \frac{3}{2}(1 - 3\eta)\frac{\varepsilon}{c^2} + O(c^{-4}) \right], \quad (10.20a)$$

$$m_K := m \left[ 1 - (6 - 7\eta)\frac{\varepsilon}{c^2} + O(c^{-4}) \right], \quad (10.20b)$$

$$h_K^2 := h^2 \left[ 1 - 2(1 - 3\eta)\frac{\varepsilon}{c^2} + 2(1 - \eta)\frac{(Gm)^2}{c^2 h^2} + O(c^{-4}) \right]. \quad (10.20c)$$

The radial equation is an exact replica of the Keplerian equation displayed in Eq. (3.13), and it therefore admits the same solution. We adopt a representation in terms of an eccentric anomaly  $u$ , according to which  $\bar{r} = \bar{a}(1 - \bar{e} \cos u)$  and  $t - T = \sqrt{\bar{a}^3/(Gm_K)}(u - \bar{e} \sin u)$ , in which  $\bar{a}$  is a post-Newtonian semi-major axis and  $\bar{e}$  a post-Newtonian eccentricity, defined in the same way as in the Keplerian problem:  $\varepsilon_K = -Gm_K/(2\bar{a})$  and  $h_K^2 = Gm_K\bar{a}(1 - \bar{e}^2)$ .

With the solution for  $\bar{r}$  obtained, we apply the transformation of Eq. (10.18) to express the motion directly in terms of  $r$ . Simple manipulations produce

$$r = a(1 - e \cos u), \quad (10.21)$$

in which  $a := \bar{a}[1 - \frac{1}{2}(8 - 3\eta)Gm/(c^2\bar{a})]$  and  $e = \bar{e}[1 + \frac{1}{2}(8 - 3\eta)Gm/(c^2\bar{a})]$  are new post-Newtonian orbital elements. Adopting these as the primary elements, it is a simple matter to express the conserved energy  $\varepsilon$  and conserved angular momentum  $h$  in terms of  $a$  and  $e$ ; we obtain

$$\varepsilon = -\frac{Gm}{2a} \left[ 1 - \frac{1}{4}(7 - \eta)\frac{Gm}{c^2 a} + O(c^{-4}) \right] \quad (10.22)$$



and

$$h^2 = Gma(1 - e^2) \left[ 1 + \frac{4 + (2 - \eta)e^2}{1 - e^2} \frac{Gm}{c^2 a} + O(c^{-4}) \right]. \quad (10.23)$$

The time function can then be expressed as

$$t - T = \frac{P}{2\pi} (u - e_t \sin u), \quad (10.24)$$

in which

$$e_t := e \left[ 1 - \frac{1}{2}(8 - 3\eta) \frac{Gm}{c^2 a} + O(c^{-4}) \right] \quad (10.25)$$

is a second eccentricity parameter (which differs from  $e$  by a post-Newtonian correction), and

$$P := 2\pi \sqrt{\frac{a^3}{Gm}} \left[ 1 + \frac{1}{2}(9 - \eta) \frac{Gm}{c^2 a} + O(c^{-4}) \right] \quad (10.26)$$

is the post-Newtonian period.

To obtain  $\phi$  as a function of  $u$  we begin with the transformation

$$r = \tilde{r} - (2 - \eta) \frac{Gm}{c^2} + O(c^{-4}), \quad (10.27)$$

which turns Eq. (10.8) into

$$\dot{\phi} = \frac{\tilde{h}}{\tilde{r}^2}, \quad (10.28)$$

where

$$\tilde{h} := \sqrt{Gma(1 - e^2)} \left[ 1 + \frac{3(1 - \eta) + (1 + 2\eta)e^2}{2(1 - e^2)} \frac{Gm}{c^2 a} + O(c^{-4}) \right]. \quad (10.29)$$

We next use Eq. (10.21) to write  $\tilde{r} = \tilde{a}(1 - \tilde{e} \cos u)$ , in which  $\tilde{a} = a[1 + (2 - \eta)Gm/(c^2 a)]$  and  $\tilde{e} = e[1 - (2 - \eta)Gm/(c^2 a)]$ , and Eq. (10.24) to obtain

$$\frac{d\phi}{du} = \frac{P}{2\pi} \frac{\tilde{h}}{\tilde{a}^2} \frac{1 - e_t \cos u}{(1 - \tilde{e} \cos u)^2} + O(c^{-4}), \quad (10.30)$$

which can be integrated to yield the orbital angle  $\phi$  in terms of eccentric anomaly  $u$ . This equation can be compared with the Keplerian version of Eq. (3.33), and here we see a substantial difference in the form of the equations. In addition to the multiplicative factor that appears on the right-hand side, we see that the post-Newtonian expression involves  $(1 - e_t \cos u)/(1 - \tilde{e} \cos u)^2$  while the Keplerian version features the simpler factor of  $(1 - e \cos u)^{-1}$ . The forms, however, can be reconciled with a simple trick. We first observe that  $e_t$  differs from  $\tilde{e}$  by a post-Newtonian correction that we denote  $\epsilon$ . We next define a third eccentricity parameter by  $e_\phi := \tilde{e} - \epsilon$ , and factorize  $(1 - \tilde{e} \cos u)^2$  as

$$\begin{aligned} (1 - \tilde{e} \cos u)^2 &= [1 - (\tilde{e} + \epsilon) \cos u][1 - (\tilde{e} - \epsilon) \cos u] + O(c^{-4}) \\ &= (1 - e_t \cos u)(1 - e_\phi \cos u) + O(c^{-4}). \end{aligned} \quad (10.31)$$

This allows us to re-express  $d\phi/du$  as

$$\frac{d\phi}{du} = (1+k) \frac{(1-e_\phi^2)^{1/2}}{1-e_\phi \cos u} + O(c^{-4}), \quad (10.32)$$

which now looks identical to Eq. (3.33), except for the factor  $1+k := (P/2\pi)(\tilde{h}/\tilde{a}^2)(1-e_\phi^2)^{-1/2}$ .

The solution to this equation can be expressed as in Eqs. (3.30) and (3.32). We have that

$$\cos\left(\frac{\phi}{1+k}\right) = \frac{\cos u - e_\phi}{1 - e_\phi \cos u}, \quad \sin\left(\frac{\phi}{1+k}\right) = \frac{(1-e_\phi^2)^{1/2} \sin u}{1 - e_\phi \cos u}, \quad (10.33)$$

or that

$$\tan\left[\frac{\phi}{2(1+k)}\right] = \sqrt{\frac{1+e_\phi}{1-e_\phi}} \tan \frac{u}{2}. \quad (10.34)$$

Simple algebra confirms that

$$e_\phi = e \left[ 1 + \frac{1}{2} \eta \frac{Gm}{c^2 a} + O(c^{-4}) \right] \quad (10.35)$$

and

$$k = \frac{3}{1-e^2} \frac{Gm}{c^2 a} + O(c^{-4}). \quad (10.36)$$

The meaning of this last quantity is easy to extract from Eqs. (10.33) and (10.34). These relations inform us that in the course of a complete radial cycle, as  $u$  runs from 0 to  $2\pi$ , the orbital angle  $\phi$  runs from 0 to  $2\pi(1+k)$ . The excess angle,

$$2\pi k = 6\pi \frac{Gm}{c^2 a(1-e^2)} + O(c^{-4}), \quad (10.37)$$

is the pericenter advance of Eq. (10.16).

It is a remarkable fact that the two-body equations of motion of post-Newtonian theory can be integrated in the same manner as the Keplerian equations, with only the small cost of introducing two additional eccentricity parameters ( $e_r$  and  $e_\phi$ , which differ from  $e$  by post-Newtonian corrections), and a pericenter-advance parameter  $k := \Delta\omega/(2\pi)$ . This implies that the post-Newtonian motion can be computed with great ease, by exploiting the tried and true methods of celestial mechanics.

The individual motion of each body can be obtained from Eq. (10.2). It is easy to show that

$$r_1 := |\mathbf{r}_1| = a_1(1 - e_1 \cos u), \quad (10.38)$$

in which  $a_1 := (M_2/m)a$  is the semi-major axis of the first body, while

$$e_1 = e \left[ 1 - \frac{M_1(M_1 - M_2)}{2m^2} \frac{Gm}{c^2 a} \right] \quad (10.39)$$

is its eccentricity. The corresponding results for the second body are obtained by a suitable exchange of labels.

### 10.1.6 de Sitter precession

The de Sitter precession of the lunar orbit is a relativistic three-body effect that involves the Moon, the Earth, and the Sun; it is a secular advance of the line of nodes of the lunar orbital plane. You will recall from our discussion in Chapter 3 that the motion of the Moon is a notoriously difficult problem in Newtonian gravity, mainly because the strong perturbations caused by the Sun lead to a poorly convergent sequence of corrections. Special formulations of the perturbation theory, such as the Hill–Brown theory, were required for high precision, and today the equations of motion are solved directly using computers. With the motion of the Moon in Newtonian gravity now under control, discrepancies between the predicted and observed motions can be attributed to relativistic effects.

Relativistic three-body effects can be investigated on the basis of the  $N$ -body equations listed in Sec. 9.3.5 – refer to Eq. (9.127). As in Sec. 3.4.1, in which we considered third-body effects in Newtonian theory, we focus our attention on a two-body system (the Earth and the Moon) and examine the perturbations produced by a third body (the Sun). We let the Moon be the first body (mass  $M_1$ , position  $\mathbf{r}_1$ ), the Earth be the second body (mass  $M_2$ , position  $\mathbf{r}_2$ ), and the Sun be the third body (mass  $M_3$ , position  $\mathbf{r}_3$ ). We let  $\mathbf{r} := \mathbf{r}_1 - \mathbf{r}_2$  be the separation between the Moon and the Earth, and  $\mathbf{R} := \mathbf{r}_2 - \mathbf{r}_3$  be the separation between the Earth and the Sun. Similarly, we let  $\mathbf{v} := \mathbf{v}_1 - \mathbf{v}_2$  be the Moon’s velocity relative to Earth’s, and  $\mathbf{V} := \mathbf{v}_2 - \mathbf{v}_3$  be the Earth’s velocity relative to the Sun. For simplicity we imagine that the Sun is at rest at the spatial origin of the coordinate system, so that  $\mathbf{r}_3 = \mathbf{0} = \mathbf{v}_3$ .

The relativistic three-body equations lead to effects of various sizes on the two-body system. We wish to focus our attention on the dominant effects, and to allow ourselves to neglect the smaller ones, or to ignore relativistic effects that would be masked by much larger Newtonian effects. To seek guidance in the identification of which terms must be kept in the equations of motion, and which terms can be neglected, we first examine the various scales of the problem.

Our first observation is that  $M_1$ , the mass of the Moon, is much smaller than both  $M_2$  and  $M_3$ , and we shall therefore neglect all terms in the equations of motion that involve  $M_1$ . Our second observation is that the acceleration of the Earth–Moon system toward the Sun is approximately twice the acceleration of the Moon toward the Earth,  $GM_3/R^2 \sim 2GM_2/r^2$ ; this comes about because  $r/R \sim 2 \times 10^{-3}$ , which counteracts the large value of  $M_3/M_2$ . Our third observation is that the post-Newtonian corrections scale as  $GM_3/(c^2 R) \sim 10^{-8}$  for the Earth–Moon system moving around the Sun, so that  $V/c \sim 10^{-4}$ , and as  $GM_2/(c^2 r) \sim 10^{-11}$  for the Moon moving around the Earth, so that  $v/c \sim 3 \times 10^{-6}$ ; the motion around the Sun is more relativistic than the motion around the Earth. We rely on this observation to neglect  $GM_2/(c^2 r)$  compared to  $GM_3/(c^2 R)$  in the equations of motion, and to neglect  $(v/c)^2$  compared to  $(V/c)^2$ . Another source of simplification is the previously noted smallness of  $r/R$ , which allows us to ignore the inhomogeneity of the Sun’s gravitational field across the Earth–Moon orbit; this means that  $\mathbf{r}_{13} = \mathbf{R} + \mathbf{r}$  can be safely approximated by  $\mathbf{R}$  in the equations of motion.

With these simplifications, it is a straightforward exercise to apply Eq. (9.127) to the Earth–Moon system perturbed by the Sun. We find that the post-Newtonian term in the

Moon's acceleration reduces to

$$\begin{aligned} \mathbf{a}_1[\text{PN}] = & \frac{GM_2}{c^2 r^2} \left[ V^2 + 2\mathbf{V} \cdot \mathbf{v} + \frac{3}{2}(\mathbf{n} \cdot \mathbf{V})^2 + \frac{5GM_3}{R} \right] \mathbf{n} + \frac{GM_2}{c^2 r^2} (\mathbf{n} \cdot \mathbf{V}) \mathbf{v} \\ & - \frac{GM_3}{c^2 R^2} \left[ V^2 + 2\mathbf{V} \cdot \mathbf{v} - \frac{4GM_3}{R} \right] \mathbf{N} \\ & + \frac{4GM_3}{c^2 R^2} \left[ (\mathbf{N} \cdot \mathbf{V}) \mathbf{V} + (\mathbf{N} \cdot \mathbf{V}) \mathbf{v} + (\mathbf{N} \cdot \mathbf{v}) \mathbf{V} \right], \end{aligned} \quad (10.40)$$

in which  $\mathbf{n} := \mathbf{r}/r$  and  $\mathbf{N} := \mathbf{R}/R$ , and that the Earth's post-Newtonian acceleration becomes

$$\mathbf{a}_2[\text{PN}] = -\frac{GM_3}{c^2 R^2} \left( V^2 - \frac{4GM_3}{R} \right) \mathbf{N} + \frac{4GM_3}{c^2 R^2} (\mathbf{N} \cdot \mathbf{V}) \mathbf{V}. \quad (10.41)$$

The relative acceleration  $\mathbf{a}[\text{PN}] := \mathbf{a}_1[\text{PN}] - \mathbf{a}_2[\text{PN}]$  is then

$$\begin{aligned} \mathbf{a}[\text{PN}] = & \frac{GM_2}{c^2 r^2} \left[ V^2 + \frac{3}{2}(\mathbf{n} \cdot \mathbf{V})^2 + \frac{5GM_3}{R} \right] \mathbf{n} \\ & + \frac{2GM_3}{c^2 R^2} \left[ 2(\mathbf{N} \cdot \mathbf{V}) \mathbf{v} + 2(\mathbf{N} \cdot \mathbf{v}) \mathbf{V} - (\mathbf{V} \cdot \mathbf{v}) \mathbf{N} \right], \end{aligned} \quad (10.42)$$

in which we have neglected  $2\mathbf{V} \cdot \mathbf{v}$  compared to  $V^2$  in the first group of terms. This can be treated as a perturbing force  $\mathbf{f}$  on the Keplerian system formed by the Earth and the Moon, and it can be used in a perturbative evolution of the orbital elements along the lines developed in Sec. 3.4.

The motion of the Moon around the Earth is described by an osculating Keplerian orbit of elements  $(p, e, \omega, \iota, \Omega)$  perturbed by the external force of Eq. (10.42). To simplify the description of the evolution of the orbital elements, we follow the approach of Sec. 3.4.1 and place the Earth on a circular orbit of radius  $R$ , angular velocity  $\Omega_{\text{orb}}$ , and orbital phase  $F = \Omega_{\text{orb}} t$  around the Sun; the orbit is situated in the fundamental  $X$ - $Y$  plane. It is then a straightforward task to calculate the components  $(\mathcal{R}, \mathcal{S}, \mathcal{W})$  of the perturbing force, to insert these into the osculating equations (3.69), to obtain the total changes  $(\Delta p, \Delta e, \Delta \omega, \Delta \iota, \Delta \Omega)$  over a complete lunar orbit, and finally, to average these over the motion of the Earth around the Sun. Our final results are that  $\langle \Delta p \rangle = \langle \Delta e \rangle = \langle \Delta \iota \rangle = \langle \Delta \omega \rangle = 0$ , but that

$$\langle \Delta \Omega \rangle = \frac{3\pi G}{c^2} \frac{M_{\odot}^{3/2} p^{3/2}}{M_{\oplus}^{1/2} R^{5/2} (1 - e^2)^{3/2}}, \quad (10.43)$$

in which we have substituted the standard symbols  $M_{\odot}$  and  $M_{\oplus}$  for the mass of the Sun and the Earth, respectively. Thus, the only long-term impact of relativistic three-body effects on the lunar orbit is a precession of the line of nodes, whose longitude  $\Omega$  advances at the averaged rate

$$\left( \frac{d\Omega}{dt} \right)_{\text{sec}} = \frac{3}{2} \sqrt{\frac{GM_{\odot}}{R^3}} \frac{GM_{\oplus}}{c^2 R}, \quad (10.44)$$

obtained by dividing  $\langle \Delta \Omega \rangle$  by the Moon's orbital period.

According to Eq. (10.44), the line of nodes of the lunar orbit advances at an averaged rate of 19.1 arcseconds per century. The effect was first predicted in 1916 by de Sitter, but at the time it was far too small to be detected on top of the 19.3 *degrees* of advance per year produced by the Sun’s Newtonian perturbations. It is a remarkable feat of modern precision instrumentation that thanks to lunar laser ranging (a technique described in some detail in Box 13.2), the de Sitter precession has been measured to a precision of better than one percent.

## 10.2 Motion of light in post-Newtonian gravity

Most of the information about astrophysical objects comes in the form of electromagnetic signals, and confrontation between theory and observations must account for the curved path of a light ray in a gravitational field. Indeed, electromagnetic waves are deflected and delayed by a massive body, and these measurable effects must be taken into account in high-accuracy astronomical observations.

### 10.2.1 Motion of a photon

In the geometric-optics approximation of electromagnetism (refer to Box 5.6), light rays behave as massless particles – photons – that move on null geodesics of a curved spacetime; the geodesic equation for a photon in a post-Newtonian spacetime was obtained in Sec. 8.1.4. The particle moves on a trajectory  $\mathbf{r}(t)$  with a velocity  $\mathbf{v} = d\mathbf{r}/dt$  that can be expressed as

$$\mathbf{v} = c \left( 1 - \frac{2}{c^2} U \right) \mathbf{n} + O(c^{-3}), \quad (10.45)$$

in terms of a unit vector  $\mathbf{n}$  that specifies the direction of propagation. This satisfies the differential equation

$$\frac{dn^j}{dt} = \frac{2}{c} (\delta^{jk} - n^j n^k) \partial_k U + O(c^{-3}), \quad (10.46)$$

in which  $U$  is the Newtonian potential evaluated at  $\mathbf{x} = \mathbf{r}(t)$ . It should be noted that in this section,  $\mathbf{r}(t)$  stands for the photon’s trajectory and not the interbody separation of a two-body system, and  $\mathbf{n}$  denotes the direction of propagation and not the vector  $\mathbf{r}/r$ .

The leading-order solution to Eq. (10.46) is  $\mathbf{n} = \mathbf{k} + O(c^{-2})$ , in which  $\mathbf{k}$  is a constant vector. At this order, the photon’s trajectory is described by the straight path

$$\mathbf{r}(t) = \mathbf{r}_e + c\mathbf{k}(t - t_e) + O(c^{-2}), \quad (10.47)$$

in which  $\mathbf{r}_e = \mathbf{r}(t = t_e)$  is the position of the source, and  $t_e$  is the emission time. At the next order we have

$$\mathbf{n} = \mathbf{k} + \boldsymbol{\alpha} + O(c^{-4}), \quad (10.48)$$

in which the deflection vector  $\alpha$  satisfies

$$\frac{d\alpha^j}{dt} = \frac{2}{c}(\delta^{jk} - k^j k^k)\partial_k U, \quad (10.49)$$

with  $\partial_k U$  evaluated at  $\mathbf{x} = \mathbf{r}(t)$ , as given by Eq. (10.47). We integrate Eq. (10.49) with the initial conditions

$$\alpha(t = t_e) = \mathbf{0}, \quad (10.50)$$

so that  $\mathbf{n}(t = t_e)$ , the initial direction of propagation, coincides with the vector  $\mathbf{k}$ .

Inserting Eq. (10.48) within Eq. (10.45), we find that

$$\mathbf{v} = c\left(1 - \frac{2}{c^2}U\right)\mathbf{k} + c\alpha + O(c^{-3}). \quad (10.51)$$

Note that  $\alpha$  is necessarily orthogonal to  $\mathbf{k}$ , so that the second term in Eq. (10.51) describes a transverse deflection of the photon; the longitudinal aspects of the correction are captured by the first term.

To integrate Eq. (10.49) we substitute  $\partial_k U = -G \int \rho' s^{-3} s^k d^3x'$  on the right-hand side, where  $\rho' := \rho(t, \mathbf{x}')$  is the mass density of the matter distribution,  $\mathbf{s} := \mathbf{r}(t) - \mathbf{x}'$ , and  $s := |\mathbf{s}|$ . This gives

$$\frac{d\alpha}{dt} = -\frac{2G}{c} \int \rho' \frac{\mathbf{b}}{s^3} d^3x', \quad (10.52)$$

in which

$$\mathbf{b} := \mathbf{s}_e - (\mathbf{s}_e \cdot \mathbf{k})\mathbf{k}, \quad \mathbf{s}_e := \mathbf{r}_e - \mathbf{x}'; \quad (10.53)$$

the vector  $\mathbf{b}$  points from  $\mathbf{x}'$  to the point of closest approach reached by a photon emitted from a position  $\mathbf{r}_e$  in a direction  $\mathbf{k}$ . We next invoke the easily established identity

$$\frac{1}{c} \frac{d}{dt} \left( \frac{\mathbf{s} \cdot \mathbf{k}}{s} \right) = \frac{b^2}{s^3} \quad (10.54)$$

to express  $d\alpha/dt$  as

$$\frac{d\alpha}{dt} = -\frac{2G}{c^2} \int \rho' \frac{\mathbf{b}}{b^2} \frac{d}{dt} \left( \frac{\mathbf{s} \cdot \mathbf{k}}{s} \right) d^3x' = -\frac{2G}{c^2} \frac{d}{dt} \int \rho' \frac{\mathbf{b}}{b^2} \frac{\mathbf{s} \cdot \mathbf{k}}{s} d^3x'. \quad (10.55)$$

Integration of this equation is immediate.

