Deflection of light to second order: A tool for illustrating principles of general relativity

Jeremiah Bodenner
Department of Physics, University of Wisconsin, River Falls, Wisconsin 54022

Clifford M. Will[a)]
McDonnell Center for the Space Sciences, Department of Physics, Washington University, St. Louis, Missouri 63130

(Received 22 January 2003; accepted 7 March 2003)

We calculate the deflection of light by a spherically symmetric body in general relativity, to second order in the quantity GM/dc^2, where M is the mass of the body and d is a measure of the distance of closest approach of the ray. Using three different coordinate systems for the Schwarzschild metric we show that the answers for the deflection, while the same at order GM/dc^2, differ at order $(GM/dc^2)^2$. We demonstrate that all three expressions are really the same by expressing them in terms of measurable, coordinate-independent quantities. These results provide concrete illustrations of the meaning of coordinates and coordinate invariance, which may be useful in teaching general relativity. © *2003 American Association of Physics Teachers.*
[DOI: 10.1119/1.1570416]

I. INTRODUCTION AND SUMMARY

The deflection of light is one of the empirical cornerstones of general relativity, from the 1919 measurements of starlight deflection during a solar eclipse, to the latest radio telescope observations of radio galaxies and quasars. Through the use of Very Long Baseline Radio Interferometry (VLBI) to observe thousands of radio galaxies and quasars over the entire celestial sphere, the bending by the Sun has now been found to agree with general relativity to a few parts in 10^4 (for a review of the latest results, see Ref. 1). The related phenomenon of gravitational lensing has become a standard astronomical tool, especially important for the study of the distribution of dark matter in the universe and for imaging the most distant galaxies.

The standard general relativistic formula for the deflection is

$$\Delta\phi = \frac{4GM}{dc^2} = 1.7504\left(\frac{M}{M_\odot}\right)\left(\frac{R_\odot}{d}\right) \quad \text{arcsec}, \tag{1}$$

where M is the mass of the body and d is the radius of closest approach of the ray, and $M_\odot$ and $R_\odot$ are the mass and radius of the Sun. This formula is actually the leading term in an expansion in powers of GM/dc^2, known as a post-Newtonian (PN) expansion. For all practical purposes to date, the first-order (1PN) expression above has been sufficient. However, plans are being developed to launch orbiting observatories that use optical interferometry to achieve angular precisions at the level of microarcseconds (μarcsec). Because, at the limb of the Sun $GM/dc^2 \approx 2\times10^{-6}$, one might expect the second-order or second post-Newtonian (2PN) correction to be of order $(1.75)\times(2\times10^{-6})$, or a few μarcsec.

Unfortunately, to measure this tiny effect using an optical device pointing almost directly toward the Sun presents significant technological challenges, not the least of which is thermal control of an on-board optical bench that must maintain picometer level metrology. For this reason (among others), the second-order deflection of light by the Sun has not been a primary goal of proposed missions such as the NASA Space Interferometry Mission (SIM)[2] or the European Space Agency's GAIA mission;[3] instead such scientific questions as the search for extra-solar planets, and improving the cosmic distance scale by extending the reach of parallax measurements, have been the main motivations for such missions. Still, it is not unreasonable to hope that these or possible follow-up missions might actually have the capability to detect the second-order term. Therefore, it is useful to have a clear understanding of what the prediction is at second order. These considerations have already motivated a number of calculations of the second-order term.[4–6]

A second motivation for studying the second-order deflection, and the main theme of this paper, is that it is a useful way for teachers to illustrate some principles of general relativity, in a relatively simple context. Chief among these principles is *general covariance*. Coordinates in general relativity are completely arbitrary; one has total freedom (subject to some simple mathematical constraints of continuity) in one's labeling of events in space–time. Only variables that relate to physically measurable quantities are meaningful. While all textbooks expound upon this principle at length, it is the rare text that provides a concrete example where an effect is calculated in two different coordinate systems and shown explicitly to be the same measurable effect. Instead, a single, convenient coordinate system is selected, and calculations of the given effect are done in that system.

The deflection of light provides such an example. The expression of Eq. (1) is frequently calculated using the Schwarzschild metric, which describes the space–time exterior to any static, spherically symmetric body. However this space–time can be expressed in an infinity of different coordinate systems, including Schwarzschild (or curvature) coordinates, isotropic coordinates, harmonic coordinates, and others. Some derivations use, instead of the exact Schwarzschild solution, an approximate solution from the linearized version of general relativity or from the post-Newtonian expansion of the field equations. All derivations give Eq. (1). The astute student might therefore ask a number of questions:

(1) If the coordinate systems mentioned are all different, why is the final expression for the deflection the same?
(2) If they do differ, where does the difference first show up?
(3) If the expressions for the deflection are ultimately different, what must be done to them so that they can be seen to represent the same physically measured result?