Taking into account Eq. (10.50), we find that the deflection vector is given by

$$\alpha(t) = -\frac{2G}{c^2} \int \rho(t, \mathbf{x}') \frac{\mathbf{b}}{b^2} \left( \frac{\mathbf{s} \cdot \mathbf{k}}{s} - \frac{\mathbf{s}_e \cdot \mathbf{k}}{s_e} \right) d^3x' \quad (10.56)$$

for any matter distribution. Further progress on evaluating the deflection vector requires specification of the mass density.

With the photon's velocity  $\mathbf{v}$  given by Eqs. (10.51) and (10.56), we may now determine the photon's trajectory  $\mathbf{r}(t)$ . We express the solution to  $d\mathbf{r}/dt = \mathbf{v}$  as

$$\mathbf{r}(t) = \mathbf{r}_e + c\mathbf{k}(t - t_e) + \mathbf{k} \delta r_{\parallel}(t) + \delta \mathbf{r}_{\perp}(t) + O(c^{-4}), \quad (10.57)$$

in which  $\delta r_{\parallel}(t)$  is the longitudinal displacement determined by  $d(\delta r_{\parallel})/dt = -2U/c$ , and  $\delta r_{\perp}(t)$  is the transverse displacement determined by  $d(\delta r_{\perp})/dt = c\boldsymbol{\alpha}$ . The longitudinal term in Eq. (10.57) does not alter the path of the photon with respect to its unperturbed description, but it affects the relationship between position and time. The transverse term represents a deviation from the path, in response to the deflection vector  $\boldsymbol{\alpha}$ .

To calculate the longitudinal displacement we insert the usual expression for the Newtonian potential and get  $d(\delta r_{\parallel})/dt = -(2G/c) \int \rho' s^{-1} d^3x'$ . The identity

$$\frac{1}{c} \frac{d}{dt} \ln(s + \mathbf{s} \cdot \mathbf{k}) = \frac{1}{s} \quad (10.58)$$

permits an immediate integration, and we obtain

$$\delta r_{\parallel}(t) = -\frac{2G}{c^2} \int \rho(t, \mathbf{x}') \ln\left(\frac{s + \mathbf{s} \cdot \mathbf{k}}{s_e + \mathbf{s}_e \cdot \mathbf{k}}\right) d^3x', \quad (10.59)$$

which reflects the initial condition  $\delta r_{\parallel}(t = 0) = 0$ . The factorization  $b^2 = (s_e - \mathbf{s}_e \cdot \mathbf{k})(s_e + \mathbf{s}_e \cdot \mathbf{k})$  allows us to write this in the alternative form

$$\delta r_{\parallel}(t) = -\frac{2G}{c^2} \int \rho(t, \mathbf{x}') \ln\left[\frac{(s + \mathbf{s} \cdot \mathbf{k})(s_e - \mathbf{s}_e \cdot \mathbf{k})}{b^2}\right] d^3x' \quad (10.60)$$

involving the vector  $\mathbf{b}$ .

To obtain the transverse displacement we invoke Eq. (10.56) and get

$$\frac{d}{dt} \delta \mathbf{r}_{\perp} = -\frac{2G}{c} \int \rho(t, \mathbf{x}') \frac{\mathbf{b}}{b^2} \left( \frac{\mathbf{s} \cdot \mathbf{k}}{s} - \frac{\mathbf{s}_e \cdot \mathbf{k}}{s_e} \right) d^3x'. \quad (10.61)$$

Once more integration is immediate thanks to the identity  $c^{-1} ds/dt = \mathbf{s} \cdot \mathbf{k}/s$  and the fact that the second term within brackets is constant in time. We arrive at

$$\delta \mathbf{r}_{\perp}(t) = -\frac{2G}{c^2} \int \rho(t, \mathbf{x}') \frac{\mathbf{b}}{b^2} \left( s - \frac{\mathbf{s} \cdot \mathbf{s}_e}{s_e} \right) d^3x', \quad (10.62)$$

which also reflects the initial condition  $\delta \mathbf{r}_{\perp}(t = 0) = \mathbf{0}$ . The motion of the photon in the post-Newtonian spacetime is now completely determined.

### 10.2.2 Deflection by a spherical body

The simplest application of light deflection involves a single spherically-symmetric body of mass  $M$ , which we place at the spatial origin of the coordinate system. Because the photon must travel outside the body to be observable, the gravitational potential  $U$  can be equated to its external expression  $GM/|\mathbf{x}|$ , and Eq. (10.49) can be integrated for this special case. Alternatively, and more simply, we can insert  $\rho(\mathbf{x}') = M\delta(\mathbf{x}')$  within Eq. (10.56) and get

$$\boldsymbol{\alpha}(t) = -\frac{2GM}{c^2} \frac{\mathbf{b}}{b^2} \left[ \frac{\mathbf{r}(t) \cdot \mathbf{k}}{r(t)} - \frac{\mathbf{r}_e \cdot \mathbf{k}}{r_e} \right], \quad (10.63)$$

in which  $\mathbf{r}(t) := \mathbf{r}_e + c\mathbf{k}(t - t_e) + O(c^{-2})$ , and

$$\mathbf{b} := \mathbf{r}_e - (\mathbf{r}_e \cdot \mathbf{k})\mathbf{k} \quad (10.64)$$

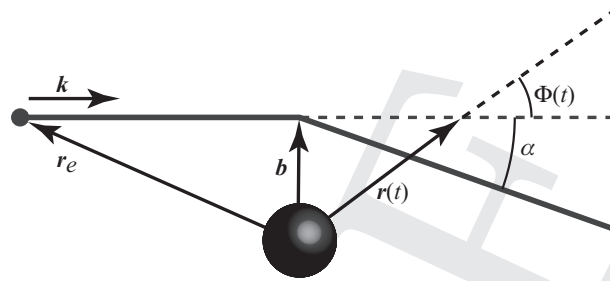


Fig. 10.1 Deflection of light by a spherical body.

is now a constant vector that points from the body's center-of-mass to the photon's point of closest approach (see Fig. 10.1);  $b := |\mathbf{b}|$  is the impact parameter. Note that the deflection vector always points in the direction of  $-\mathbf{b}$ , corresponding to a deflection toward the massive body, and varies in magnitude in response to the change in the photon's position.

The deflection vector can also be expressed as

$$\boldsymbol{\alpha}(t) = -\frac{2GM}{c^2} \frac{\mathbf{b}}{b^2} \left[ \cos \Phi(t) + \sqrt{1 - (b/r_e)^2} \right], \quad (10.65)$$

in which  $\cos \Phi(t) := (\mathbf{r} \cdot \mathbf{k})/r$ , so that  $\Phi(t)$  is the angle between the photon's current position  $\mathbf{r}(t)$  and its initial direction  $\mathbf{k}$  (see Fig. 10.1), and the square root is an alternative expression for  $-(\mathbf{r}_e \cdot \mathbf{k})/r_e$ . If we specialize to a situation in which the distance of closest approach  $b$  is very much smaller than the distance to the source  $r_e$ , then the deflection vector simplifies to

$$\boldsymbol{\alpha}(t) = -\frac{4GM}{c^2} \frac{\mathbf{b}}{b^2} \frac{1 + \cos \Phi(t)}{2}. \quad (10.66)$$

Evaluating this as  $t \rightarrow \infty$ , long after the photon has passed the point of closest approach, we find that  $\mathbf{r}(t)$  becomes increasingly aligned with  $\mathbf{k}$ , so that  $\cos \Phi(t \rightarrow \infty) \rightarrow 1$  and

$$\boldsymbol{\alpha}(t \rightarrow \infty) = -\frac{4GM}{c^2} \frac{\mathbf{b}}{b^2}. \quad (10.67)$$

From Fig. 10.1 it is easy to see that the photon's total angle of deflection  $\alpha$  is given by  $\alpha \simeq \tan \alpha = |\mathbf{n} \cdot \mathbf{b}|/b / (\mathbf{n} \cdot \mathbf{k})$ , in which  $\mathbf{n} = \mathbf{k} + \boldsymbol{\alpha} + O(c^{-4})$  is evaluated at late times. This gives  $\alpha = |\boldsymbol{\alpha} \cdot \mathbf{b}|/b = |\boldsymbol{\alpha}|$ , and we find that

$$\alpha = \frac{4GM}{c^2 b}. \quad (10.68)$$

This is the famous *deflection angle* of a light ray passing near a spherical body, as first calculated by Einstein in 1915.

### 10.2.3 Measurement of light deflection

The preceding results indicate that light is indeed deflected by a massive body, but they do not yet offer a means to put the prediction to a test. The reason is that while the final direction of propagation  $\mathbf{k} + \boldsymbol{\alpha}$  of a photon can easily be measured by a telescope, the initial



direction  $\mathbf{k}$  is unknown, and the extraction of  $\boldsymbol{\alpha}$  from the observations is impossible. To perform a test it is necessary to manipulate the equations to produce a relationship between quantities that can be measured directly. The way to proceed is to involve two sources of light, the first a *reference source* and the second a *target source*, which represents the emitter of interest. The angle  $\theta$  between the two light rays, when they are received simultaneously at a telescope or a radio interferometer, is a measurable quantity. It can be given a precise mathematical expression independent of any coordinate system, and it is this relationship that can be tested by observations.

In the following we assume that all measurements are carried out by an observer at rest in the post-Newtonian spacetime. In reality, of course, observations would be performed by an astronomer on the moving Earth, and her motion would have to be incorporated in the analysis of the measurement; we ignore such effects here. The observer has a spacetime velocity vector  $u^\alpha$ , and her reference frame is spanned by the spatial unit vectors  $e_{(j)}^\alpha$  first introduced in Sec. 4.1.5; the label  $(j)$  runs from (1) to (3). The vectors are mutually orthogonal, they are also orthogonal to  $u^\alpha$ , and they satisfy the identity

$$P^{\alpha\beta} := g^{\alpha\beta} + u^\alpha u^\beta / c^2 = e_{(1)}^\alpha e_{(1)}^\beta + e_{(2)}^\alpha e_{(2)}^\beta + e_{(3)}^\alpha e_{(3)}^\beta. \quad (10.69)$$

The tensor  $P_\beta^\alpha$  projects any vector  $A^\beta$  in the directions orthogonal to  $u^\alpha$ , which define the observer's reference frame; it was first introduced in Sec. 4.1.5 in the context of flat spacetime, and this is its incarnation in a curved spacetime with metric  $g_{\alpha\beta}$ .

Let the target source emit a photon with velocity vector  $v^\alpha = (c, \mathbf{v})$ , as given by Eq. (10.51), and let the reference source emit another photon with velocity vector  $v'^\alpha = (c, \mathbf{v}')$ . With no loss of generality we orient the vectorial basis  $e_{(j)}^\alpha$  in such a way that both photons move in the  $x$ - $y$  plane of the observer's reference frame. As shown back in Sec. 4.1.6, see Eq. (4.28), the angle  $\phi$  made by the target photon with respect to the observer's  $x$ -axis is determined by

$$\cos \phi = c \frac{v_\alpha e_{(1)}^\alpha}{-v_\beta u^\beta}, \quad \sin \phi = c \frac{v_\alpha e_{(2)}^\alpha}{-v_\beta u^\beta}. \quad (10.70)$$

Similarly, the angle  $\phi'$  made by the reference photon is given by

$$\cos \phi' = c \frac{v'_\alpha e_{(1)}^\alpha}{-v'_\beta u^\beta}, \quad \sin \phi' = c \frac{v'_\alpha e_{(2)}^\alpha}{-v'_\beta u^\beta}. \quad (10.71)$$

The angle between the two photons is  $\theta := \phi - \phi'$ , and this is given by

$$\cos \theta = \cos \phi \cos \phi' + \sin \phi \sin \phi' = \frac{v_\alpha v'_\beta (e_{(1)}^\alpha e_{(1)}^\beta + e_{(2)}^\alpha e_{(2)}^\beta)}{v_\mu v'_\nu (u^\mu u^\nu / c^2)}. \quad (10.72)$$

Because  $v_\alpha e_{(3)}^\alpha = 0 = v'_\alpha e_{(3)}^\alpha$ , the quantity within brackets can be replaced by the projector of Eq. (10.69), and we arrive at

$$\cos \theta = 1 + \frac{g_{\alpha\beta} v^\alpha v'^\beta}{v^\mu v'^\nu (u_\mu u_\nu / c^2)}, \quad (10.73)$$

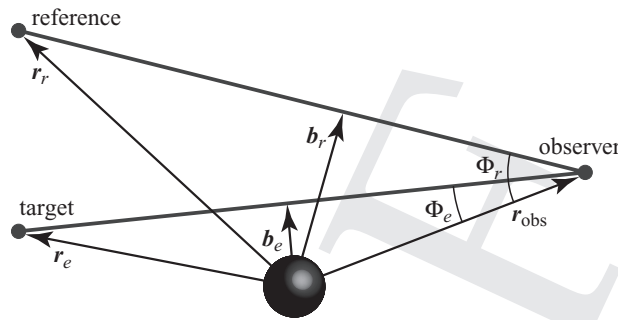


Fig. 10.2 Geometry of light deflection measurements.

our final expression for the relative angle. Note that the right-hand side of Eq. (10.73) is a spacetime invariant that can be evaluated in any coordinate system; this guarantees that  $\theta$  is a precisely defined, observable quantity.

Our next task is to evaluate Eq. (10.73) for our situation, which involves the post-Newtonian spacetime, the observer at rest, and the target and reference photons. The relevant components of the metric tensor are  $g_{00} = -(1 - 2U/c^2) + O(c^{-4})$ ,  $g_{jk} = (1 + 2U/c^2)\delta_{jk} + O(c^{-4})$ , the only non-vanishing component of the observer's velocity vector is  $u^0 = c(1 + U/c^2) + O(c^{-3})$ , and the photon's velocities are given by  $\mathbf{v} = c(1 - 2U/c^2)\mathbf{k} + c\boldsymbol{\alpha} + O(c^{-3})$  and  $\mathbf{v}' = c(1 - 2U/c^2)\mathbf{k}' + c\boldsymbol{\alpha}' + O(c^{-3})$ . Inserting all this within Eq. (10.73) gives

$$\cos \theta = \mathbf{k} \cdot \mathbf{k}' + \mathbf{k}' \cdot \boldsymbol{\alpha} + \mathbf{k} \cdot \boldsymbol{\alpha}' + O(c^{-4}). \quad (10.74)$$

The relative angle is thus given a simple expression in terms of  $\mathbf{k}$  and  $\boldsymbol{\alpha}$ , the initial direction and deflection of the target photon, as well as  $\mathbf{k}'$  and  $\boldsymbol{\alpha}'$ , the initial direction and deflection of the reference photon.

We next insert Eq. (10.66) within Eq. (10.74) and obtain

$$\cos \theta = \cos \theta_0 - \frac{4MG}{c^2} \frac{\mathbf{k}' \cdot \mathbf{b}_e}{b_e^2} \frac{1 + \cos \Phi_e}{2} - \frac{4MG}{c^2} \frac{\mathbf{k} \cdot \mathbf{b}_r}{b_r^2} \frac{1 + \cos \Phi_r}{2}, \quad (10.75)$$

in which  $\cos \theta_0 := \mathbf{k} \cdot \mathbf{k}'$ , so that  $\theta_0$  is the relative angle in the absence of deflection,  $\cos \Phi_e := \mathbf{r}_{\text{obs}} \cdot \mathbf{k}/r_{\text{obs}}$ , so that  $\Phi_e$  is the angle between the observer's position  $\mathbf{r}_{\text{obs}}$  and the direction  $\mathbf{k}$  of the target photon, and  $\cos \Phi_r := \mathbf{r}_{\text{obs}} \cdot \mathbf{k}'/r_{\text{obs}}$ , so that  $\Phi_r$  is the angle between the observer's position and the direction of the reference photon (see Fig. 10.2); we also have  $\mathbf{b}_e := \mathbf{r}_e - (\mathbf{r}_e \cdot \mathbf{k})\mathbf{k}$ , where  $\mathbf{r}_e$  is the position of the target source (the emitter), and  $\mathbf{b}_r := \mathbf{r}_r - (\mathbf{r}_r \cdot \mathbf{k}')\mathbf{k}'$ , where  $\mathbf{r}_r$  is the position of the reference source.

Our expression for  $\cos \theta$  can be cleaned up if we express  $\mathbf{k}' \cdot \mathbf{b}_e$  and  $\mathbf{k} \cdot \mathbf{b}_r$  in terms of the angles  $\Phi_e$ ,  $\Phi_r$ , and  $\theta_0$ . To accomplish this we note that since the vector  $\mathbf{r}_{\text{obs}} - \mathbf{r}_e$  is directed along  $\mathbf{k}$  (apart from the small correction from  $\boldsymbol{\alpha}$ ), we can write  $\mathbf{b}_e = \mathbf{r}_{\text{obs}} - (\mathbf{r}_{\text{obs}} \cdot \mathbf{k})\mathbf{k}$ ; similarly we have that  $\mathbf{b}_r = \mathbf{r}_{\text{obs}} - (\mathbf{r}_{\text{obs}} \cdot \mathbf{k}')\mathbf{k}'$ . These expressions imply  $\mathbf{k}' \cdot \mathbf{b}_e = r_{\text{obs}}(\cos \Phi_r - \cos \Phi_e \cos \theta_0)$  and  $\mathbf{k} \cdot \mathbf{b}_r = r_{\text{obs}}(\cos \Phi_e - \cos \Phi_r \cos \theta_0)$ . Noting in addition

that  $\sin \Phi_e = b_e/r_{\text{obs}}$  and  $\sin \Phi_r = b_r/r_{\text{obs}}$ , as can be gleaned from Fig. 10.2, we arrive at

$$\begin{aligned} \cos \theta = \cos \theta_0 - \frac{4MG}{c^2 b_e} \left( \frac{\cos \Phi_r - \cos \Phi_e \cos \theta_0}{\sin \Phi_e} \right) \frac{1 + \cos \Phi_e}{2} \\ - \frac{4MG}{c^2 b_r} \left( \frac{\cos \Phi_e - \cos \Phi_r \cos \theta_0}{\sin \Phi_r} \right) \frac{1 + \cos \Phi_r}{2}. \end{aligned} \quad (10.76)$$

The difference between the unperturbed angle  $\theta_0$  and the measured angle  $\theta$  is denoted  $\delta\theta$ , and since  $\cos \theta = \cos(\theta_0 + \delta\theta) \approx \cos \theta_0 - \sin \theta_0 \delta\theta$ , we find that

$$\begin{aligned} \delta\theta = \frac{4MG}{c^2 b_e} \left( \frac{\cos \Phi_r - \cos \Phi_e \cos \theta_0}{\sin \Phi_e \sin \theta_0} \right) \frac{1 + \cos \Phi_e}{2} \\ + \frac{4MG}{c^2 b_r} \left( \frac{\cos \Phi_e - \cos \Phi_r \cos \theta_0}{\sin \Phi_r \sin \theta_0} \right) \frac{1 + \cos \Phi_r}{2}. \end{aligned} \quad (10.77)$$

We recall that  $\theta_0$  is the angle subtended by the target and reference sources, that  $\Phi_e$  is the angle subtended by the target source and the deflecting body, and that  $\Phi_r$  is the angle subtended by the reference source and the deflecting body. In our discussion so far, these angles have been defined by the unperturbed, straight motion of the light rays in flat spacetime, as depicted in Fig. 10.2. But since the angles occur within a post-Newtonian expression, and the difference between the perturbed and unperturbed angles is itself of post-Newtonian order, we can actually interpret  $\theta_0$ ,  $\Phi_e$ , and  $\Phi_r$  in Eq. (10.77) as if they were the *observed angles* on the sky; the difference manifests itself at second post-Newtonian order only, and is negligible.

**Box 10.2**

**Spherical trigonometry**

Angular measurements in astronomy are performed on the celestial sphere, a fictitious sphere of large radius centered at the position of the observer, on which all astronomical bodies are imagined to be situated. Relationships between angles are then clarified with the rules of spherical trigonometry.

For the angles shown in Fig. 10.3, we have the relations

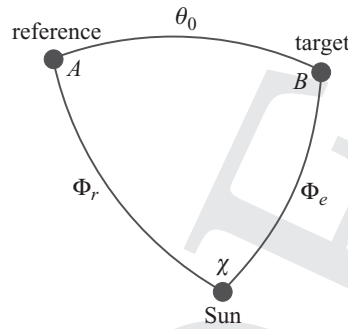
$$\begin{aligned} \cos \theta_0 &= \cos \Phi_r \cos \Phi_e + \sin \Phi_r \sin \Phi_e \cos \chi, \\ \cos \Phi_r &= \cos \theta_0 \cos \Phi_e + \sin \theta_0 \sin \Phi_e \cos B, \\ \cos \Phi_e &= \cos \theta_0 \cos \Phi_r + \sin \theta_0 \sin \Phi_r \cos A, \end{aligned}$$

and

$$\frac{\sin \Phi_e}{\sin A} = \frac{\sin \Phi_r}{\sin B} = \frac{\sin \theta_0}{\sin \chi}.$$

With these we find that Eq. (10.77) simplifies to

$$\delta\theta = \frac{4MG}{c^2 b_e} \cos B \frac{1 + \cos \Phi_e}{2} + \frac{4MG}{c^2 b_r} \cos A \frac{1 + \cos \Phi_r}{2}.$$



**Fig. 10.3** Angles of sources on the celestial sphere as seen from the Earth. When the Sun is behind the Earth, the point labelled “Sun” becomes the extension of the Sun–Earth line into the sky, and the angles  $\Phi_r$  and  $\Phi_e$  become  $\pi - \Phi_r$  and  $\pi - \Phi_e$ , respectively.

This equation states that the change in angular separation between the target and reference sources is just the sum of the apparent displacement of each image away from the deflecting body, projected along the line joining the reference and target sources.

Equation (10.77) is what we need to put the prediction of light deflection to a test, except for the fact that while we can always measure  $\theta$  for the target and reference sources, we still do not have access to  $\theta_0$ , its unperturbed value. To get around this, all measurements of the light deflection are carried out *differentially*. In this method, the angle between the target and reference stars is measured at different times, first when the Sun is nowhere in the vicinity of the stars, and again when the light from the target star passes near the Sun. The first measurement yields the unperturbed angle, which does not change with time unless the stars have significant proper motion. The second measurement yields the deflected angle  $\theta$ , and a measurement of  $\delta\theta = \theta - \theta_0$  can finally be compared with Eq. (10.77).

Equation (10.77) simplifies a little when the target source is very close to the Sun as seen from Earth, so that  $\Phi_e \ll 1$ . An expansion of  $\cos \theta_0 = \cos \Phi_r \cos \Phi_e + \sin \Phi_r \sin \Phi_e \cos \chi$ , see Box 10.2, in powers of  $\Phi_e$  reveals that

$$\theta_0 = \Phi_r - \cos \chi \Phi_e + \frac{\cos \Phi_r \sin^2 \chi}{2 \sin \Phi_r} \Phi_e^2 + O(\Phi_e^3), \quad (10.78)$$

and substitution within Eq. (10.77) produces

$$\delta\theta = \frac{4GM}{c^3 b_e} \left[ -\cos \chi + \frac{1 + (1 + 2 \sin^2 \chi) \cos \theta_0}{2 \sin \theta_0} \Phi_e + O(\Phi_e^2) \right], \quad (10.79)$$

in which  $\chi$  is the angle subtended by the reference and target sources as seen from the Sun (as shown in Fig. 10.3). This expression reveals clearly how the angular separation between the target and reference sources changes with time as the Sun moves across the sky, causing  $\chi$  and  $\Phi_e$  to vary with time. Inserting  $b_e = r_{\text{obs}} \sin \Phi_e = r_{\text{obs}} \Phi_e + O(\Phi_e^3)$  reveals also that the theoretical prediction can be expressed entirely in terms of observable quantities.

The first successful measurement of the bending of light by the Sun was carried out by British astronomer Arthur Stanley Eddington and his colleagues during the total solar eclipse of May 29, 1919. Two expeditions were sent out to measure the eclipse, one to Brazil, the other to the small island of Principe, off the west coast of Africa. In both cases it was a differential measurement: Photographs of the stars near the Sun taken during the eclipse were compared with photographs of the same stars taken at night from the same locations later in the year, and the changes in angles between pairs of stars were carefully measured. Eddington's announcement in November 1919 that the bending measurements were in agreement with general relativity helped make Einstein an international celebrity. The observations, however, had an accuracy of approximately 30 percent, and succeeding eclipse measurements were not much better. The results were scattered between one half and twice the Einstein prediction, but in spite of such limited success, Einstein was declared victorious.

The subsequent development of long-baseline radio interferometry greatly improved the measurement of the light deflection. These techniques now have the capability of measuring angular separations and changes in angles to accuracies of tens of micro-arcseconds. Early measurements took advantage of the fact that a few quasistellar radio sources – quasars – pass very close to the Sun as seen from the Earth. As the Earth moves in its orbit, changing the lines of sight of the quasars relative to the Sun, the angular separation  $\delta\theta$  between pairs of quasars varies. The time evolution of  $\chi$  and  $\Phi_e$  in Eq. (10.79) is determined using an accurate ephemeris for the Earth and initial directions for the quasars, and the resulting prediction for  $\delta\theta$  as a function of time is compared with the measured values. A number of measurements of this kind over the period 1969–1975 yielded results in agreement with general relativity to a few parts in  $10^3$ . In recent years, transcontinental and intercontinental VLBI observations of quasars and radio galaxies have been made primarily to monitor the Earth's rotation and to establish a highly accurate reference frame for astronomy and navigation. These measurements are sensitive to the deflection of light over almost the entire celestial sphere. For example, the deflection of a ray approaching the Earth from a direction  $90^\circ$  away from the Sun is 4 milli-arcseconds, easily measurable by modern VLBI techniques. A 2004 analysis of nearly 2 million VLBI observations of 541 radio sources, made by 87 VLBI sites over a 20-year period, verified Einstein's prediction to a few parts in  $10^4$ . Analysis of observations made by the Hipparcos optical astrometry satellite yielded a test at the level of 0.3 percent, and a future orbiting observatory named GAIA will have the capability of testing the deflection to parts per million.

**Box 10.3****The "Newtonian" deflection of light**

At the time of Eddington's attempt to measure the deflection of light, people envisioned three possible outcomes for the experiment: no deflection, the Einsteinian deflection, or one half of Einstein's prediction, commonly called the "Newtonian" deflection. The Newtonian deflection can be derived in a variety of ways. One way is to assume that light behaves as a particle, to recall that the trajectory of a particle is independent of its mass (weak equivalence principle), and to calculate the deflection of its trajectory in the limit in which the particle's speed approaches the speed of light. Such an approach would have made sense in Newton's day,

when light was really viewed as a “corpuscle,” and indeed, Newton himself speculated on the possible effect of gravity on light.

The English physicist Henry Cavendish may have been the first person to calculate the bending explicitly, possibly as early as 1784, although evidence for this was not discovered until around 1914, during an effort to compile and publish his entire body of work – publication never being high on Cavendish’s list of priorities. In fact, all that was found was a scrap of paper in Cavendish’s handwriting stating that he had done the calculation, and giving the answer.

Independently of Cavendish, the Bavarian astronomer Johann von Soldner did publish in 1803 a detailed calculation of the Newtonian bending in a German astronomical journal. Strangely, von Soldner’s calculation was largely forgotten until it was resurrected in 1921 by Phillip Lenard as part of a campaign to discredit the “Jewish” relativity of Einstein by publicizing the earlier work of the “Aryan” von Soldner. Apparently, Lenard was not deterred by the fact that the 1919 observations actually favored general relativity over the Newtonian deflection.

Unaware of the earlier work, Einstein himself derived the “Newtonian” deflection in 1911. He argued, as we have back in Sec. 5.1.3, that gravity requires replacement of the Minkowski metric of flat spacetime by the Newtonian metric of Eq. (5.12),

$$ds^2 = -(1 - 2U/c^2)d(ct)^2 + dx^2 + dy^2 + dz^2.$$

Geodesic motion for a test particle in this spacetime reproduces Newtonian gravity, and geodesic motion for a photon gives the Newtonian deflection. Another derivation using only the equivalence principle imagines a sequence of freely falling frames through which a light ray passes as it travels near a gravitating body. Each frame is momentarily at rest at the moment the light ray enters it. Although the path of the ray is a straight line within each frame, the frame picks up a downward velocity during the ray’s traversal, because of the body’s gravitational attraction. When the adjacent frame receives the light ray, it is deflected toward the body because the downward motion of the previous frame induces aberration on the received ray. By adding up all the tiny aberrations over a sequence of frames, one arrives at the Newtonian deflection.

But the full theory of general relativity doubles the deflection, because the spatial part of the metric now comes with the multiplying factor  $(1 + 2U/c^2)$ . This represents spatial curvature, which could not be taken into account either by Newtonian gravity or by the principle of equivalence. So the total deflection can be viewed as a sum of a Newtonian deflection relative to locally straight lines, plus the bending of locally straight lines relative to straight lines at infinity; each effect contributes exactly half the total deflection.

### 10.2.4 Gravitational lenses

The deflection of light has become a cornerstone of the empirical edifice that supports general relativity. But in 1979 the phenomenon became much more than that. That year, astronomers Dennis Walsh, Robert Carswell and Ray Weymann discovered the “double quasar” Q0957+561, which consisted of two quasar images about 6 arcseconds apart, with almost the same redshift ( $z = 1.41$ ) and very similar spectra. Given that quasars are

thought to be among the most distant objects in the universe, the probability of finding two so close together was low. It was soon realized that there was in fact just one quasar, but that intervening matter in the form of a galaxy or a cluster of galaxies was bending the light from the quasar and producing two separate images.

Since then, over 60 lensed quasars have been discovered. But more importantly, gravitational lensing has become a major tool in efforts to map the distribution of mass around galaxies and clusters, and in searches for dark matter, dark energy, compact objects, and extrasolar planets. Many subtopics of gravitational lensing have been developed to cover different astronomical realms: microlensing for the search for dim compact objects and extra-solar planets, the use of luminous arcs to map the distribution of mass and dark matter, and weak lensing to measure the properties of dark energy. Lensing has to be taken into account in interpreting certain aspects of the cosmic microwave background radiation, and in extracting information from gravitational waves emitted by sources at cosmological distances. These topics are beyond the scope of this book, but we can extend our discussion of the deflection of light to provide many of the basic concepts and results.

### Deflection vector

We consider light rays emitted by a remote source like a quasar to be observed by an astronomer on Earth. The rays pass through a distribution of matter – the gravitational lens – on their way to the observer, and undergo a deflection described by the vector  $\alpha$  of Eq. (10.56). In this situation the source is almost directly behind the lens, so that  $s_e \cdot \mathbf{k}/s_e \simeq -1$ , the observer is almost directly in front of the lens, so that  $s_{\text{obs}} \cdot \mathbf{k}/s_{\text{obs}} \simeq 1$ , and the deflection vector can be simplified to

$$\alpha = -\frac{4G}{c^2} \int \rho(\mathbf{x}') \frac{\mathbf{b}}{b^2} d^3x', \quad (10.80)$$

in which  $\mathbf{b} := s_e - (s_e \cdot \mathbf{k})\mathbf{k}$ , where  $s_e := \mathbf{r}_e - \mathbf{x}'$ ; here  $\mathbf{r}_e$  is the position of the source relative to the center-of-mass of the matter distribution, and  $\mathbf{k}$  is the ray's initial direction; its final direction as measured by the observer is  $\mathbf{k} + \alpha$ . We have assumed that the distribution of matter is time independent, so that  $\rho$  depends on the spatial variables  $\mathbf{x}'$  only.

To express the deflection vector in a more convenient form we introduce the vectors

$$\xi := \mathbf{r}_e - (\mathbf{r}_e \cdot \mathbf{k})\mathbf{k}, \quad \xi' := \mathbf{x}' - (\mathbf{x}' \cdot \mathbf{k})\mathbf{k}, \quad (10.81)$$

so that  $\xi$  is the projection of the source's position  $\mathbf{r}_e$  in the *lens plane*, defined as the plane perpendicular to the direction of propagation  $\mathbf{k}$  ( $\xi$  is also the photon's point of closest approach to the center-of-mass of the matter distribution), and similarly,  $\xi'$  is the projection of  $\mathbf{x}'$  in the lens plane. In terms of these vectors we have that  $\mathbf{b} = \xi - \xi'$ , and the deflection vector becomes

$$\alpha = -\frac{4G}{c^2} \int \rho(\mathbf{x}') \frac{\xi - \xi'}{|\xi - \xi'|^2} d^3x'. \quad (10.82)$$

To simplify this further we implement a coordinate transformation from the old system  $\mathbf{x}'$  to a new system  $(\xi', \ell')$ , in which  $\ell' := \mathbf{x}' \cdot \mathbf{k}$  is the distance along the line of sight,

perpendicular to the lens plane. Integration over  $d\ell'$  involves  $\rho(\xi', \ell')$  only, and it gives rise to

$$\Sigma(\xi') := \int \rho(\xi', \ell') d\ell', \quad (10.83)$$

the projected, two-dimensional mass density (per unit area) of the matter distribution. Our final expression for the deflection vector is

$$\alpha(\xi) = -\frac{4G}{c^2} \int \Sigma(\xi') \frac{\xi - \xi'}{|\xi - \xi'|^2} d^2\xi'; \quad (10.84)$$

it reveals that  $\alpha$  is a function of  $\xi$  only, and that it lies within the lens plane. Alternative expressions for the deflection vector are formulated in Exercise 10.5.

The deflection vector acquires a particularly simple form when the matter distribution is axially symmetric, so that the surface density  $\Sigma$  depends only on the magnitude  $\xi'$  of the vector  $\xi'$ . In this case we can adopt a system of polar coordinates  $(\xi', \phi')$  in the lens plane, and carry out the integration over the angle  $\phi'$ . The result is

$$\alpha(\xi) = -\frac{4G}{c^2} \frac{m(\xi)}{\xi^2} \xi, \quad (10.85)$$

in which

$$m(\xi) := 2\pi \int_0^\xi \Sigma(\xi') \xi' d\xi' \quad (10.86)$$

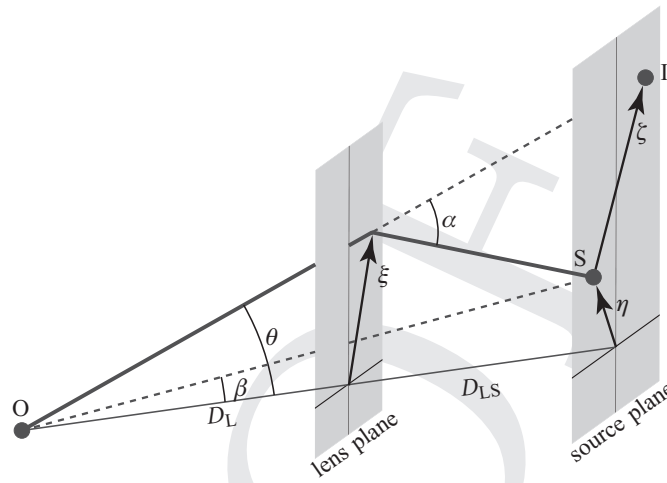
is the mass inside a circle of radius  $\xi$  in the lens plane. In this case the deflection vector necessarily points in the direction opposite to  $\xi$ .

### Lens equation

We now consider the situation depicted in Fig. 10.4. We have a source at a distance  $D_S$  from the observer, and a lens between the source and observer, at a distance  $D_L$  from the observer;  $D_{LS} := D_S - D_L$  is the distance between the lens and the source. The figure displays the lens plane, which is perpendicular to the initial direction of propagation of the light rays, and a source plane that contains the source, which is parallel to the lens plane. The figure also shows the optical axis, which passes through the observer and the center-of-mass of the lens; the optical axis is not necessarily perpendicular to the planes. The vector  $\eta$  gives the position of the source in the source plane, relative to the optical axis, and  $\zeta$  is the position of the image relative to the source; as usual  $\xi$  is the point of closest approach in the lens plane. We introduce the vector  $\beta := \eta/D_S$ , whose magnitude  $\beta$  is the angle between the source and the optical axis. We introduce also the vector  $\theta := (\eta + \zeta)/D_S$ , whose magnitude  $\theta$  is the angle between the image and the optical axis.

Our main goal is to determine  $\theta$  for a given  $\beta$ , and the key step is to recognize from Fig. 10.4 that the deflection vector  $\zeta$  can be expressed as  $-\alpha D_{LS}$ , where the negative sign accounts for the fact that  $\alpha$  points in the direction opposite to  $\xi$ . The definition of  $\theta$  then





**Fig. 10.4** Geometry of a gravitational lens. The observer is labeled O, the source S, and the image I.

implies  $\theta = (D_S\beta - D_{LS}\alpha)/D_S$ , or

$$\theta + \frac{D_{LS}}{D_S}\alpha = \beta. \quad (10.87)$$

This *lens equation* can be solved for  $\theta$  once we insert the deflection vector of Eq. (10.84), in which we substitute  $\xi = D_L\theta$ , another relation that can be inferred from Fig. 10.4. Because the deflection vector is a non-linear function of its argument, the lens equation typically admits more than one solution for a given source position  $\beta$ . To obtain these solutions it is necessary to know  $\alpha(\xi)$ , and this requires an evaluation of the integral in Eq. (10.84). This can be accomplished when the density profile of the lens is known, but this information is usually not available in an astronomical context. The power of gravitational lensing in astronomy resides in the fact that the lens equation can be turned around: the measured positions of the multiple images of a given source can be used, through the deflection vector  $\alpha$ , to deduce some features of the mass distribution.

### Schwarzschild lens

The simplest instance of a gravitational lens is one created by a single, spherically-symmetric body of mass  $M$ . In this case the deflection vector is given by Eq. (10.67), or is obtained by letting  $m(\xi) = M$  in Eq. (10.85). We have

$$\alpha(\xi) = -\frac{4GM}{c^2} \frac{\xi}{\xi^2}, \quad (10.88)$$

and substitution within Eq. (10.87) gives rise to

$$\theta - \frac{\theta_E^2}{\theta} = \beta, \quad (10.89)$$

in which

$$\theta_E^2 := \frac{4GM}{c^2} \frac{D_{LS}}{D_S D_L}. \quad (10.90)$$

In this case the lens equation reduces to a scalar equation, because by virtue of the axial symmetry of the lens (which follows from its spherical symmetry), all vectors point in the same direction. The parameter  $\theta_E$  is known as the *Einstein angle*, and the corresponding length scale

$$\xi_E := D_L \theta_E = \sqrt{\frac{4GM}{c^2} \frac{D_{LS} D_L}{D_S}} \quad (10.91)$$

is the *Einstein radius*. The Einstein angle gives the characteristic scale of the deflection. For lenses of galactic scales with  $M \sim 10^{12} M_\odot$ ,  $\theta_E \simeq 1.8$  as, while for solar-mass lenses,  $\theta_E \simeq 0.5$  mas. The lens equation is consistent whenever  $\xi \gg 4GM/c^2$ , so that higher post-Newtonian corrections are not required, and this condition implies that  $\theta$  must be much larger than  $(D_S/D_{LS})\theta_E^2$ .

The solutions to the lens equations are

$$\theta_\pm = \frac{1}{2} \left( \beta \pm \sqrt{\beta^2 + 4\theta_E^2} \right), \quad (10.92)$$

and we see that the lens produces two images of the source. For large angles  $\beta \gg \theta_E$  the solutions reduce to  $\theta_+ = \beta + \theta_E^2/\beta + \dots$  and  $\theta_- = -\theta_E^2/\beta + \dots$ , which indicates that one image is displaced beyond the source, while the second image occurs below the optical axis and is closer to the lens than is the source itself. When  $\beta$  increases beyond unity, the second solution becomes smaller than  $\theta_E^2$  and enters a regime in which the lens equation is no longer consistent, because for this solution  $\xi = D_L \theta = -D_L \theta_E^2/\beta \sim (4GM/c^2)(D_{LS}/D_S)/\beta$ , which violates the condition  $\xi \gg 4GM/c^2$ . For small angles  $\beta \ll \theta_E$  the solutions reduce to  $\theta_+ = \theta_E + \frac{1}{2}\beta + \dots$  and  $\theta_- = -\theta_E + \frac{1}{2}\beta + \dots$ , which indicates that the images are at an angle approximately equal to  $\theta_E$  above and below the optical axis. When  $\beta = 0$ , that is, when the source is directly behind the lens, the deflection angle is precisely equal to  $\pm\theta_E$ , and the axial symmetry of the situation forces the image to take the shape of a ring – known as an *Einstein ring* – around the optical axis.

When the source has a non-zero angular size, the lens continues to displace its images, but there is also a distortion of its shape. Points on opposite sides of the source perpendicular to the optical axis are displaced by an angle  $\theta$ , and are therefore stretched by a corresponding factor, while points on either side parallel to the optical axis are stretched only by the difference in  $\theta$ . When the source is circular, and in the limit where its angular size  $\delta\beta$  is small compared to  $\beta$ , the image is an ellipse with an axis ratio  $a/b = \beta/\sqrt{\beta^2 + 4\theta_E^2}$ , with  $a$  denoting the semi-axis in the direction of the lens, and  $b$  the semi-axis in the perpendicular direction. For sources of larger angular size, the image can actually be distorted into an arc with a convex side; this occurs when

$$\delta\beta < \beta \quad \text{and} \quad \theta_E \geq \frac{\beta - \delta\beta}{2\delta\beta} \sqrt{\beta^2 - \delta\beta^2}. \quad (10.93)$$

The orientation and shapes of such luminous arcs have been used to deduce the mass distribution of the galaxies or clusters that act as lenses, a procedure sometimes dubbed *gravitational tomography*. Even when the lens produces elliptical distortions that are too small to be measured individually, there is a systematic effect, averaged over large collections of images, that is sensitive to the evolution of the universe over an epoch when dark energy began to be important; this is the realm of weak gravitational lensing.

Because the number of photons emitted per unit area and unit time is constant for a steady source, the brightness of the image is proportional to its observed area, which for small angles is proportional to  $\int \theta d\theta d\phi$ . And because the intrinsic area of the source is proportional to  $\int \beta d\beta d\phi$ , we find that each image is magnified by a factor

$$\mu_{\pm} = \frac{\theta_{\pm} d\theta_{\pm}}{\beta d\beta} = \pm \frac{1}{4} \left( \frac{\beta}{\sqrt{\beta^2 + 4\theta_E^2}} + \frac{\sqrt{\beta^2 + 4\theta_E^2}}{\beta} \pm 2 \right). \quad (10.94)$$

The term in parentheses is always greater than zero, and we see that  $\mu_+$  is positive, while  $\mu_-$  is negative, indicating that the image is inverted relative to the source. In addition, we see that  $\mu_+$  is always greater than unity, but that  $|\mu_-|$  is smaller than unity, indicating that the second image is actually demagnified by the lens.

In microlensing situations the images are too close together to be resolved individually by the observer, and in such cases the total magnification is measured by

$$|\mu_+| + |\mu_-| = \frac{1}{2} \left( \frac{\beta}{\sqrt{\beta^2 + 4\theta_E^2}} + \frac{\sqrt{\beta^2 + 4\theta_E^2}}{\beta} \right), \quad (10.95)$$

which is always greater than unity. The technique of monitoring the variable brightness of lensing images was used in a series of experiments to search for “massive compact halo objects” (MACHOs) in our galaxy. If the galaxy contained a population of dark objects (black holes, neutron stars, brown dwarf stars, or other exotic objects) with masses comparable to  $M_{\odot}$ , then the brightness of a star transiting behind such an object should behave in a way consistent with Eq. (10.95). This effect can be distinguished from the star’s own variability, or from the absorption of starlight by intervening matter, because these tend to depend on wavelength, while the lensing is independent of wavelength. Searches for dark objects passing in front of the dense field of stars in the Large Magellanic Cloud and in the galactic center were carried out between 1993 and 2007, placing a stringent upper limit on the amount of halo mass that could be made up of such objects. This strengthened the conclusion that the vast majority of the halo mass must be made of non-baryonic dark matter.

In 2003, an extra-solar planetary system was discovered by microlensing. The combined lensing of a distant source by a Jupiter-scale companion and its host star was measured and could be deconvolved to determine the mass ratio and the approximate distance between the planet and the star. Additional systems were discovered subsequently, and gravitational lensing is proving to be a key tool in the search for exoplanets.

### 10.2.5 Shapiro time delay

Our discussion of the motion of light in a gravitational field has so far emphasized the transverse aspects of the motion – the deflection of a light ray with respect to its unperturbed, straight path. The longitudinal aspects – the changed relationship between position and time – are also important, and we conclude this section with an examination of the delay suffered by a photon as it travels across a gravitational potential well. We consider a light source situated at  $\mathbf{r}_e$  relative to the center-of-mass of a spherical body of mass  $M$ , and an observer situated at  $\mathbf{r}_{\text{obs}}$ . We imagine that the body is situated between the source and the observer, near the line of sight, and that the photon is at a distance  $b$  from the body when it reaches its point of closest approach. We wish to calculate the light-travel time between the source and observer.

The information is contained in Eq. (10.57),

$$\mathbf{r}(t) = \mathbf{r}_e + c\mathbf{k}(t - t_e) + \mathbf{k}\delta r_{\parallel}(t) + \delta\mathbf{r}_{\perp}(t) + O(c^{-4}), \quad (10.96)$$

in which we may insert Eq. (10.60) for the longitudinal displacement  $\delta r_{\parallel}$ , and Eq. (10.62) for the transverse displacement  $\delta\mathbf{r}_{\perp}$ . The equation is evaluated at  $t = t_{\text{obs}}$  so that  $\mathbf{r}(t = t_{\text{obs}}) = \mathbf{r}_{\text{obs}}$ , and the displacements must be specialized to the case at hand, for which the mass density can be expressed as  $\rho(t, \mathbf{x}') = M\delta(\mathbf{x}')$ . We do not need an explicit expression for  $\delta\mathbf{r}_{\perp}$ , but we find that the longitudinal displacement becomes

$$\delta r_{\parallel} = -\frac{2GM}{c^2} \ln \left[ \frac{(r_{\text{obs}} + \mathbf{r}_{\text{obs}} \cdot \mathbf{k})(r_e - \mathbf{r}_e \cdot \mathbf{k})}{b^2} \right]. \quad (10.97)$$

The travel time  $t_{\text{obs}} - t_e$  is computed by squaring Eq. (10.96), which yields

$$|\mathbf{r}_{\text{obs}} - \mathbf{r}_e|^2 = c^2(t_{\text{obs}} - t_e)^2 + 2c(t_{\text{obs}} - t_e)\delta r_{\parallel} + O(c^{-4}), \quad (10.98)$$

or

$$c(t_{\text{obs}} - t_e) = |\mathbf{r}_{\text{obs}} - \mathbf{r}_e| - \delta r_{\parallel} + O(c^{-4}). \quad (10.99)$$

Inserting our expression for  $\delta r_{\parallel}$ , we arrive at

$$t_{\text{obs}} - t_e = \frac{1}{c}|\mathbf{r}_{\text{obs}} - \mathbf{r}_e| + \frac{2GM}{c^3} \ln \left[ \frac{(r_{\text{obs}} + \mathbf{r}_{\text{obs}} \cdot \mathbf{k})(r_e - \mathbf{r}_e \cdot \mathbf{k})}{b^2} \right] + O(c^{-5}). \quad (10.100)$$

The first term is obviously the time required to travel a distance  $|\mathbf{r}_{\text{obs}} - \mathbf{r}_e|$  in the absence of gravity, while the second term is the delay produced by the massive body. With the source assumed to be at a large distance behind the body, and the observer at a large distance in front of the body, we have that  $\mathbf{r}_e \cdot \mathbf{k}/r_e \simeq -1$  and  $\mathbf{r}_{\text{obs}} \cdot \mathbf{k}/r_{\text{obs}} \simeq 1$ , and Eq. (10.100) simplifies to

$$t_{\text{obs}} - t_e = \frac{1}{c}|\mathbf{r}_{\text{obs}} - \mathbf{r}_e| + \frac{2GM}{c^3} \ln \frac{4r_{\text{obs}}r_e}{b^2} + O(c^{-5}). \quad (10.101)$$

This will be our final expression for the light-travel time.

The situation examined thus far does not yet give rise to a means to measure the time delay, because the time of emission  $t_e$  is typically not known. A slight variation on the theme, however, gives us what we need. Imagine that the light source at  $\mathbf{r}_e$  is replaced by a

reflector, and that the observer at  $\mathbf{r}_{\text{obs}}$  sends a pulse of light to the reflector at a time  $t_0$ , to receive it back at a later time  $t_1$ . The light-travel time during the return trip is twice what was calculated previously, and we find that

$$t_1 - t_0 = \frac{2}{c} |\mathbf{r}_{\text{obs}} - \mathbf{r}_e| + \frac{4GM}{c^3} \ln \frac{4r_{\text{obs}}r_e}{b^2} + O(c^{-5}). \quad (10.102)$$

The last term is the famous *Shapiro time delay*. For solar-system situations it can be expressed as

$$\Delta t_{\text{Shapiro}} = \frac{M}{M_\odot} \left\{ 240 - 20 \ln \left[ \left( \frac{b}{R_\odot} \right)^2 \left( \frac{\text{AU}^2}{r_{\text{obs}}r_e} \right) \right] \right\} \mu\text{s}, \quad (10.103)$$

in which  $r_{\text{obs}}$  and  $r_e$  are measured in astronomical units, the distance between the Earth and the Sun; the Shapiro time delay is measured in hundreds of microseconds.

While we are now closer to a measurement protocol, we are not there yet. For one thing, our expression for the time delay, Eq. (10.102), is given in terms of coordinate time and distances, and it should be converted to an expression that involves observable quantities only. For another, we do not have access to the unperturbed, Euclidean distance  $|\mathbf{r}_{\text{obs}} - \mathbf{r}_e|$  in order to separate out the relativistic effect. Just as in the case of the deflection of light reviewed in Sec. 10.2.3, it is essential to do a differential measurement of the variations in round-trip travel times during many repetitions of the experiment. Particularly important are “superior conjunction” configurations in which the reflector transits behind the Sun as viewed from Earth, leading to a strong modulation of the travel times by the logarithmic term in Eq. (10.102) as  $b$  changes with time. In order to do this accurately, however, one must take into account the variations in travel times that are due to the orbital motion of the reflector. This is done by accumulating radar-ranging data on the reflector when it is far from superior conjunction (so that the time-delay term is negligible) to determine its orbit, using the computed orbit to predict its coordinate trajectory  $\mathbf{r}_e(t)$  near superior conjunction, and then combining this with the trajectory of the Earth  $\mathbf{r}_\oplus(t)$  to determine the Newtonian round-trip time and the logarithmic term in Eq. (10.102). The prediction made on this basis can then be compared with the actual measurements obtained during superior conjunction.

It was radio astronomer Irwin I. Shapiro who first discovered, in 1964, that the time delay was a prediction of general relativity. (It was independently discovered by Jet Propulsion Laboratory scientists Duane Muhleman and Paul Reichley.) He called it the “fourth” test of general relativity, after the three classical tests – the periastron advance of Mercury, the light deflection, and the redshift of light as it climbs up a gravitational field. Shapiro and his colleagues carried out the first measurement of the time delay in 1967, by bouncing radar signals off the surface of Mercury. Later experiments involved radar echos from Venus and the tracking of the Mars exploration spacecraft, Mariners 6, 7 and 9, as well as the 1976 Viking landers and orbiters.

The most recent measurement of the Shapiro time delay exploited the Cassini spacecraft while it was on its way to Saturn. Several circumstances made this mission particularly favorable. One was the ability to do tracking measurements using both X-band (7175 MHz) and Ka-band (34316 MHz) radar, thereby significantly reducing the dispersive effects of the solar corona. Another was that the distance of closest approach of the radar signals to the

Sun was only  $1.6 R_{\odot}$  during Cassini's 2002 superior conjunction, when the spacecraft was at 8.43 AU from the Sun. Because the tracking involved Doppler measurements only, and not time delays, it was the rate of change of the Shapiro delay that was actually measured. The result was in agreement with general relativity to two parts in  $10^5$ .

The Shapiro delay now figures in a wide range of astrophysical phenomena. In its one-way incarnation, it has been measured in a number of binary-pulsar systems, most notably in the double pulsar J0737-3039A/B, where the orbit is seen almost edge-on, and the pulsed radio signals from each pulsar pass close to the other pulsar once per orbit. It is also relevant in analyses of the spectra and time variations of X-ray emissions from accretion disks around black holes; in this case the modeling must be performed in full general relativity, because the post-Newtonian approximation breaks down near black holes.

## 10.3 Post-Newtonian gravity in timekeeping and navigation

In our development of the foundations of special and general relativity, we have emphasized that only physically measurable quantities are relevant, and that the selection of coordinates is completely arbitrary, devoid of physical meaning. On the other hand, to calculate such things as equations of motion or the Shapiro time delay, it is essential to have a suitable coordinate system. No one should consider doing such calculations using a proper time variable or using physically-measured lengths, because this would introduce unnecessary complications into the calculations. In this section we explore the relationship between the coordinates employed to chart a spacetime and the physical measurements carried out by observers moving in the spacetime.

### 10.3.1 "A brief history of time"

[The title of this section is obviously borrowed from Stephen Hawking's bestseller. We could not resist.]

Most of the calculations presented in this book employ a coordinate system  $x^{\alpha}$  defined so that, far from any gravitating system, the coordinate  $x^0/c := t$  represents proper time as measured by an inertial observer at rest with respect to the coordinate system, and  $x^j$  represent spatial coordinates whose scale is measured by rigid rulers also at rest. But we rarely have access to those ideal observers, and we make our measurements using instruments that reside in the gravitational field of the Earth, which is itself moving in the gravitational environment of the solar system.

Most measurements involve time. We determine the orbit of a planet by measuring the time at which it passes in front of a star or transits across the Sun. We determine the orbit of a spacecraft by measuring either the round-trip travel time of a radar signal (radar ranging) or the change in frequency of the return signal compared to the emitted signal (Doppler tracking). We determine the position of a quasar by measuring the difference in arrival time of a given radio wave front at two radio telescopes separated by a known baseline. We navigate around the countryside by measuring the times of arrival of coded signals from

four or more GPS satellites orbiting the Earth. So a question is: what is the relationship between time as measured on Earth and the *coordinate time*  $t$  that we use to calculate the motion of planets and satellites or the trajectories of light rays?

Before about 1950, this question did not have much practical importance. Time was determined initially by the regular rotation of the Earth, so that the second was defined to be  $1/86\,400$  of a day. But as measurements of the orbits of the Earth and planets improved, it was realized that the length of the day fluctuates seasonally because of large-scale atmospheric motions, and it increases with time because of tidal dissipation, making it ill suited as a standard of time for precision measurements. The standard of time was then replaced by ephemeris time, for which the second is defined to be a fraction of the period of the Earth's orbit around the Sun. But even the Earth's orbital period is not precisely constant, because of planetary perturbations, and a better time standard had to be identified.

The development of atomic clocks during the period 1949–1955 made it possible to envision a standard of time based on fundamental physics instead of astronomy, with a potential for unprecedented accuracy. Today the second is defined to be the time elapsed during 9192 631 770 cycles of the hyperfine transition in the ground state of cesium-133. In addition, the speed of light has now been *defined* to be exactly equal to 299 792 458 m/s, so that the meter is now defined in terms of the second, and is no longer tied to the length of a certain platinum rod in Paris.

During the 1960s and early 1970s, two developments signaled that general relativity would have to be incorporated into the practical definition of time. One was the development of numerical methods to integrate the equations of motion for bodies in the solar system, which replaced the largely analytic perturbative methods reviewed in Chapter 3. These techniques made it straightforward to include post-Newtonian corrections in the equations of motion, so that ephemeris time could be regarded as akin to the coordinate time  $t$  that appears in the post-Newtonian metric. The second development was the ongoing improvement in the accuracy and stability of atomic clocks, in particular those based on cesium-133, rubidium-87, and the 21-cm hyperfine transition of hydrogen. With stabilities better than a few parts in  $10^{13}$ , these new standards were sensitive to the effects of gravity on time. For example, two clocks on the surface of the Earth differ in their rates by one part in  $10^{13}$  for every kilometer in height difference, and this effect is easily measured by the latest clocks. The rate difference between the surface of the Earth and interplanetary space is a few parts in  $10^{10}$ , so that the effect of gravity on time is very important in the determination of planetary orbits.

As a result, in 1976 the International Astronomical Union, one of whose functions is to establish the world-wide agreements and conventions for the establishment and dissemination of precise time, resolved to establish atomic time as the primary international standard, known as *Temps Atomique International* (TAI), for which the second is defined by a cesium-133 clock at rest on the Earth's geoid (see below). They also established the relationship between TAI and the time coordinate  $t$  involved in the post-Newtonian metric and in calculations of orbital motion. In 1991 the IAU endeavored to make these definitions as rigorous as possible from a relativistic point of view, and established working groups to monitor the definitions and their accuracy in the face of improving technology and more

exacting requirements by users. This is a great example of the practical importance of general relativity.

The resolutions of the IAU are framed in the context of the post-Newtonian metric and equations of motion, and it has not yet proved necessary to extend these considerations to second post-Newtonian order. This is because the maximum size of the 2PN corrections is of order  $(U/c^2)^2 \sim 10^{-16}$ , which is just below the accuracy of the best current clocks (here the gravitational potential is dominated by the Sun, and it is evaluated at the Earth's orbit). One can well imagine, however, that the next generation of atomic clocks, based on phenomena like cold atom fountains, Bose-Einstein condensates, or atom interferometers, will be so accurate that 2PN corrections will be needed.

### 10.3.2 Reference frames

We begin our exploration of the measurement of time on, or near, a moving and rotating Earth with the specification of three reference frames. The first is the *barycentric frame*  $(t_{\text{bary}}, \mathbf{x}_{\text{bary}})$ , which is attached to the barycenter of the entire solar system; this is the frame in which the motion of planets would be calculated. The second is the *non-rotating geocentric frame*  $(\bar{t}, \bar{\mathbf{x}})$ , which is attached to the moving Earth, but is non-rotating with respect to the barycentric frame; this is the frame in which the motion of satellites around the Earth would be calculated, and the behavior of their clocks analyzed. And the third is the *rotating geocentric frame*  $(t, \mathbf{x})$ , which is also attached to the moving Earth, but rotating rigidly with Earth's angular velocity; this is the frame in which the behavior of clocks on the surface of the Earth would be analyzed.

The transformation between the barycentric and non-rotating geocentric frames is a special case of the class of post-Newtonian transformations developed in Sec. 8.3. This case was examined in Sec. 8.3.5, where it was shown that

$$t_{\text{bary}} = \bar{t} + \frac{1}{c^2}(A + \mathbf{v}_{\oplus} \cdot \bar{\mathbf{x}}) + O(c^{-4}), \quad (10.104a)$$

$$\mathbf{x}_{\text{bary}} = \bar{\mathbf{x}} + \mathbf{r}_{\oplus} + O(c^{-2}), \quad (10.104b)$$

in which  $\mathbf{r}_{\oplus}(\bar{t})$  is the Earth's position in the barycentric frame, expressed as a function of geocentric time  $\bar{t}$ ,  $\mathbf{v}_{\oplus}(\bar{t})$  its velocity, and  $A(\bar{t})$  is determined by

$$\frac{dA}{d\bar{t}} = \frac{1}{2}v_{\oplus}^2 + U_{\text{ext}}, \quad (10.105)$$

with  $U_{\text{ext}}(\bar{t}, \mathbf{r}_{\oplus})$  denoting the piece of the Newtonian potential produced by the bodies external to the Earth, including the Moon, Sun, and other solar-system bodies, evaluated at the Earth's position.

The discussion of Sec. 8.3.5 implies that in the non-rotating geocentric frame, the metric is given by

$$ds^2 = -\left[1 - \frac{2}{c^2}U + O(c^{-4})\right]d(c\bar{t})^2 + \left[1 + O(c^{-2})\right](d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2), \quad (10.106)$$

in which  $U$  (which should be properly denoted  $\bar{U}$ ) is the Earth's gravitational potential augmented by the tidal potential produced by the external bodies. For reasons that will be



clarified shortly, we have neglected the post-Newtonian terms in the metric, and retained only the Newtonian term in  $\bar{g}_{00}$ .

We wish to use the metric of Eq. (10.106) to calculate the interval of proper time  $d\tau$  measured by a clock moving with a velocity  $\bar{\mathbf{v}}$  in the non-rotating geocentric frame. Because  $d\tau$  is related to  $ds^2$  by  $c^2 d\tau^2 = -ds^2$ , and because  $d\bar{\mathbf{x}} = \bar{\mathbf{v}} d\bar{t}$  as we follow the clock's world line, we find that

$$d\tau = \left[ 1 - \frac{1}{c^2} \left( \frac{1}{2} \bar{v}^2 + U \right) + O(c^{-4}) \right] d\bar{t}. \quad (10.107)$$

This is a key equation that allows us to relate  $d\tau$ , a quantity measured by an actual clock, to  $d\bar{t}$ , an interval of coordinate time that is measured by no one, but could be employed in many calculations. The relation involves the clock's speed  $\bar{v}$  – a special relativistic effect – and the gravitational potential  $U$  – a general relativistic effect. The neglect of post-Newtonian terms in the metric can now be justified: they would lead to corrections of order  $c^{-4}$  in  $d\tau$ , which are negligible in the present context. We shall also neglect the tidal terms in  $U$ , because  $U_{\text{tidal}}/c^2 \sim 10^{-17}$  is too small to be of practical significance, at least for today's technological state of the art.

The transformation between the rotating and non-rotating geocentric frames involves a rigid rotation of the coordinates around the Earth's rotation axis (which we choose to align with the  $z$ -direction), at the Earth's angular velocity  $\omega$ . The transformation is described in great detail in Sec. 2.3.1, where we show that the velocities transform as  $v^{\bar{a}} = \Lambda^{\bar{a}}_j (v^j + \epsilon^j_{kn} \omega^k x^n)$ , in which  $\bar{\mathbf{v}}$  is the clock's velocity in the non-rotating frame,  $\mathbf{v}$  its velocity in the rotating frame,  $\boldsymbol{\omega} = [0, 0, \omega]$  is the angular-velocity vector, and  $\Lambda$  is the transformation matrix. The relation implies

$$\bar{v}^2 = v^2 + 2\mathbf{v} \cdot (\boldsymbol{\omega} \times \mathbf{x}) + \omega^2(x^2 + y^2), \quad (10.108)$$

and substitution within Eq. (10.107) returns

$$d\tau = \left[ 1 - \frac{1}{c^2} \left( \frac{1}{2} v^2 + \mathbf{v} \cdot (\boldsymbol{\omega} \times \mathbf{x}) + \Phi \right) + O(c^{-4}) \right] dt, \quad (10.109)$$

in which

$$\Phi = U + \Phi_C \quad (10.110)$$

is the Newtonian potential augmented by the centrifugal potential  $\Phi_C := \frac{1}{2} \omega^2(x^2 + y^2)$ . It should be noted that the time coordinate employed in the rotating frame is the same as in the non-rotating frame:  $t := \bar{t}$ . We can rely on Eq. (10.109) to relate a clock time interval  $d\tau$  to the coordinate time interval  $dt$ .

### 10.3.3 Geoid

We now focus our attention on a clock situated on Earth's surface, or at a low elevation above the surface; the clock may be at rest or moving with a velocity  $\mathbf{v}$ . We shall perform our calculations in the rotating geocentric frame  $(t, \mathbf{x})$ , and therefore work with Eq. (10.109).

Our first item of business is to specify a reference surface from which all elevations will be measured. Because of the important role that the generalized potential  $\Phi$  plays in Eq. (10.109), it is wise to choose a surface  $\Phi = \text{constant}$  as our reference, which is then known as the *geoid*. From Sec. 2.3.1 we know that surfaces of constant  $\rho$ ,  $p$ , and  $\Phi$  all coincide for a fluid body, and since the body's boundary is situated at  $p = 0$ , the reference surface can be made to coincide with the boundary. The Earth is not quite a fluid body, but most of its surface is covered by oceans which do behave as perfect fluids. For the Earth, therefore, the geoid is chosen to coincide with the mean ocean surface extended through the continents.

To model the gravitational field of the Earth we take into account its rotational deformation, but neglect all other deformations produced by inhomogeneities in its mass density; this is a gross oversimplification of the actual situation. We assume that the deformation is axially symmetric, and express the gravitational potential as in Eq. (1.144),

$$U(r, \theta) = \frac{GM}{r} \left[ 1 - J_2 \left( \frac{R}{r} \right)^2 P_2(\cos \theta) \right], \quad (10.111)$$

in which the multipole expansion was truncated to  $\ell = 2$ ; here  $M$  is the mass of the Earth,  $R$  its equatorial radius, and  $J_2$  its dimensionless quadrupole moment. The centrifugal potential is

$$\Phi_C = \frac{1}{2} \omega^2 r^2 \sin^2 \theta, \quad (10.112)$$

and according to our discussion in Sec. 2.4.5,  $J_2 = k_2 \zeta$ , in which  $k_2$  is the Earth's gravitational Love number, and

$$\zeta := \frac{2\omega^2 R^3}{3GM} \quad (10.113)$$

provides the fractional scale of the rotational deformation. We shall assume that  $\zeta \ll 1$ .

Constructing  $\Phi = U + \Phi_C$ , it is easy to show that the geoid  $\Phi(r, \theta) = \Phi_{\text{geoid}} = \text{constant}$  is situated on the surface  $r = R_{\text{geoid}}(\theta)$ , where

$$R_{\text{geoid}}(\theta) = R(1 - f \cos^2 \theta) \quad (10.114)$$

with

$$f = \frac{3}{2} J_2 + \frac{\omega^2 R^3}{2GM} = \frac{3}{4} (1 + 2k_2) \zeta. \quad (10.115)$$

The expression implies that  $f = 1 - R(0)/R(\frac{\pi}{2})$ , and the condition  $\zeta \ll 1$  implies that  $f$  is small.

Another description of the geoid is given by the Cartesian form  $x^2 + y^2 + (1 + 2f)z^2 = R^2$ , and the unit normal  $\mathbf{n}$  to the geoid is easily shown to have the components  $n_x = (1 - 2f \cos^2 \theta) \sin \theta \cos \phi$ ,  $n_y = (1 - 2f \cos^2 \theta) \sin \theta \sin \phi$ , and  $n_z = (1 + 2f \sin^2 \theta) \cos \theta$ . The *geographical latitude*  $\psi$  of a point on the geoid is the angle between  $\mathbf{n}$  and the equatorial plane, so that  $\sin \psi = n_z$ . The relation implies

$$\cos \theta = (1 - 2f \cos^2 \psi) \sin \psi. \quad (10.116)$$

In terms of the geographical latitude we have that  $R_{\text{geoid}} = R(1 - f \sin^2 \psi)$ .

For future reference we wish to find an approximate expression for the potential  $\Phi$  for a point  $\mathbf{x}$  at a height  $h \ll R$  just above (or below) the geoid; the displacement is in the direction of the normal vector  $\mathbf{n}$ , at a constant geographical latitude  $\psi$ . Expressing  $\Phi$  in terms of  $r$  and  $\psi$ , setting  $r = R_{\text{geoid}} + h$ , and expanding in powers of  $h$ , we find that

$$\Phi = \Phi_{\text{geoid}} - hg(\psi) + O(h^2), \quad (10.117)$$

where  $g(\psi) := -\partial\Phi/\partial r$ , given by

$$\begin{aligned} g(\psi) &= \frac{GM}{R^2} \left[ 1 + \frac{3}{2}J_2 - \frac{\omega^2 R^3}{GM} + \left( \frac{2\omega^2 R^3}{GM} - \frac{3}{2}J_2 \right) \sin^2 \psi \right] \\ &= \frac{GM}{R^2} \left[ 1 - \frac{3}{2}(1 - k_2)\zeta + \frac{3}{2}(2 - k_2)\zeta \sin^2 \psi \right], \end{aligned} \quad (10.118)$$

is the local acceleration of gravity. For the Earth this evaluates to  $g = (9.7803 + 0.0519 \sin^2 \psi) \text{ m/s}^2$ .

### 10.3.4 Temps Atomique International

By definition, TAI is the time measured by an atomic clock at rest on the geoid. By virtue of Eq. (10.109), in which we insert  $\mathbf{v} = \mathbf{0}$ , we have that the relationship between TAI and coordinate time  $t$  is given by

$$d(\text{TAI}) = (1 - \Phi_{\text{geoid}}/c^2) dt, \quad (10.119)$$

and constancy of  $\Phi$  on the geoid guarantees that clocks at rest on the geoid mark time at exactly the same rate.

Actual atomic clocks are typically not situated on the geoid. Instead they tend to be housed in more convenient places at a range of altitudes above sea level, such as Boulder, Colorado (NIST), Washington, DC (USNO), Paris, France (BIPM), and so on. These clocks, therefore, do not keep time at the rate specified by TAI, and a correction must be introduced to account for their different placement in the Earth's potential. To calculate this we return to Eq. (10.109), insert Eq. (10.117), and get

$$d\tau = [1 - \Phi_{\text{geoid}}/c^2 + hg(\psi)/c^2] dt \quad (10.120)$$

for the time  $\tau$  measured by a clock at rest at a geographical latitude  $\psi$  and elevation  $h$  above the geoid. Combining this with Eq. (10.119) produces

$$d(\text{TAI}) = [1 - hg(\psi)/c^2] d\tau, \quad (10.121)$$

an equation that relates the measured time  $\tau$  to the time standard TAI. The correction evaluates to  $g/c^2 = (1.0882 + 0.00577 \sin^2 \psi) \times 10^{-16} \text{ m}^{-1}$ . Thus, any atomic clock at rest on Earth can have its rate linked directly to TAI by a simple correction factor. In practice, TAI is defined by an average measurement over an ensemble of the best atomic clocks at different locations around the world, suitably corrected via Eq. (10.121).

The establishment of TAI as a precise time standard relies on measurements made by atomic clocks at multiple locations. How are these clocks synchronized? We shall not delve

into the practical aspects of this matter, but wish to raise an important question of principle: How can clocks be synchronized in a rotating reference frame?

Suppose that we have a clock at position  $B$  that we wish to synchronize with another clock at position  $A$ . For simplicity we ignore effects associated with elevation, and assume that both locations are on the geoid. Each clock marks time at the rate indicated by Eq. (10.119), and integration yields  $\tau[A] = (1 - \Phi_{\text{geoid}}/c^2)t$  and  $\tau[B] = (1 - \Phi_{\text{geoid}}/c^2)t + \Delta\tau$ , in which  $\Delta\tau$  is an integration constant. The synchronization procedure is meant to achieve  $\Delta\tau = 0$ , so that  $\tau[B] = \tau[A]$ .

A possible way of establishing synchronization would be to rely on a portable clock that is transported slowly from  $A$  to  $B$ . This clock is initially synchronized with the  $A$ -clock, and during its trip to  $B$  it marks time at a rate  $d\tau^*$  given by Eq. (10.109),

$$d\tau^* = \left[ 1 - \frac{1}{c^2} (\mathbf{v} \cdot (\boldsymbol{\omega} \times \mathbf{r}) + \Phi_{\text{geoid}}) \right] dt, \quad (10.122)$$

in which  $\mathbf{r}(t)$  is the path of the portable clock; we have neglected the  $\frac{1}{2}v^2$  term under the assumption that the transport is slow. Integrating this for the whole trip, which begins at time  $t_1$  and ends at time  $t_2$ , we find that

$$\tau_2^* - \tau_1^* = \left[ 1 - \Phi_{\text{geoid}}/c^2 \right] (t_2 - t_1) - \frac{1}{c^2} \int_A^B (\boldsymbol{\omega} \times \mathbf{r}) \cdot d\mathbf{r}. \quad (10.123)$$

Because the clock is synchronized with the  $A$ -clock at the beginning of the trip, we have that  $\tau_1^* = \tau_1[A] = (1 - \Phi_{\text{geoid}}/c^2)t_1$ , in which  $\tau_1[A]$  is the time recorded by the  $A$ -clock at departure. Expressing  $t_2$  in terms of  $\tau_2[B]$ , the time recorded by the  $B$ -clock at the arrival of the portable clock, and solving for  $\tau_2^*$ , we find that

$$\tau_2^* = \tau_2[B] - \Delta\tau - \frac{1}{c^2} \int_A^B (\boldsymbol{\omega} \times \mathbf{r}) \cdot d\mathbf{r}. \quad (10.124)$$

We see that the path integral prevents us from achieving synchronization ( $\Delta\tau = 0$ ) by doing the obvious, setting the  $B$ -clock so that  $\tau_2[B] = \tau_2^*$ . Instead, we must synchronize the  $B$ -clock by setting

$$\tau_2[B] = \tau_2^* + \frac{1}{c^2} \int_A^B (\boldsymbol{\omega} \times \mathbf{r}) \cdot d\mathbf{r}; \quad (10.125)$$

a correction must be applied to reflect the transport of the clock in the rotating frame of the Earth. Note that the path integral is independent of  $\mathbf{v}$ ; the effect survives even in the limit of a quasi-static transport.

This failure of simple-minded synchronization, known as the *Sagnac effect* after the French physicist Georges Sagnac (1869–1926), is a consequence of the Earth's rotation. For a path  $\mathbf{r}$  confined to the geoid, it is easy to show that

$$\frac{1}{c^2} \int_A^B (\boldsymbol{\omega} \times \mathbf{r}) \cdot d\mathbf{r} = \frac{\omega R^2}{c^2} \int_A^B \sin^2 \theta d\phi, \quad (10.126)$$

and the integral (multiplied by  $R^2$ ) is twice the area swept out by a line joining the center of the Earth to a projection of the path onto the equatorial plane. For a closed path around the

equator, the Sagnac term amounts to 207 ns, and the effect is very noticeable. Note that the Sagnac effect disappears when the clock is transported along a meridian, so that  $d\phi = 0$ .

Clock synchronization on a rotating Earth can nevertheless be achieved provided that one applies the Sagnac correction, which requires one to keep careful track of the path taken by the portable clock. As a result, all clocks on the geoid can be consistently synchronized to a single master clock, which could be situated in Washington DC, Greenwich UK, Beijing China, or at the North Pole (the latter being the most politically neutral). This method now forms the basis for all terrestrial timekeeping.

### 10.3.5 Orbiting clocks

Atomic clocks orbiting the Earth have become an important part of our lives thanks to the Global Positioning System (GPS; see Box 10.4), which would fail utterly if it did not account for relativistic effects in timekeeping. To analyze the behavior of orbiting clocks we put ourselves in the non-rotating geocentric frame  $(\bar{t}, \bar{\mathbf{x}})$  and use Eq. (10.107),

$$d\tau[\text{orbit}] = \left[ 1 - \frac{1}{c^2} \left( \frac{1}{2} \bar{v}^2 + U \right) \right] d\bar{t}, \quad (10.127)$$

to relate the clock's proper time  $\tau[\text{orbit}]$  to the coordinate time  $\bar{t}$ ; here  $\bar{v}$  is the satellite's orbital velocity and  $U$  is the Earth's gravitational potential. In this case we may ignore the multipolar deformations, which are small at the high altitudes considered here ( $r \simeq 4R_\oplus$  for GPS satellites), and set  $U = GM/r$ .

Each satellite moves on a Keplerian orbit around the Earth, and we may rely on the orbital equations derived in Sec. 3.2.4 to integrate Eq. (10.127). It is convenient to employ a representation of the motion in terms of the eccentric anomaly  $u$ , as displayed in Eqs. (3.34) and (3.35). We have that  $\bar{r} = a(1 - e \cos u)$  and  $\bar{v}^2 = (GM/a)(1 + e \cos u)/(1 - e \cos u)$ , where  $a$  and  $e$  are the semi-major axis and eccentricity, respectively, and the eccentric anomaly is related to  $\bar{t}$  by Kepler's equation

$$\bar{t} - T = \sqrt{\frac{a^3}{GM}}(u - e \sin u), \quad (10.128)$$

in which  $T$  is the time of perigee passage. It follows that

$$\frac{1}{2} \bar{v}^2 + \frac{GM}{r} = \frac{3}{2} \frac{GM}{a} + \frac{2GM}{a} \frac{e \cos u}{1 - e \cos u}, \quad (10.129)$$

and substitution within Eq. (10.127) yields

$$\tau[\text{orbit}] = \left( 1 - \frac{3}{2} \frac{GM}{c^2 a} \right) \bar{t} - \frac{2GM}{c^2 a} \sqrt{\frac{a^3}{GM}} e \sin u + \tau_0, \quad (10.130)$$

where  $\tau_0$  is a constant of integration set by synchronization of the orbiting clock.

The time marked by the orbiting clock is expressed in terms of the coordinate time  $\bar{t}$ . To complete the description we should relate it to the time measured by a clock at rest (or moving slowly inside a car) at a geographical latitude  $\psi$  and elevation  $h$  on the surface of

the Earth. The translation is provided by Eq. (10.120), which integrates to

$$\tau[\text{ground}] = [1 - \Phi_{\text{geoid}}/c^2 + hg(\psi)/c^2]\bar{t} + \tau_1, \quad (10.131)$$

with  $\tau_1$  another constant of integration. Making the substitution in Eq. (10.130), we find that

$$\begin{aligned} \tau[\text{orbit}] = & \left\{ 1 - \frac{1}{c^2} \left[ \frac{3}{2} \frac{GM}{a} - \Phi_{\text{geoid}} + hg(\psi) \right] \right\} \tau[\text{ground}] \\ & - \frac{2GM}{c^2 a} \sqrt{\frac{a^3}{GM}} e \sin u + \tau_2, \end{aligned} \quad (10.132)$$

in which  $u$  is now expressed implicitly as a function of  $\tau[\text{ground}]$ .

Equation (10.132) relates the time measured by a clock on board a satellite to the time measured by a clock on the ground. The first group of terms describes a constant difference between the clock rates. The first correction term  $-3GM/(2c^2a)$  comes with a negative sign, and represents a time-dilation effect produced by the orbital motion. The second term  $+\Phi_{\text{geoid}}/c^2$  comes with a positive sign, and represents a gravitational blueshift associated with the transfer of the time signal from high to low altitude. The third term  $-hg(\psi)/c^2$  is an elevation-dependent correction, and its sign depends on the sign of  $h$ . The second group of terms is a time-dependent modulation of  $\tau[\text{orbit}]$  associated with the orbital eccentricity; this effect is usually very small, because the GPS satellites are in highly circular orbits ( $e \sim 0.01$ ). A typical orbit has a period of 12 hours, corresponding to  $a = 26,630$  km, or about  $4.2R_{\oplus}$ , and this implies that the geoid term dominates the orbital term in the constant rate factor. This means that the gravitational blueshift dominates time dilation, so that GPS clocks tick faster than clocks on the geoid by a factor of  $4.467 \times 10^{-10}$ , or by about 38.6 microseconds per day. Note that the geoid and satellite terms cancel each other at an orbital radius of about  $1.5R_{\oplus}$ , and that time dilation would dominate for a lower orbit.

**Box 10.4**

**Global Positioning System**

The Global Positioning System (GPS) has become a ubiquitous fixture of our daily lives, being found in many smartphones and helping us navigate our cars in unfamiliar neighborhoods. The GPS consists of 24 satellites orbiting the Earth, with groups of four satellites on six distinct orbital planes oriented at 55 degrees relative to Earth's equatorial plane. The system operates in such a way that at least four satellites are visible above the horizon at any given time. Each satellite harbours a very accurate atomic clock; the accumulated time lapse of a given clock over a full day is measured in nanoseconds. The system is operated by the United States Department of Defense; it was deployed for military purposes before it was made available to civilians.

Each GPS satellite emits a sequence of radio pulses that contain information about the satellite's position and the time of emission. This information is used by a receiving unit on Earth to determine its own position. The receiver acquires the position  $\mathbf{r}_A$  and time of emission  $t_A$  from each of the four visible satellites. Ignoring the relativistic effects described in the main text, the receiver determines its own position  $\mathbf{r}$  and time  $t$  by solving the system of equations

$$c^2(t - t_A)^2 = |\mathbf{r} - \mathbf{r}_A|^2;$$

solving for  $t$  is necessary because the clock inside the receiver is not nearly as accurate as the atomic clocks, and because this clock has not been synchronized with sufficient precision. In this way the receiving unit can determine its position on Earth with an accuracy of a few meters and local time with an accuracy of 30 nanoseconds.

As explained in the main text, relativistic effects such as time dilation and gravitational blueshift must be taken into account for the successful operation of the GPS. For example, according to Eq. (10.132), the size of the constant rate difference between  $\tau$  [orbit] and  $\tau$  [ground] is given by

$$\frac{3}{2} \frac{GM}{c^2 a} - \frac{\Phi_{\text{geoid}}}{c^2} = -4.4647 \times 10^{-10}.$$

This may seem like a small number, but in the course of a full day it builds up to a time difference of 38.6  $\mu\text{s}$ . This is *much, much larger* than the intrinsic accuracy of the atomic clocks, and translates (after multiplication by  $c$ ) to a distance error of 12 km. Think of Einstein the next time your GPS saves you from an unfortunate detour in downtown Los Angeles.

### 10.3.6 Timing of binary pulsars

As a final application of timekeeping in general relativity, we examine binary pulsars, those important astrophysical systems that have been so fruitful in delivering tests of our favorite theory. A binary pulsar is a two-body system, of which at least one is an active pulsar that emits radio pulses at very regular intervals. The pulses are received on Earth by a radio telescope, and a careful timing allows the astronomer to extract detailed information about the orbital motion of the binary pulsar, including its post-Newtonian aspects. The companion star is generally assumed to be a compact object as well, either a neutron star or a white dwarf (both types of system have been observed) or a black hole (no evidence so far).

We wish to relate the proper time of emission  $\tau_e$  of a radio pulse, as measured by a clock attached to the pulsar, to its arrival time  $t_a$ , as measured by a remote observer. Our timing model will include all the relevant relativistic effects, but for simplicity we shall neglect a number of effects that are also important to the radio astronomer. These include the motion of the observer relative to the barycenter of the binary system, the aberration associated with the rotational motion of the pulsar, and the dispersion of the radio pulses in the interstellar medium.

We consider a binary system consisting of a pulsar of mass  $M_1$  and position  $\mathbf{r}_1$ , and a companion of mass  $M_2$  and position  $\mathbf{r}_2$ . The motion of the binary is described in the system's barycentric frame, and the observer is situated at a remote position  $\mathbf{r}_{\text{obs}}$  relative to the barycenter; the observer is imagined to be at rest at that position, and her clock measures coordinate time  $t$ . As usual we let  $\mathbf{r} := \mathbf{r}_1 - \mathbf{r}_2$  denote the orbital separation, and  $r := |\mathbf{r}|$ .

Our first task is to relate the pulsar's proper time  $\tau$  to the coordinate time  $t$ . Working to the leading, Newtonian order, the relation is provided by Eq. (10.107), which we evaluate at the position of the pulsar. Removing the overbars, which are no longer required, we have

that

$$d\tau = \left[ 1 - \frac{1}{c^2} \left( \frac{1}{2} v_1^2 + U \right) + O(c^{-4}) \right] dt, \quad (10.133)$$

in which  $\mathbf{v}_1$  is the pulsar's velocity and  $U$  is the binary's gravitational potential, consisting of the pulsar's own constant potential  $U_1$  and the companion's potential  $GM_2/r$ . To integrate this we once again make use of the Keplerian relations obtained in Sec. 3.2.4. After some algebra we find that

$$\frac{1}{2} v_1^2 + \frac{GM_2}{r} = \frac{GM_2}{a} \left( \frac{2M_1 + 3M_2}{2m} + \frac{M_1 + 2M_2}{m} \frac{e \cos u}{1 - e \cos u} \right), \quad (10.134)$$

in which  $m := M_1 + M_2$  is the total mass,  $a$  is the binary's semi-major axis,  $e$  its eccentricity, and  $u$  is the eccentric anomaly, which is related to  $t$  by Kepler's equation. Making the substitution and integrating, we find that

$$\tau = \left( 1 - \frac{2M_1 + 3M_2}{2m} \frac{GM_2}{c^2 a} - \frac{U_1}{c^2} \right) t - \frac{M_1 + 2M_2}{m} \sqrt{\frac{a^3}{Gm}} \frac{GM_2}{c^2 a} e \sin u + \tau_0, \quad (10.135)$$

in which  $\tau_0$  is a constant of integration. In practical applications  $\tau_0$  is unobservable, and so is the constant factor in front of  $t$ , which can be absorbed into a re-definition of the pulsar's intrinsic frequency, which is unknown; the key information is contained in the time-dependent term, which is proportional to the binary's eccentricity.

Equation (10.135) allows us to relate the proper time of emission  $\tau_e$  to the coordinate time  $t_e$ . We must now obtain the relation between  $t_e$  and the time of arrival  $t_a$ , by calculating the time required by an electromagnetic signal to travel from the pulsar at  $\mathbf{r}_1$  to the observer at  $\mathbf{r}_{\text{obs}}$ . The work was already carried out in Sec. 10.2.5, and Eq. (10.100) informs us that

$$c(t_a - t_e) = |\mathbf{r}_{\text{obs}} - \mathbf{r}_1| + \frac{2GM_2}{c^2} \ln \left[ \frac{|\mathbf{r}_{\text{obs}} - \mathbf{r}_2| + (\mathbf{r}_{\text{obs}} - \mathbf{r}_2) \cdot \mathbf{k}}{r + \mathbf{r} \cdot \mathbf{k}} \right] + O(c^{-4}), \quad (10.136)$$

where

$$\mathbf{k} := \frac{\mathbf{r}_{\text{obs}} - \mathbf{r}_1}{|\mathbf{r}_{\text{obs}} - \mathbf{r}_1|} \quad (10.137)$$

is the direction of propagation. Some work was required to go from Eq. (10.100) to Eq. (10.136). First, the version here accounts for the fact that the delay is produced by the companion at  $\mathbf{r}_2$  instead of a body situated at the spatial origin of the coordinate system. And second, the factor  $b^2$  in Eq. (10.100) was factorized as  $b^2 = (r - \mathbf{r} \cdot \mathbf{k})(r + \mathbf{r} \cdot \mathbf{k})$ , a relation that follows directly from the definition  $\mathbf{b} := \mathbf{r} - (\mathbf{r} \cdot \mathbf{k})\mathbf{k}$ . Our result can be simplified if we recognize that the observer is situated at a large distance from the binary system. With  $r_{\text{obs}}/r_1 \gg 1$  and  $r_{\text{obs}}/r_2 \gg 1$ , we find that Eq. (10.136) reduces to

$$c(t_a - t_e) = r_{\text{obs}} - \mathbf{n}_{\text{obs}} \cdot \mathbf{r}_1 - \frac{2GM_2}{c^2} \ln \left[ \frac{(1 + \mathbf{n}_{\text{obs}} \cdot \mathbf{n})r}{2r_{\text{obs}}} \right] + O(c^{-4}), \quad (10.138)$$

in which  $\mathbf{n} := \mathbf{r}/r$  and  $\mathbf{n}_{\text{obs}} := \mathbf{r}_{\text{obs}}/r_{\text{obs}}$ . It is understood that  $\mathbf{r}$  and  $\mathbf{r}_1$  are evaluated at time  $t = t_e$ .

We now wish to put Eq. (10.138) in a more concrete form. We adopt the point of view of the astronomer, who looks up at the binary system and assigns to it the orbital parameters



$(a, e, \iota, \Omega, \omega)$ , in which  $\iota$  is the orbital inclination,  $\Omega$  the longitude of the ascending node, and  $\omega$  the longitude of pericenter. She erects a fundamental  $(X, Y, Z)$  frame and aligns it in such a way that the  $Z$ -axis points toward the binary. In this description we have that  $\mathbf{n}_{\text{obs}} = -\mathbf{e}_Z$ , and  $\mathbf{n}$  is given by Eq. (3.42); with this it follows that  $\mathbf{n}_{\text{obs}} \cdot \mathbf{n} = -\sin \iota \sin(\omega + f)$ , in which  $f$  is the orbit's true anomaly.

In the term  $\mathbf{n}_{\text{obs}} \cdot \mathbf{r}_1$  on the right-hand side of Eq. (10.138) we insert the post-Newtonian description of the orbital motion given by Eq. (10.21). In the logarithmic term we may insert the usual Keplerian relations. Making the substitutions, we arrive at

$$c(t_a - t_e) = r_{\text{obs}} + \sin \iota \sin(\omega + f_e) a_1 (1 - e_1 \cos u_e) - \frac{2GM_2}{c^2} \ln \left\{ [1 - \sin \iota \sin(\omega + f_e)] (1 - e \cos u_e) \right\} + O(c^{-4}), \quad (10.139)$$

in which  $a_1$  and  $e_1$  are the pulsar's post-Newtonian orbital parameters, as defined in Eq. (10.38),  $f_e$  is the true anomaly at the time of emission, and  $u_e$  is the eccentric anomaly. The relation between  $f_e$  and  $u_e$  is given by Eq. (10.34), and the relation between  $u_e$  and  $t_e$  is expressed by Eq. (10.24).

Our task is completed. With Eq. (10.135) we relate the proper time of emission  $\tau_e$  of a radio pulse to the coordinate time  $t_e$ , and with Eq. (10.139) we relate  $t_e$  to the arrival time  $t_a$ . Combining these expressions, we obtain the desired relation between  $t_a$  and  $\tau_e$ . The relation is provided in a rather implicit form, but the building blocks are all sufficiently simple that it can easily be turned into a practical timing tool.

In Sec. 10.1.4 we pointed out that the measurement of the pericenter advance in a binary-pulsar system yields a very accurate determination of the system's total mass. The timing formulae of Eqs. (10.135) and (10.139) provide additional information. The time-dependent term in Eq. (10.135) involves a different combination of  $M_1$  and  $M_2$ , and combining this measurement with a pericenter-advance measurement returns values for each mass. When the Shapiro delay is significant (which requires a nearly edge-on orbit), the amplitude of the delay reveals the companion mass  $M_2$  straight away, while its detailed shape depends on  $\sin \iota$ . Such independent measurements of the masses in many binary pulsars turn out to give consistent results to high precision, affirming the correctness of the general relativistic description of the physics. In Chapter 12 we uncover yet another way to measure the masses, via the effect of gravitational radiation damping.

## 10.4 Spinning bodies

As relativistic effects go, those associated with a body's spin tend to be suppressed compared to the effects reviewed in Sec. 10.1. For example, in the spin-orbit contributions to the equations of motion, Eqs. (9.189) or Eq. (9.245), a relativistic factor  $v_A^2/c^2$  has been replaced by  $S_A v_A / (M_A c^2 r_{AB})$ , which is smaller by a factor of order  $(V_A/v_A)(R_A/r_{AB})$ , where  $V_A$  is the equatorial rotation velocity of the spinning body and  $R_A$  is its radius. The rotation velocity for a body is limited by the Keplerian velocity  $(GM_A/R_A)^{1/2}$ , at which matter on the equator is no longer bound to the body, and the orbital velocity is itself

comparable to  $(GM_A/r_{AB})^{1/2}$ , so that one would expect a spin-orbit effect to be smaller than an orbital post-Newtonian effect by a factor of order  $(R_A/r_{AB})^{1/2}$ . In the context of the theory of motion developed in Chapter 9, in which each body was assumed to be small compared with the inter-body separation, this is necessarily a small number. Furthermore, most bodies do not rotate at anything close to the Keplerian limit. For example, the Earth's equatorial rotational velocity is approximately 0.5 km/s, which is 16 times smaller than the Keplerian limit. Even the most rapidly spinning pulsar rotates at less than half its Keplerian limit. The sole exception to this rule is a rapidly rotating black hole, for which the effective  $V_A$  can be comparable to the speed of light.

So it would seem that to detect the effects of spin in the solar system would be hopeless, unless we could relax the constraint that  $R_A \ll r_{AB}$ , which means putting one body close to the surface of the rotating body. We can do this and retain the validity of the equations of motion obtained in Chapter 9 by demanding that the first body have a negligible mass. This is the context of Gravity Probe B, in which gyroscopes are placed on a low Earth orbit with  $r_{AB} = R_A + 650$  km, and of the LAGEOS experiment, which tracked satellites on near-Earth orbits with  $r_{AB} \approx 2R_A$ . On the other hand, by virtue of the remarkable accuracy afforded by pulsar timing, relativistic spin-orbit effects have been measured in a number of binary pulsar systems.

### 10.4.1 Frame dragging and Gravity Probe B

Gravity Probe B will very likely go down in the history of science as one of the most ambitious, difficult, expensive, and controversial relativity experiments ever performed.<sup>1</sup> In fact, the list of superlatives associated with the project is formidable: it fabricated the world's most perfect spheres, it achieved the lowest magnetic fields, it built the best drag-free control, and so on. All these achievements and more were essential to the successful measurement of the minute precession of a spinning body as it orbits the Earth.

#### Precession of a gyroscope

We consider a binary system that consists of the Earth, with mass  $M_\oplus$  and spin  $\mathcal{S}_\oplus$ , and a gyroscope of negligible mass and spin  $\mathcal{S}$ . The gyroscope orbits the Earth, and its spin precesses. The relevant equation was listed in Sec. 9.5.6. According to Eq. (9.198), we have that  $d\mathcal{S}/dt = \boldsymbol{\Omega} \times \mathcal{S}$ , where  $\boldsymbol{\Omega} = \boldsymbol{\Omega}[\text{so}] + \boldsymbol{\Omega}[\text{ss}]$  is the precessional angular velocity decomposed in spin-orbit and spin-spin contributions, with

$$\boldsymbol{\Omega}[\text{so}] = \frac{3}{2} \frac{GM_\oplus}{c^2 r^2} \mathbf{n} \times \mathbf{v}, \quad (10.140a)$$

$$\boldsymbol{\Omega}[\text{ss}] = \frac{G}{c^2 r^3} \left[ 3(\mathcal{S}_\oplus \cdot \mathbf{n})\mathbf{n} - \mathcal{S}_\oplus \right], \quad (10.140b)$$

in which  $\mathbf{r} = r\mathbf{n}$  is the gyroscope's orbital position, and  $\mathbf{v}$  its orbital velocity.

<sup>1</sup> Full disclosure: one of us (CMW) served as Chair of an external NASA Science Advisory Committee for Gravity Probe B from 1998 to 2010.

We describe the orbital motion in the fundamental  $(X, Y, Z)$  frame introduced in Sec. 3.2.5. The frame is oriented in such a way that the  $X$ - $Y$  plane coincides with the Earth's equatorial plane, and  $\mathbf{S}_\oplus$  points in the  $Z$ -direction. We place the gyroscope on a circular orbit of radius  $a$ , inclination  $\iota$ , and longitude of ascending node  $\Omega$ . The direction to the gyroscope is given by Eq. (3.42),

$$\begin{aligned} \mathbf{n} = & [\cos \Omega \cos(\omega + f) - \cos \iota \sin \Omega \sin(\omega + f)] \mathbf{e}_X \\ & + [\sin \Omega \cos(\omega + f) + \cos \iota \cos \Omega \sin(\omega + f)] \mathbf{e}_Y \\ & + \sin \iota \sin(\omega + f) \mathbf{e}_Z, \end{aligned} \quad (10.141)$$

in which  $f = 2\pi t/P$  is the orbital phase;  $P = 2\pi a^{3/2}(GM_\oplus)^{-1/2}$  is the orbital period.

It is convenient to introduce a vectorial basis  $(\mathbf{e}_p, \mathbf{e}_q, \mathbf{e}_z)$  adapted to the gyroscope's orbital plane. The normal to the plane is given by Eq. (3.45),

$$\mathbf{e}_z = \sin \iota \sin \Omega \mathbf{e}_X - \sin \iota \cos \Omega \mathbf{e}_Y + \cos \iota \mathbf{e}_Z. \quad (10.142)$$

As a second basis vector we select

$$\mathbf{e}_p = \cos \Omega \mathbf{e}_X + \sin \Omega \mathbf{e}_Y, \quad (10.143)$$

which lies in the plane and points toward the ascending node. And as a third vector, orthogonal to the first two, we get

$$\mathbf{e}_q = -\cos \iota \sin \Omega \mathbf{e}_X + \cos \iota \cos \Omega \mathbf{e}_Y + \sin \iota \mathbf{e}_Z. \quad (10.144)$$

In terms of the orbital basis we have that  $\mathbf{n} = \cos f \mathbf{e}_p + \sin f \mathbf{e}_q$ ,  $\mathbf{e}_Z = \sin \iota \mathbf{e}_q + \cos \iota \mathbf{e}_z$ , and  $\mathbf{n} \times \mathbf{v} = \sqrt{GM_\oplus/a} \mathbf{e}_z$ .

We insert the foregoing relations within the precessional angular velocities, obtaining

$$\boldsymbol{\Omega}[\text{so}] = \frac{3}{2c^2 a} \left( \frac{GM_\oplus}{a} \right)^{3/2} \mathbf{e}_z, \quad (10.145a)$$

$$\boldsymbol{\Omega}[\text{ss}] = \frac{GS_\oplus}{c^2 a^3} \left[ \frac{3}{2} \sin \iota \sin 2f \mathbf{e}_p + \frac{1}{2} \sin \iota (1 - 3 \cos 2f) \mathbf{e}_q - \cos \iota \mathbf{e}_z \right], \quad (10.145b)$$

and substitute these into the precession equation. Because the changes in the gyroscope spin  $\mathbf{S}$  are small, we may insert its initial value  $\mathbf{S}_0$  on the right-hand side of the equation. And because we are interested only in the secular evolution of the spin, we average the precession equation over a complete cycle of the orbital motion. We arrive at

$$\left\langle \frac{d\mathbf{S}}{dt} \right\rangle_{\text{so}} = \frac{3}{2c^2 a} \left( \frac{GM_\oplus}{a} \right)^{3/2} \mathbf{e}_z \times \mathbf{S}_0, \quad (10.146a)$$

$$\left\langle \frac{d\mathbf{S}}{dt} \right\rangle_{\text{ss}} = \frac{GS_\oplus}{c^2 a^3} \left( \frac{1}{2} \sin \iota \mathbf{e}_q - \cos \iota \mathbf{e}_z \right) \times \mathbf{S}_0. \quad (10.146b)$$

Inserting the appropriate numerical values for the Earth, the equations reveal that the precessional angular velocities are given by

$$\langle \Omega[\text{so}] \rangle = \frac{3}{2c^2 a} \left( \frac{GM_\oplus}{a} \right)^{3/2} = 8.43 \left( \frac{R_\oplus}{a} \right)^{5/2} \text{ as/yr}, \quad (10.147a)$$

$$\langle \Omega[\text{ss}] \rangle = \frac{GS_\oplus}{c^2 a^3} = 0.109 \left( \frac{R_\oplus}{a} \right)^3 \text{ as/yr}; \quad (10.147b)$$

we see that the spin-orbit precession amounts to ~~in~~ a few arcseconds per year, while the spin-spin precession is smaller by almost two orders of magnitude.

### Geodetic precession; Schiff precession

The spin-orbit precession is frequently called the “geodetic precession” of a spin vector. Its underlying origin is the fact that spacetime is curved in the vicinity of the Earth, and that parallel transport of a vector around a closed path returns a vector that points in a direction that differs from the initial direction. The change is given by the product of components of the curvature tensor with the vector itself, multiplied by the area enclosed by the orbit. In order of magnitude we have that  $\delta S \sim (\text{Riemann})(\text{area})S_0$ , and with  $\text{Riemann} \sim GM_\oplus/(c^2 a^3)$ ,  $\text{area} \sim a^2$ ,  $P \sim \sqrt{a^3/(GM_\oplus)}$ , we find that

$$\left\langle \frac{dS}{dt} \right\rangle_{\text{SO}} \sim \frac{\delta S}{P} \sim \frac{1}{c^2 a} \left( \frac{GM_\oplus}{a} \right)^{3/2} S_0, \quad (10.148)$$

in agreement with the more detailed calculation. The geodetic precession is independent of the orientation of the orbit, but is maximized when the spin  $S_0$  lies within the orbital plane; it vanishes when the spin is parallel to  $e_z$ .

The spin-spin precession is also called the “Schiff precession” of a spin vector. In 1960, Leonard Schiff calculated the spin-orbit and spin-spin precessions of a gyroscope and suggested the possibility of measuring them. Unbeknownst to Schiff, a researcher at the US Pentagon named George Pugh had performed the same calculations a few months earlier. Pugh was working for the Weapons Systems Evaluation Group, assessing the use of high-performance gyroscopes in missile and aircraft guidance. He wondered how large the relativistic effects would be, and what it would take of a space experiment to measure them. Pugh’s classified work could not be published in the open literature, and so Schiff was initially given credit for the idea; only later was Pugh’s work discovered and recognized. It is worth noting that all this occurred in the very early days of space exploration, only two years after the launch of *Sputnik*.

The phrases “frame dragging” and “Lense-Thirring effects” are sometimes attached to the spin-spin precession of a spin vector. Lense and Thirring were the first, in 1918, to examine the metric of a rotating body in the weak-field limit, and to work out the effects of rotation on the orbits of bodies, a topic to which we return in the next subsection. They did not consider the motion of gyroscopes explicitly, but the basic phenomenon of frame dragging applies to them just as well: an inertial frame with axes defined by an array of perpendicularly oriented gyroscopes will precess about the axis defined by  $\Omega[\text{ss}]$  relative to

an inertial frame at infinity. This phenomenon is directly linked to the dragging of a particle moving on a geodesic near a rotating body, as discussed back in Sec. 9.5.3.

The Maxwell-like formulation of post-Newtonian theory summarized in Box. 8.1 provides additional insight into the phenomenon of spin–spin precession. We recall that a rotating body produces a gravitomagnetic field  $\mathbf{B}_g = \nabla \times \mathbf{A}_g$ , in which  $\mathbf{A}_g = -4U/c^2$  is a rescaled version of the post-Newtonian vector potential. For a spinning Earth at the spatial origin of the coordinate system, we have that

$$\mathbf{A}_g = \frac{2G}{c^2} \frac{\mathbf{x} \times \mathbf{S}_\oplus}{|\mathbf{x}|^3}, \quad (10.149)$$

and this produces a dipolar gravitomagnetic field given by

$$\mathbf{B}_g = -\frac{2G}{c^2 |\mathbf{x}|^3} \left[ \frac{3(\mathbf{S}_\oplus \cdot \mathbf{x})\mathbf{x}}{|\mathbf{x}|^2} - \mathbf{S}_\oplus \right]. \quad (10.150)$$

Comparing this with Eq. (10.140), we see that  $\boldsymbol{\Omega}[\text{ss}] = -\frac{1}{2}\mathbf{B}_g(\mathbf{x} = \mathbf{r})$ . In electrodynamics, the torque exerted by a magnetic field on a magnetic dipole moment  $\mathbf{m}$  is  $\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B}$ . In gravity we would expect that the torque exerted by a gravitomagnetic field on a gravitomagnetic dipole moment  $\mathbf{m}_g := \frac{1}{2}\mathbf{S}$  is given by

$$\boldsymbol{\tau} = \mathbf{m}_g \times \mathbf{B}_g = \left( \frac{1}{2}\mathbf{S} \right) \times \left( -2\boldsymbol{\Omega}[\text{ss}] \right) = \boldsymbol{\Omega}[\text{ss}] \times \mathbf{S}, \quad (10.151)$$

in agreement with the precession equation.

### Gravity Probe B

We now return to the description of the Gravity Probe B experiment. In order to maximize the spin–orbit precession, we align the gyroscope so that it points in the orbital plane. We write

$$\mathbf{S}_0 = S_0(\cos \psi \mathbf{e}_p + \sin \psi \mathbf{e}_q), \quad (10.152)$$

with  $\psi$  denoting the angle between  $\mathbf{S}_0$  and the line of nodes. It is then easy to show that

$$\begin{aligned} \left\langle \frac{d\mathbf{S}}{dt} \right\rangle_{\text{SS}} &\propto \left( \frac{1}{2} \sin \iota \mathbf{e}_q - \cos \iota \mathbf{e}_z \right) \times \mathbf{S}_0 \\ &\propto \cos \iota \sin \psi \mathbf{e}_p - \cos \iota \cos \psi \mathbf{e}_q - \frac{1}{2} \sin \iota \cos \psi \mathbf{e}_z, \end{aligned} \quad (10.153)$$

and that the length of this vector is proportional to  $[1 - \sin^2 \iota (1 - \frac{1}{4} \cos^2 \psi)]^{1/2}$ . It becomes clear that the effect is maximized by setting  $\iota = 0$  and therefore placing the orbit in the Earth’s equatorial plane. This, however, is not the optimal choice for the operations of Gravity Probe B, because in this case  $\langle d\mathbf{S}/dt \rangle_{\text{SO}}$  and  $\langle d\mathbf{S}/dt \rangle_{\text{SS}}$  are both aligned in the direction of  $\mathbf{e}_z \times \mathbf{S}_0$ . In this orientation the spin–spin precession would be very difficult to distinguish from the dominant spin–orbit precession, and the configuration does not permit a clean measurement of each effect separately. A much better strategy is to select an orbit that produces spin–orbit and spin–spin precessions that are perpendicular to each other. It is easy to show that the inner product between  $\langle d\mathbf{S}/dt \rangle_{\text{SO}}$  and  $\langle d\mathbf{S}/dt \rangle_{\text{SS}}$  is proportional to

$\cos \iota$ , and setting  $\iota = \frac{\pi}{2}$  forces this to be zero. The optimal orbit, therefore, is a *polar orbit*. With these specifications the spin-precession equations become

$$\left\langle \frac{d\mathbf{S}}{dt} \right\rangle_{\text{SO}} = \frac{3}{2c^2 a} \left( \frac{GM_{\oplus}}{a} \right)^{3/2} \left( -\sin \psi \mathbf{e}_p + \cos \psi \mathbf{e}_q \right), \quad (10.154a)$$

$$\left\langle \frac{d\mathbf{S}}{dt} \right\rangle_{\text{SS}} = -\frac{GS_{\oplus}}{2c^2 a^3} \cos \psi \mathbf{e}_z. \quad (10.154b)$$

In this geometry the spin-orbit precession is in the orbital plane, and the spin-spin precession is normal to the plane.

The Gravity Probe B spacecraft was launched into an almost perfectly circular polar orbit at an altitude of 642 km, with the orbital plane parallel to the direction of a guide star known as *IM Pegasi*. The spacecraft contained four spheres made of fused quartz, all spinning about the same axis (two were spun in the opposite direction), which was oriented to be in the orbital plane, pointing toward the guide star. An onboard telescope pointed continuously at the guide star, and the direction of each spin was compared with the direction to the star, which was at a declination of  $16^\circ$  relative to the Earth's equatorial plane. With  $\psi$  chosen such that the angle between  $\mathbf{e}_q$  (now parallel to the Earth's rotation axis) and the spins is equal to  $74^\circ$ , we obtain a value of 6630 mas/yr for the spin-orbit precession, and of 38 mas/yr for the spin-spin precession, the first in the orbital plane (in the north-south direction) and the second perpendicular to it (in the east-west direction).

In order to reduce the non-relativistic torques on the rotors to an acceptable level, the rotors were fabricated to be both spherical and homogenous to better than a few parts in 10 million. Each rotor was coated with a thin film of niobium, and the experiment was conducted at cryogenic temperatures inside a dewar containing 2200 litres of superfluid liquid helium. As the niobium film becomes a superconductor, each rotor develops a magnetic moment parallel to its spin axis. Variations in the direction of the magnetic moment relative to the spacecraft were then measured using current loops surrounding each rotor. As the spacecraft orbits the Earth, the aberration of light from the guide star causes an artificial but predictable change in direction between the rotors and the onboard telescope; this was an essential tool for calibrating the conversion between the voltages read by the current loops and the actual angle between the rotors and the guide star.

The spacecraft was launched on April 20, 2004, and the mission ended in September 2005, as scheduled, when the remaining liquid helium boiled off. Although all subsystems of the spacecraft and the apparatus performed extremely well, they were not perfect. Calibration measurements carried out during the mission, both before and after the science phase, revealed unexpectedly large torques on the rotors, believed to be caused by electrostatic interactions between surface imperfections on the niobium films and the spherical housings surrounding each rotor. These effects and other anomalies greatly contaminated the data and complicated its analysis, but finally, in October 2010, the Gravity Probe B team announced that the experiment had successfully measured both the geodetic and frame-dragging precessions. The outcome was in agreement with general relativity, with a precision of 0.3% for the spin-orbit precession, and 20% for the spin-spin precession.

### 10.4.2 Frame dragging and LAGEOS satellites

We next consider the influence of the Earth's spin on the orbital motion of a satellite, which we take to have a negligible mass and spin. The spin-orbit contribution to the relative acceleration between the satellite and the Earth is given by Eq. (9.252), and with our assumptions we find that it becomes

$$\mathbf{a}[\text{so}] = \frac{2G}{c^2 r^3} \left[ 3(\mathbf{n} \times \mathbf{v}) \cdot \mathbf{S}_\oplus \mathbf{n} + 3(\mathbf{n} \cdot \mathbf{v}) \mathbf{n} \times \mathbf{S}_\oplus - 2\mathbf{v} \times \mathbf{S}_\oplus \right], \quad (10.155)$$

in which  $\mathbf{r}$  is the satellite's position relative to Earth,  $\mathbf{v}$  its velocity, and as usual  $\mathbf{n} := \mathbf{r}/r$ . Recalling that  $\mathbf{h} := \mathbf{r} \times \mathbf{v} = h\mathbf{e}_z$  is the conserved angular momentum (per unit reduced mass) of the Keplerian motion, and writing  $\mathbf{S}_\oplus = S_\oplus \mathbf{e}_z$ , the spin-orbit acceleration becomes

$$\mathbf{a}[\text{so}] = -\frac{2GS_\oplus}{c^2 r^3} \left[ 2\mathbf{v} \times \mathbf{e}_z - 3\dot{r} \mathbf{n} \times \mathbf{e}_z - \frac{3h}{r} (\mathbf{e}_z \cdot \mathbf{e}_z) \mathbf{n} \right], \quad (10.156)$$

with  $\dot{r} := \mathbf{n} \cdot \mathbf{v}$  denoting the radial component of the velocity vector.

We take the spin-orbit acceleration as a perturbing force  $\mathbf{f}$  to be inserted in the formalism of osculating Keplerian orbits reviewed in Sec. 3.3.2. A short computation using  $\mathbf{v} = \dot{r} \mathbf{n} + (h/r) \boldsymbol{\lambda}$  reveals that the components of  $\mathbf{f}$  in the orbital basis  $(\mathbf{n}, \boldsymbol{\lambda}, \mathbf{e}_z)$  introduced in Sec. 3.2.2 are

$$\mathcal{R} = \frac{2GS_\oplus h}{c^2 r^3} \frac{1}{r} \mathbf{e}_z \cdot \mathbf{e}_z, \quad (10.157a)$$

$$\mathcal{S} = -\frac{2GS_\oplus}{c^2 r^3} \dot{r} \mathbf{e}_z \cdot \mathbf{e}_z, \quad (10.157b)$$

$$\mathcal{W} = \frac{2GS_\oplus}{c^2 r^3} \left( \dot{r} \boldsymbol{\lambda} \cdot \mathbf{e}_z + \frac{2h}{r} \mathbf{n} \cdot \mathbf{e}_z \right). \quad (10.157c)$$

In this we insert the Keplerian relations  $h = \sqrt{Gmp}$ ,  $r = p/(1 + e \cos f)$ ,  $\dot{r} = \sqrt{Gm/p} e \sin f$ , in which  $m \simeq M_\oplus$  is the total mass,  $p$  the semi-latus rectum,  $e$  the eccentricity, and  $f$  is the true anomaly. We also insert the vectorial relations listed in Sec. 3.2.5, according to which  $\mathbf{e}_z \cdot \mathbf{e}_z = \cos \iota$ ,  $\boldsymbol{\lambda} \cdot \mathbf{e}_z = \sin \iota \cos(\omega + f)$ , and  $\mathbf{n} \cdot \mathbf{e}_z = \sin \iota \sin(\omega + f)$ , in which  $\iota$  is the orbit's inclination relative to Earth's equatorial plane, and  $\omega$  is the longitude of perigee. Finally, we insert the components of the perturbing force in the osculating equations listed in Sec. 3.3.2, refer to Eqs. (3.69), and integrate over a complete orbital cycle to obtain the secular changes in the Keplerian orbital elements.

The final outcome of this sequence of steps is the statement that  $\Delta p = \Delta e = \Delta \iota = 0$ , but that

$$\Delta \omega = -12\pi \frac{GS_\oplus}{\sqrt{GM_\oplus} p^3 c^2} \cos \iota, \quad (10.158a)$$

$$\Delta \Omega = 4\pi \frac{GS_\oplus}{\sqrt{GM_\oplus} p^3 c^2}. \quad (10.158b)$$

These results can be combined to give

$$\Delta \omega + \cos \iota \Delta \Omega = -8\pi \frac{GS_\oplus}{\sqrt{GM_\oplus} p^3 c^2} \cos \iota, \quad (10.159)$$

**Table 10.1** Orbital elements of laser-ranged satellites.

Satellite	Semi-major axis (km)	Orbital period (min)	Eccentricity	Inclination to equator (°)	Mass (kg)
LAGEOS I	12,257	225	0.0045	109.84	407
LAGEOS II	12,168	223	0.0135	52.64	405
LARES	7,821	115	0.0007	69.5	387

which describes the secular advance of the perigee relative to a fixed reference direction – the  $X$ -axis – in the fundamental plane. The nodal advance  $\Delta\Omega$  over a complete orbit can be converted to an averaged rate of advance by dividing by the orbital period, and inserting numerical values appropriate for an Earth-orbiting satellite, we obtain

$$\left(\frac{d\Omega}{dt}\right)_{\text{sec}} = 0.2188 \left(\frac{R_{\oplus}}{a}\right)^3 (1 - e^2)^{-3/2} \text{ as/yr.} \quad (10.160)$$

This very small effect on the satellite’s motion, measured in fractions of arcseconds per year, is not beyond the reach of modern spacecraft tracking using lasers. In particular, a pair of “Laser Geodynamics Satellites” (LAGEOS) are ideal for this purpose. Launched in 1976 and 1992, the satellites are massive spheres, 60 cm in diameter and weighing about 400 kg, placed in nearly circular orbits with semi-major axes approximately equal to  $2R_{\oplus}$  (the precise orbital elements are listed in Table 10.1). The spheres are covered with laser retroreflectors, conical mirrors designed to reflect a laser beam back to the same direction from which it came. Because of their large mass-to-area ratio, the spheres are less affected by atmospheric drag than other satellites at similar altitudes. This, combined with the high precision of laser ranging (which routinely achieves millimeter-level precision), means that their orbits can be determined extremely precisely. The LAGEOS satellites were launched primarily to carry out studies in geodesy and geodynamics, but it was soon recognized that they were potentially capable of measuring the relativistic nodal advance, which under the stated conditions amounts to approximately 32 mas/yr.

A significant challenge in performing the experiment comes from the fact that the deformed figure of the Earth produces a huge contribution to the nodal advance. The effect of oblateness on orbital motion was examined back in Sec. 3.4.3, and an application of these results to the LAGEOS orbits reveals that the Newtonian perturbation produces an advance of approximately 120 degrees per year, more than 10 million times larger than the relativistic advance. To measure the relativistic contribution to 10% accuracy requires a very precise subtraction of the Newtonian effect, which requires knowledge of the Earth’s quadrupole moment ( $J_2 \sim 10^{-3}$ ) to better than a part in  $10^5$ . The higher multipole moments  $J_4$ ,  $J_6$ , and so on, also contribute substantially to the nodal advance. The Earth’s multipole moments were not known sufficiently well at the time of the launch of the first LAGEOS satellite to permit a clean measurement of the relativistic contribution.

After the launch of LAGEOS I in 1976, it was recognized by Italian physicist Ignazio Ciufolini and others, building on an earlier idea by Stanford physicists Richard van Patten and Francis Everitt, that the Newtonian and relativistic contributions to  $\Delta\Omega$  could actually be distinguished from one another. The key observation is that the Newtonian contribution



is proportional to  $\cos \iota$ , while the relativistic contribution is independent of inclination. If a second LAGEOS satellite were to have an inclination supplementary to that of LAGEOS I, so that  $\iota_2 = 180^\circ - \iota_1$ , then the Newtonian advance would be equal and opposite for the two satellites. (This turns out to be true for the contributions of *all* even- $\ell$  multipole moments; see Exercise 3.5.) This means that if one measured the nodal advance of both satellites simultaneously and added them together, the large Newtonian effect would cancel out exactly, leaving the relativistic effect in full view.

Ciufolini and other relativists campaigned vigorously to have LAGEOS II launched with  $\iota_2 = 70.16^\circ$ , but other considerations prevailed in the end. LAGEOS II was launched with  $\iota_2 = 52.64^\circ$ , mainly to optimize coverage by the world's network of laser tracking stations, which was important for geophysics and geodynamics research. The fall-back option was then to combine the data from the two satellites as they were. One could still eliminate the largest Newtonian contribution coming from  $J_2$  with a suitable linear combination of the two measured nodal advances, thereby revealing the relativistic contribution and those coming from higher-order multipole moments, whose uncertainties would contribute to the error made in measuring the relativistic effect. For a time, Ciufolini and collaborators tried to include a third piece of data, the perigee advance of LAGEOS II (which has a small eccentricity), as a way to also eliminate  $J_4$ , but this turned out to be plagued with systematic errors that were large and hard to control.

But then along came CHAMP and GRACE. Europe's CHAMP (Challenging Minisatellite Payload) and NASA's GRACE (Gravity Recovery and Climate Experiment) missions, launched in 2000 and 2002, respectively, use precision tracking of spacecraft to measure variations in the Earth's gravity on scales as small as several hundred kilometers, with unprecedented accuracies. GRACE consists of a pair of satellites flying in close formation (200 kilometers apart) on polar orbits. Each satellite carries an on-board accelerometer to measure non-gravitational perturbations, a satellite to satellite K-band radar to measure variations in the Earth's gravity gradient on short scales, and a GPS tracking unit to measure larger scale variations. With the dramatic improvements on  $J_\ell$  obtained by CHAMP and GRACE, Ciufolini and his colleagues could now treat  $J_4$  and higher multipole moments as known, and use the two LAGEOS nodal advances to determine  $J_2$  and the relativistic contribution. The final outcome was a successful test of the relativistic prediction at the level of 10% accuracy, which was later confirmed by independent analyses of the LAGEOS/GRACE/CHAMP data.

On February 13, 2012, a third laser-ranged satellite, known as LARES (Laser Relativity Satellite) was launched by the Italian Space Agency. Its inclination was  $69.5^\circ$ , very close to the required supplementary angle relative to LAGEOS I, and its eccentricity was very nearly zero (see Table 10.1). Combining data from all three satellites with continually improving Earth data from GRACE, the LARES team led by Ciufolini hopes to achieve a test of frame-dragging at the 1% level.

### 10.4.3 Binary systems of spinning bodies

In our final exploration of the influence of spin in post-Newtonian gravity, we examine the motion of a binary system of spinning bodies.

### Equations of motion and conserved quantities

The equations of motion for the relative orbit and spin vectors of a binary system were derived back in Sec. 9.5.9. The relative acceleration  $\mathbf{a} := \mathbf{a}_1 - \mathbf{a}_2$  is expressed as  $\mathbf{a} = \mathbf{a}[0\text{PN}] + \mathbf{a}[1\text{PN}] + \mathbf{a}[\text{SO}] + \mathbf{a}[\text{SS}] + O(c^{-4})$ , with the two first terms given by Eq. (10.1), the spin-orbit contribution given by

$$\mathbf{a}[\text{SO}] = \frac{G}{c^2 r^3} \left\{ \frac{3}{2} (\mathbf{n} \times \mathbf{v}) \cdot (3\boldsymbol{\sigma} + 4\mathbf{S}) \mathbf{n} + \frac{3}{2} (\mathbf{n} \cdot \mathbf{v}) \mathbf{n} \times (3\boldsymbol{\sigma} + 4\mathbf{S}) - \mathbf{v} \times (3\boldsymbol{\sigma} + 4\mathbf{S}) + \frac{3\lambda}{2} \left[ \mathbf{n}(\mathbf{n} \times \mathbf{v}) \cdot \boldsymbol{\sigma} - (\mathbf{n} \cdot \mathbf{v}) \mathbf{n} \times \boldsymbol{\sigma} \right] \right\}, \quad (10.161)$$

and the spin-spin contribution given by

$$\mathbf{a}[\text{SS}] = -\frac{3G}{\mu c^2 r^4} \left[ (\mathbf{S}_1 \cdot \mathbf{S}_2) \mathbf{n} - 5(\mathbf{S}_1 \cdot \mathbf{n})(\mathbf{S}_2 \cdot \mathbf{n}) \mathbf{n} + (\mathbf{S}_1 \cdot \mathbf{n}) \mathbf{S}_2 + (\mathbf{S}_2 \cdot \mathbf{n}) \mathbf{S}_1 \right]. \quad (10.162)$$

The accelerations are expressed in terms of the combinations  $\boldsymbol{\sigma} := (M_2/M_1)\mathbf{S}_1 + (M_1/M_2)\mathbf{S}_2$  and  $\mathbf{S} := \mathbf{S}_1 + \mathbf{S}_2$  of the spin vectors, and the arbitrary parameter  $\lambda$  was introduced back in Sec. 9.5.8 to characterize the choice of representative world line within each body;  $m := M_1 + M_2$  and  $\mu := M_1 M_2 / (M_1 + M_2)$  are respectively the total and reduced masses of the binary system.

The spin-precession equation for the first body is  $d\mathbf{S}_1/dt = \boldsymbol{\Omega}_1 \times \mathbf{S}_1$  with  $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_1[\text{SO}] + \boldsymbol{\Omega}_1[\text{SS}]$ , where

$$\boldsymbol{\Omega}_1[\text{SO}] = \frac{2G\mu}{c^2 r^2} \left( 1 + \frac{3M_2}{4M_1} \right) \mathbf{n} \times \mathbf{v}, \quad (10.163a)$$

$$\boldsymbol{\Omega}_1[\text{SS}] = \frac{G}{c^2 r^3} [3(\mathbf{S}_2 \cdot \mathbf{n}) \mathbf{n} - \mathbf{S}_2], \quad (10.163b)$$

are the precessional angular velocities. The equations for the second body are obtained by a simple exchange of labels  $1 \leftrightarrow 2$ .

Back in Sec. 10.1.1 we established that the equations of motion  $\mathbf{a} = \mathbf{a}[0\text{PN}] + \mathbf{a}[1\text{PN}] + O(c^{-4})$  admit a conserved energy  $E = \mu \varepsilon$  given by Eq. (10.3), and a conserved angular momentum  $\mathbf{L} = \mu \mathbf{h}$  given by Eq. (10.4). It is possible to extend these results and show that the equations of motion with spin-orbit and spin-spin terms included also admit a conserved energy and a conserved angular momentum. The modified energy (per unit reduced mass) takes the form of  $\varepsilon = \varepsilon[0\text{PN}] + \varepsilon[1\text{PN}] + \varepsilon[\text{SO}] + \varepsilon[\text{SS}] + O(c^{-4})$ , with the first two terms as listed previously, the spin-orbit term given by

$$\varepsilon[\text{SO}] = \lambda \frac{G}{c^2 r^2} (\mathbf{n} \times \mathbf{v}) \cdot \boldsymbol{\sigma}, \quad (10.164)$$

and the spin-spin term given by

$$\varepsilon[\text{SS}] = -\frac{G}{\mu c^2 r^3} \left[ \mathbf{S}_1 \cdot \mathbf{S}_2 - 3(\mathbf{n} \cdot \mathbf{S}_1)(\mathbf{n} \cdot \mathbf{S}_2) \right]. \quad (10.165)$$

Note that the spin-orbit contribution to the conserved energy vanishes when  $\lambda = 0$ , that is, when the representative world line is chosen to track the body's center-of-mass in the barycentric frame (as opposed to the moving frame of each body). The modified total

angular momentum takes the form

$$\mathbf{J} = \mathbf{L} + \mathbf{S} + O(c^{-4}) \quad (10.166)$$

with  $\mathbf{L} = \mathbf{L}[0\text{PN}] + \mathbf{L}[1\text{PN}] + \mathbf{L}[\text{SO}]$ , where the Newtonian and post-Newtonian terms were given by Eq. (10.4), and the spin-orbit term is given by

$$\begin{aligned} \mathbf{L}[\text{SO}] = & \frac{G\mu}{2c^2 r} \mathbf{n} \times [\mathbf{n} \times (3\boldsymbol{\sigma} + 4\mathbf{S})] \\ & - \frac{\lambda}{2c^2} \left[ \frac{G\mu}{r} \mathbf{n} \times (\mathbf{n} \times \boldsymbol{\sigma}) + \frac{\mu}{m} \mathbf{v} \times (\mathbf{v} \times \boldsymbol{\sigma}) \right]; \end{aligned} \quad (10.167)$$

there is no spin-spin contribution to the total angular momentum. The facts that  $d\boldsymbol{\varepsilon}/dt = 0$  and  $d\mathbf{J}/dt = \mathbf{0}$  by virtue of the post-Newtonian dynamics can be verified with a straightforward computation involving the complete set of equations of motion; for this purpose it is convenient to also make use of the evolution equation

$$\frac{d\mathbf{S}}{dt} = \frac{G\mu}{2c^2 r^2} (\mathbf{n} \times \mathbf{v}) \times (3\boldsymbol{\sigma} + 4\mathbf{S}) + \frac{3G}{c^2 r^3} [(\mathbf{S}_1 \cdot \mathbf{n})\mathbf{n} \times \mathbf{S}_2 + (\mathbf{S}_2 \cdot \mathbf{n})\mathbf{n} \times \mathbf{S}_1] \quad (10.168)$$

for the total spin vector.

### Simple precession

To explore the effects produced by the spin-orbit coupling in the equations of motion of a binary system, we examine first the case in which only one of the bodies has spin; this is sometimes called the “simple precession” case. Taking the first body to have spin, and setting  $\mathbf{S}_2 = \mathbf{0}$ , the precession equation for  $\mathbf{S} = \mathbf{S}_1$  can be expressed as

$$\frac{d\mathbf{S}}{dt} = \frac{2G}{c^2 r^3} \left( 1 + \frac{3M_2}{4M_1} \right) \mathbf{L}[0\text{PN}] \times \mathbf{S}, \quad (10.169)$$

where  $\mathbf{L}[0\text{PN}] = \mu \mathbf{r} \times \mathbf{v}$  is the Newtonian expression for the orbital angular momentum. Expressing  $1 + 3M_2/(4M_1)$  as  $\frac{1}{4}(1 + 3m/M_1)$ , and  $\mathbf{L}[0\text{PN}]$  as  $\mathbf{J} - \mathbf{S} + O(c^{-2})$ , we find that the equation becomes

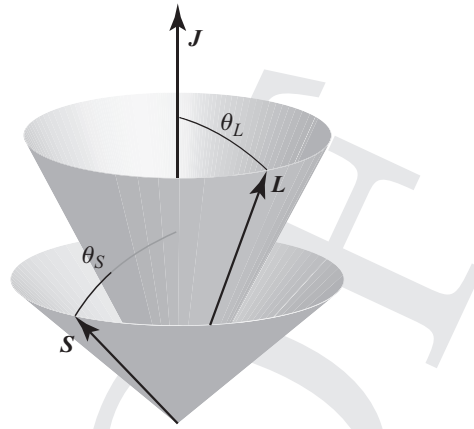
$$\frac{d\mathbf{S}}{dt} = \frac{G}{2c^2 r^3} \left( 1 + \frac{3m}{M_1} \right) \mathbf{J} \times \mathbf{S}, \quad (10.170)$$

in which corrections of order  $c^{-4}$  are neglected. Because  $\mathbf{J} = \mathbf{L} + \mathbf{S}$  is conserved, this equation implies that  $d\mathbf{L}/dt = -d\mathbf{S}/dt$  and is therefore proportional to  $\mathbf{J} \times \mathbf{S}$ . Replacing  $\mathbf{S}$  by  $\mathbf{J} - \mathbf{L}$  in this expression, we arrive at

$$\frac{d\mathbf{L}}{dt} = \frac{G}{2c^2 r^3} \left( 1 + \frac{3m}{M_1} \right) \mathbf{J} \times \mathbf{L} \quad (10.171)$$

for the rate of change of the orbital angular momentum. These equations reveal that both  $\mathbf{S}$  and  $\mathbf{L}$  precess about a fixed  $\mathbf{J}$  at a common angular frequency

$$\Omega_{\text{prec}} = \frac{GJ}{2c^2 r^3} \left( 1 + \frac{3m}{M_1} \right); \quad (10.172)$$



**Fig. 10.5** Simple precession of  $\mathbf{L}$  and  $\mathbf{S}$  about a fixed total angular momentum  $\mathbf{J}$ .

the precession preserves the angle  $\theta_L$  between  $\mathbf{L}$  and  $\mathbf{J}$ , and the angle  $\theta_S$  between  $\mathbf{S}$  and  $\mathbf{J}$  (see Fig. 10.5). The precession of  $\mathbf{L}$  about  $\mathbf{J}$  implies that the binary’s orbital plane no longer has a fixed orientation in space, as it does in the Newtonian and spinless post-Newtonian dynamics. The normal to the orbital plane is aligned with  $\mathbf{L}[0\text{PN}] = \mu \mathbf{r} \times \mathbf{v}$ , and because this deviates from  $\mathbf{L}$  by terms of order  $c^{-2}$ , we see that the orbital plane does not quite follow the precessional motion; there is an extra wobble produced by the spin–orbit contribution to the orbital angular momentum. On average, however, the motion of the orbital plane is described by a steady precession, and this phenomenon has already been encountered in our discussion of the LAGEOS satellites. It is easy to show that the precession angle integrated over a complete orbit corresponds precisely to the nodal advance given by Eq. (10.158).

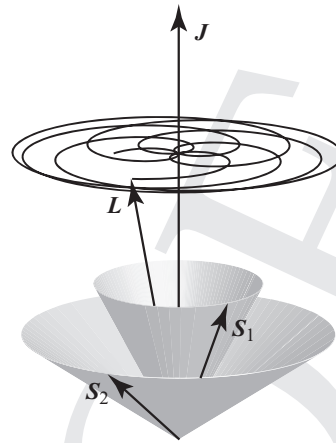
Simple precession can also occur when both bodies are spinning, but in this case the orbital angular momentum is required to be *much larger* than the spin angular momenta, and the masses are required to be equal. The demand for a large  $\mathbf{L}$  is to ensure that the spin–spin coupling continues to be negligible compared to the spin–orbit coupling, and the demand for equal masses is to ensure equality of the precessional angular velocities, which are proportional to  $(1 + 3m/M_1)$  and  $(1 + 3m/M_2)$ , respectively. In such circumstances the precession equations become

$$\frac{d\mathbf{S}_1}{dt} \simeq \frac{7G}{2c^2 r^3} \mathbf{L}[0\text{PN}] \times \mathbf{S}_1, \quad \frac{d\mathbf{S}_2}{dt} \simeq \frac{7G}{2c^2 r^3} \mathbf{L}[0\text{PN}] \times \mathbf{S}_2. \quad (10.173)$$

These equations can be combined to give a precession equation for the total spin  $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$ , and the  $\mathbf{L}[0\text{PN}]$  factor on the right-hand side can again be expressed as  $\mathbf{J} - \mathbf{S} + O(c^{-2})$ . The result can then be used to derive a precession equation for  $\mathbf{L}$ , and we arrive at

$$\frac{d\mathbf{S}}{dt} \simeq \frac{7G}{2c^2 r^3} \mathbf{J} \times \mathbf{S}, \quad \frac{d\mathbf{L}}{dt} \simeq \frac{7G}{2c^2 r^3} \mathbf{J} \times \mathbf{L}. \quad (10.174)$$

These equations imply that  $\mathbf{S}$  and  $\mathbf{L}$  both precess around  $\mathbf{J}$  with a common angular velocity  $\Omega_{\text{prec}} = 7GJ/(2c^2 r^3)$ .



**Fig. 10.6** Precession of  $L$  about a fixed  $J$  for two bodies with  $M_1 = 2M_2$ , with spins of equal magnitudes whose projections perpendicular to  $J$  are initially at right angles.

The precessions remain reasonably simple even when the masses are unequal, provided that we maintain the condition that  $L$  is much larger than both  $S_1$  and  $S_2$ . In this case the factor  $L[0PN] \times S_1$  in  $dS_1/dt$  can be expressed as  $[J - S_1 - S_2 + O(c^{-2})] \times S_1$ , and since  $J$  is dominated by orbital angular momentum, this can be approximated by  $J \times S_1$ . The resulting equations are

$$\frac{dS_1}{dt} \simeq \frac{G}{2c^2 r^3} \left(1 + \frac{3m}{M_1}\right) J \times S_1, \quad (10.175a)$$

$$\frac{dS_2}{dt} \simeq \frac{G}{2c^2 r^3} \left(1 + \frac{3m}{M_2}\right) J \times S_2, \quad (10.175b)$$

and the spins precess with different angular velocities. The motion of  $L = J - S_1 - S_2$  incorporates both precessions, and this leads to a complicated trajectory of the type displayed in Fig. 10.6.

### Spin–spin coupling

Thus far we have considered situations in which the spin–spin aspects of the orbital dynamics could be neglected. To conclude our survey of the dynamics of a binary system of spinning bodies, we examine the conditions under which the spin–spin aspects of the dynamics can become important.

The importance of the spin–spin coupling on body 1 (say) can be measured by the ratio

$$\frac{ss}{so} := \frac{\Omega_1[ss]}{\Omega_1[so]}, \quad (10.176)$$

in which  $\Omega_1[ss]$  is the magnitude of the spin–spin precessional angular velocity vector, and  $\Omega_1[so]$  the magnitude of the spin–orbit angular velocity. In order of magnitude we have

that  $\Omega_1[\text{so}] = GL(1 + 3m/M_1)/(2c^2r^3)$  and  $\Omega_1[\text{ss}] = 2GS_2/(c^2r^3)$ , so that

$$\frac{\text{ss}}{\text{so}} = \frac{4}{1 + 3m/M_1} \frac{S_2}{L}. \quad (10.177)$$

To estimate this we take body 2 to have a radius  $R_2$  and a rotational frequency  $\omega_2$ , so that  $S_2 \sim M_2 R_2^2 \omega_2$ . We also let  $L \sim \mu r^2 \Omega_K$ , in which  $\Omega_K = \sqrt{Gm/r^3}$  is the Keplerian angular velocity of the orbital motion. Making the substitutions, we obtain

$$\frac{\text{ss}}{\text{so}} = \frac{4m}{3m + M_1} \left(\frac{R_2}{r}\right)^2 \frac{\omega_2}{\Omega_K}. \quad (10.178)$$

A useful bound on this is obtained by recalling that  $\omega_2$  is limited by the Keplerian angular velocity at the body's surface, so that  $\omega_2 < \sqrt{GM_2/R_2^3}$ . This gives

$$\frac{\text{ss}}{\text{so}} < \frac{4m}{3m + M_1} \left(\frac{M_2}{m}\right)^{1/2} \left(\frac{R_2}{r}\right)^{1/2}. \quad (10.179)$$

This bound, together with our usual requirement that  $R_2 \ll r$ , implies that  $\text{ss}/\text{so} \ll 1$ , and that the spin–spin aspects of the dynamics are typically small.

An exception to this rule arises when body 1 is very small (in both size and mass), so that it can approach body 2 without suffering from tidal disruption. In such circumstances Eq. (10.178) simplifies to  $\text{ss}/\text{so} \simeq (4/3)(\omega_2/\Omega_K)$ , which can be large whenever  $\omega_2$  is comparable to  $\Omega_K$ . This is precisely the situation for Gravity Probe B: plugging in the angular velocities for the Earth and for a low Earth orbit, we find that  $\text{ss}/\text{so}$  is comparable to the ratio identified in Sec. 10.4.1 between the 43 mas frame-dragging precession and the 6600 mas geodetic precession.

Another exception to the general rule arises in the case of a compact object, a neutron star or a black hole, for which the rotational velocity can approach the speed of light, producing a very large spin. The spin of a compact object scales as

$$S_2 = \chi \frac{GM_2^2}{c}, \quad (10.180)$$

in which  $\chi$  is a dimensionless parameter that ranges between 0 and approximately 0.6 for a neutron star, and between 0 and 1 for a Kerr black hole. Inserting this and  $L \sim \mu\sqrt{Gmr}$  within Eq. (10.177), we find that

$$\frac{\text{ss}}{\text{so}} = \frac{4M_2}{4M_1 + 3M_2} \chi \sqrt{\frac{Gm}{c^2r}} \quad (10.181)$$

for a binary system involving compact objects. For a very tight binary at the natural endpoint of its orbital evolution under gravitational radiation reaction (refer to Chapter 12 below), the spin–spin aspects of the dynamics become important when  $r$  drops below approximately  $10Gm/c^2$ . This is especially true when one of the spins becomes aligned with  $L$  so that the spin–orbit term passes through zero. The system then loses its gyroscopic bearing, and  $L$ ,  $S_1$ , and  $S_2$  can all precess wildly. The orbital angular momentum can even change sign, so that an initially clockwise orbit flips over and becomes counter-clockwise. These strange gyrations, however, occur in a strong-field regime where post-Newtonian theory is

gradually breaking down, and it is only with the advent of numerical relativity that the rich dynamical behavior of spinning binaries could be explored to the fullest extent.

## 10.5 Bibliographical notes

Post-Newtonian celestial mechanics provides a vast field of study that we have sampled only sparsely in this chapter. For a more comprehensive survey the reader is invited to consult the treatises by Soffel (1989) and Brumberg (1991).

The methods introduced in Sec. 10.1.5 to integrate the post-Newtonian equations of motion of a two-body system originate from Damour and Deruelle (1985). The relativistic precession of the line of nodes of the lunar orbit, reviewed in Sec. 10.1.6, was first predicted by de Sitter (1916), and the effect was first measured by Shapiro *et al.* (1988).

The story of the 1919 eclipse expedition to measure the relativistic deflection of light, mentioned in Sec. 10.2.3, is engagingly told in the book *Einstein's Jury* by Jeffrey Crelinsten (2006), and in Daniel Kennefick's *Physics Today* article from 2009. The official report on the expedition is published as Dyson, Eddington, and Davidson (1920). The results of light-deflection measurements carried out by the VLBI between 1979 and 1999 are compiled in Shapiro *et al.* (2004), and the Hipparcos measurements are reported in Froeschlé, Mignard, and Arenou (1997). The “Newtonian” calculation described in Box 10.3, which originates in the independent work of Cavendish and von Soldner, is described in more detail in Will (1988).

The theory of gravitational lenses, touched upon in Sec. 10.2.4, is developed systematically in the excellent monograph by Schneider, Ehlers, and Falco (1992). The discovery of the first lens, in the form of the double quasar Q0957+561, was reported by Walsh, Carswell, and Weymann (1979). Results of the MACHO experiment are recorded in Alcock *et al.* (2000), and the first discovery of an exoplanet via gravitational lensing is described in Bond *et al.* (2004).

The calculation of the Shapiro time delay (Sec. 10.2.5) was first published in Shapiro (1964), and the results of the first measurement were reported in Shapiro *et al.* (1968). The improved experiment involving the Cassini spacecraft is described in Bertotti, Iess, and Tortora (2003).

The relativistic aspects of the Global Positioning System, described in Sec. 10.3.5 and Box 10.4, are treated systematically in Neil Ashby's excellent *Living Reviews* article from 2003. Our discussion of the timing of binary pulsars in Sec. 10.3.6 borrows heavily from Blandford and Teukolsky (1976) and Damour and Deruelle (1986).

The final results of the Gravity Probe B experiment, described in Sec. 10.4.1, were reported in Everitt *et al.* (2011). The dragging of inertial frames was first studied by Thirring and Lense (1918), and the calculation of the spin–spin precession was first presented in the open literature in Schiff (1960). The independent classified calculation by Pugh (1959) was later reprinted in 2003. The successful measurement, of the dragging of inertial frames on the orbital motion of the LAGEOS satellites, as described in Sec. 10.4.2, were reported in Ciufolini and Pavlis (2004).

## 10.6 Exercises

- 10.1** Show that the post-Newtonian energy of Eq. (10.3) and the post-Newtonian angular momentum of Eq. (10.4) are conserved under the two-body dynamics reviewed in Sec. 10.1.1.
- 10.2** In the formulation of the post-Newtonian two-body problem in Sec. 10.1.3, express the post-Newtonian energy  $\varepsilon$  and post-Newtonian angular momentum  $\mathbf{h}$  explicitly in terms of the osculating orbital elements. Using Eq. (10.14), show that  $\varepsilon$  and  $\mathbf{h}$  are constant at 1PN order.
- 10.3** Consider the osculating equations (10.14) in the limit of small eccentricity. Show that a circular orbit (an orbit with  $r$  constant) does *not* correspond to  $e = 0$ . Find a solution for  $p$ ,  $e$ , and  $\omega$  that corresponds to a circular post-Newtonian orbit, and give an interpretation of the orbit in the language of osculating Keplerian orbits.
- 10.4** Calculate the “Newtonian deflection of light” according to the prescription of Box 10.3. Show that the result is half the deflection predicted by general relativity.
- 10.5** In this problem we formulate alternative expressions for the deflection vector of a gravitational lens.
- (a) Show that  $\boldsymbol{\alpha}(\boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}} \psi$ , in which  $\nabla_{\boldsymbol{\xi}}$  is the gradient operator associated with the vector  $\boldsymbol{\xi}$ , and

$$\psi(\boldsymbol{\xi}) = -\frac{4G}{c^2} \int \Sigma(\boldsymbol{\xi}') \ln |\boldsymbol{\xi} - \boldsymbol{\xi}'| d^2\xi'$$

is a deflection potential.

- (b) Show that the deflection angle can also be expressed as

$$\boldsymbol{\alpha}(\boldsymbol{\xi}) = \frac{2}{c^2} \int \nabla_{\boldsymbol{\xi}} U d\ell,$$

in which  $U$  is the Newtonian potential expressed in terms of the coordinates  $(\boldsymbol{\xi}, \ell)$  introduced in Sec. 10.2.4.

- 10.6** Show that the magnification of the images of a Schwarzschild gravitational lens can be written in the form

$$\mu_{\pm} = \frac{1}{1 - (\theta_E/\theta_{\pm})^4},$$

in which  $\theta_E$  is the Einstein angle and  $\theta_{\pm}$  are the two solutions to the lens equation.

- 10.7** Consider a Schwarzschild gravitational lens, and a circularly symmetric source whose center is at an undeflected angle  $\beta_0$  from the lens in the  $x$  direction. Assume that the source has an angular diameter  $2\chi$ , with  $\chi < \beta_0$ , and model any point on the edge of the source as being on a circle described by  $\boldsymbol{\beta}(\phi) = (\beta_0 + \chi \cos \phi)\mathbf{e}_x + \chi \sin \phi \mathbf{e}_y$ , with  $\phi$  ranging from 0 to  $2\pi$ .



- (a) In the limit  $\chi \ll \beta_0$ , show that the image is distorted into an ellipse, with a minor axis parallel to the direction of the image displacement, and with the ratio of minor to major axes given by  $\beta_0/\sqrt{\beta_0^2 + 4\theta_E^2}$ .
- (b) As  $\chi$  increases for fixed  $\beta_0$ , show that the ellipse becomes concave, i.e. becomes an arc, when

$$\frac{\chi}{(\beta_0 - \chi)\sqrt{\beta_0^2 - \chi^2}} \geq \frac{1}{2\theta_E}.$$

**10.8** The Schwarzschild lens is a rather poor model of a gravitational lens produced by a galaxy, and a better approximation of a galactic mass distribution is provided by the Plummer expression

$$m(\xi) = M \frac{(\xi/\xi_c)^2}{1 + (\xi/\xi_c)^2},$$

in which  $M$  is the total mass of the galaxy, and  $\xi_c$  is a “core radius” inside which most of the mass density resides.

- (a) Calculate the mass density  $\Sigma(\xi)$  and plot it as a function of  $\xi/\xi_c$ .
- (b) Show that the lens equation for a Plummer lens is given by

$$\theta - \left(\frac{\theta_E}{\theta_c}\right)^2 \frac{\theta}{1 + (\theta/\theta_c)^2} = \beta,$$

in which  $\theta_E$  is the Einstein angle of Eq. (10.90), and  $\theta_c := \xi_c/D_L$ .

- (c) The lens equation is now a cubic equation for  $\theta$ , and the number of solutions will depend on the sign of the discriminant  $\Delta$ : the lens produces three images when  $\Delta > 0$ , and a single image when  $\Delta < 0$ . Show that the discriminant is given by

$$\Delta = -4\beta^4 - (8 + 20\theta_E^2 - \theta_E^4)\beta^2 + 4(\theta_E^2 - 1)^3,$$

in which  $\beta$  and  $\theta_E$  are now expressed in units of  $\theta_c$ .

- (d) Show that the sign of  $\Delta$  is dictated by the parameter  $\beta_*$  defined by

$$\beta_*^2 = -1 - \frac{5}{2}\theta_E^2 + \frac{1}{8}\theta_E^4 + \frac{1}{8}\theta_E(\theta_E^2 + 8)^{3/2},$$

so that  $\Delta > 0$  when  $\beta < \beta_*$ , while  $\Delta < 0$  when  $\beta > \beta_*$ .

The Plummer lens therefore produces three images when  $\beta$  is smaller than the critical value  $\beta_*$ . It is known that the production of an *odd number of images* is typical in axially symmetric lenses with an extended distribution of mass.

**10.9** A rocket is launched from the surface of a body of mass  $M$  and radius  $R$  on a radial orbit, reaching a maximum distance  $r_{\max}$  before returning to the planet. The rocket carries an atomic clock, while an identical clock remains on the body’s surface. During the flight, the clock on the ground emits a signal with a frequency  $f_0$ , and this signal is received by the rocket with a frequency  $f'$  as measured onboard the rocket. A signal of that same frequency is then generated and returned to the ground (the device that achieves this is called a transponder); there it is measured to have a

frequency  $f_1$ . This is the method used for Doppler tracking. At the same moment as the tracking signal is transponded back, the rocket emits its own signal of frequency  $f_0$  as measured onboard. This signal is received on the ground, and the frequency is measured to be  $f_2$ .

(a) Show that

$$\frac{f_1}{f_0} = \frac{1 - \dot{r}/c}{1 + \dot{r}/c} + O(c^{-3}),$$

$$\frac{f_2}{f_0} = 1 - \frac{\dot{r}}{c} + \frac{GM}{c^2} \left( \frac{1}{R} - \frac{1}{r_{\max}} \right) + O(c^{-3}),$$

where  $\dot{r}$  is the radial velocity of the rocket.

(b) Define the combination of frequency ratios  $\Delta := (f_2/f_0) - (f_1/2f_0) - 1/2$ . Show that this combination eliminates the first-order Doppler effect and is given by

$$\Delta = \frac{GM}{c^2} \left( \frac{1}{R} - \frac{2}{r} + \frac{1}{r_{\max}} \right) + O(c^{-3}). \quad (10.182)$$

Discuss the qualitative behavior of the frequency shift during the experiment.

(c) For the Earth, and for a rocket reaching a maximum altitude of 10 000 km, determine the maximum value of  $\Delta$ . This was the basis for the 1976 Gravity Probe A experiment of Robert Vessot and colleagues, using hydrogen maser clocks and a Scout rocket launched from Wallops Island, Virginia. The experiment verified the prediction for  $\Delta$  to about 70 parts per million.

**10.10** Using the recipe concocted in Sec. 10.3.6, write a computer code that calculates  $t_a$ , the time of arrival of a radio pulse emitted by a binary pulsar, as a function of  $\tau_e$ , the proper time of emission. The code should have the pulsar's intrinsic frequency  $\omega$ , the masses  $M_1$  and  $M_2$ , the post-Newtonian orbital elements  $a$  and  $e$ , and the positional elements  $\iota$  and  $\omega$  as input parameters. You have complete latitude over the design and implementation of the algorithm. Be clever!

**10.11** This exercise approaches the de Sitter precession of the lunar orbit from a different viewpoint.

(a) Treat the Earth–Moon system as a giant gyroscope with spin  $\mathbf{S} = \mu \mathbf{r} \times \mathbf{v}$ , where  $\mu$  is the reduced mass. Imagine that this gyroscope is on a circular orbit around the Sun. Show that the spin–orbit precession of  $\mathbf{S}$  produces the de Sitter precession as given by Eq. (10.44).

(b) Using Eq. (10.42), show that the rate of change of the Moon's angular-momentum vector  $\mathbf{h} = \mathbf{r} \times \mathbf{v}$  can be expressed as

$$\frac{d\mathbf{h}}{dt} = \frac{3}{2} \frac{GM_3}{c^2 R^2} (\mathbf{N} \times \mathbf{V}) \times \mathbf{h}, \quad (10.183)$$

where  $M_3$  is the mass of the Sun,  $R$  the Earth–Sun distance,  $\mathbf{N}$  a unit vector pointing from the Sun to the Earth, and  $\mathbf{V}$  the Earth's velocity. Assume that the orbit of the Earth around the Sun is circular. Compare this result with your simple model in the first part of this exercise. *Hint:* Any post-Newtonian expression

that can be converted to total time derivatives can be absorbed into the definition of  $\mathbf{h}$ ; be careful to stay within the approximations used to construct Eq. (10.42), which amount to keeping only terms linear in  $\mathbf{v}$ , the Moon's velocity relative to Earth.

- 10.12** The gravitomagnetic viewpoint elaborated in Box 8.1 provides useful insights into a number of post-Newtonian effects, most notably the behavior of a gyroscope around a rotating body. Review this material and then provide qualitative answers to the following questions, using only “right-hand-rule” techniques and lessons learned from electromagnetism. To keep everything simple, model both the Earth and a gyroscope as loops of matter rotating about an axis perpendicular to the plane of the loop.
- (a) Show that the gravitomagnetic field of the Earth has the same dipole form as the magnetic field of a loop of positive charge rotating in the same direction, but with the field lines emanating from the south pole and returning to the north pole.
  - (b) A gyroscope (loop) sits in a gravitomagnetic field that is perpendicular to the plane of the loop. Show that the torque on the gyroscope vanishes.
  - (c) A gyroscope sits in a gravitomagnetic field that lies in the plane of the loop. Show that the torque  $d\mathbf{S}/dt$  on the gyroscope is in the direction of  $\mathbf{S} \times \mathbf{B}_g$ , where  $\mathbf{S}$  is the spin vector of the gyroscope.
  - (d) Show that a gyroscope at the Earth's equator, with its spin axis parallel to the equator, precesses in a direction opposite to the Earth's rotation. Show that a similar gyroscope at the pole precesses in the same direction as the Earth's rotation. How do these qualitative conclusions compare with the corresponding precessions derived from Eq. (10.140b)?
- 10.13** In the formulation of the dynamics of spinning binaries provided in Sec. 10.4.3, show that the energy per unit reduced mass  $\varepsilon$  and the total angular momentum  $\mathbf{J}$  are conserved quantities.