In this paper, we answer these questions by calculating the deflection of light to 2PN order using the Schwarzschild metric, written in three different coordinate systems. The metrics in these three coordinate systems have the following forms (henceforth we use units in which $G=c=1$).

Schwarzschild coordinates:

$$ds^2=-(1-2M/r_S)dt^2+\frac{dr_S^2}{1-2M/r_S}+r_S^2d\Omega^2. \tag{2}$$

Isotropic coordinates:

$$ds^2=-\frac{(1-M/2r_I)^2}{(1+M/2r_I)^2}dt^2+(1+M/2r_I)^4(dr_I^2+r_I^2d\Omega^2). \tag{3}$$

Harmonic coordinates:

$$ds^2=-\frac{1-M/r_H}{1+M/r_H}dt^2+\frac{1+M/r_H}{1-M/r_H}dr_H^2 +r_H^2(1+M/r_H)^2d\Omega^2, \tag{4}$$

where $d\Omega^2\equiv d\theta^2+\sin^2\theta d\phi^2$, which is the standard metric on the two-sphere. The transformations among these different coordinate systems involve only the radial coordinate, and are given by

$$r_S=r_I(1+M/2r_I)^2,\qquad r_S=r_H+M. \tag{5}$$

The results for the deflection in these three coordinate sytems are

$$\begin{aligned}\Delta\phi&=\frac{4M}{d_S}+\left(\frac{M}{d_S}\right)^2\left[\frac{15\pi}{4}-4\right]\quad\text{Schwarzschild},\\ &=\frac{4M}{d_I}+\left(\frac{M}{d_I}\right)^2\left[\frac{15\pi}{4}-8\right]\quad\text{Isotropic},\\ &=\frac{4M}{d_H}+\left(\frac{M}{d_H}\right)^2\left[\frac{15\pi}{4}-8\right]\quad\text{Harmonic},\end{aligned} \tag{6}$$

where d represents in each case the coordinate radius at the point of closest approach of the ray. These results agree with other work.[4–7]

With these results, we may answer the astute student's questions.

(1) At first order, the deflection results from the first-order corrections to flat space–time in the metric. But, from Eq. (5), the three coordinate systems are themselves the same, to first order in M/r, i.e., $r_S=r_I[1+O(M/r_I)]=r_H[1+O(M/r_H)]$, consequently the deflection is the same in all three coordinates, to first order.

(2) The difference first shows up at 2PN order, as can be seen in Eq. (6). Interestingly, the 2PN results for isotropic and harmonic coordinates are identical. This is because these coordinates are actually the same to *second* order, i.e., from Eq. (5), $r_I=r_H[1+O(M/r_H)^2]$. Consequently the difference between isotropic and harmonic coordinates will not show up in the deflection until 3PN order.

(3) To express the three results in terms of something physically measurable (at least in principle), one option is to choose the so-called "circumferential radius," r_C, which is $1/2\pi$ times the physically measured circumference of a circle of a given coordinate radius (holding t, r, and θ fixed). From the three forms of the metric, Eqs. (2), (3) and (4), it is simple to show that $r_C=r_S=r_I(1+M/2r_I)^2=r_H+M$. (Schwarzschild coordinates are in fact *defined* such that r_S is the circumferential radius.) By using these expressions to transform each formula for the deflection in (6) to d_C, the *circumferential* distance of closest approach, we obtain the single expression

$$\Delta\phi=\frac{4M}{d_C}+\left(\frac{M}{d_C}\right)^2\left[\frac{15\pi}{4}-4\right]. \tag{7}$$

This is the second-order deflection of light, in coordinate independent language.

The remainder of this paper provides the details and further discussion. In Sec. II, we write down a general form for the static spherical metric and derive the equation of motion for photons in a form that permits a straightforward solution via second-order perturbation theory. In Sec. III, we specialize to the three coordinate systems and derive the deflection in each. Section IV provides concluding remarks.

II. EQUATION OF MOTION FOR PHOTONS IN A GENERAL SPHERICAL STATIC METRIC

We begin by writing the metric for a static, spherical system in the form

$$ds^2=-A(r)dt^2+B(r)dr^2+r^2C(r)d\Omega^2. \tag{8}$$

For photons, the equations of motion can be obtained by varying the Lagrangian $L=(ds/d\lambda)^2$ with respect to a parameter λ, or from the geodesic equation $d^2x^\alpha/d\lambda^2+\Gamma^\alpha_{\beta\gamma}(dx^\beta/d\lambda)(dx^\gamma/d\lambda)=0$, subject to the constraint that $ds=0$ along the world line of a photon. Because of the spherical symmetry, we can choose the equatorial plane ($\theta=\pi/2$) to be the plane of the motion. The resulting equations of motion are

$$A\frac{dt}{d\lambda}=\text{const}\equiv E, \tag{9}$$

$$Cr^2\frac{d\phi}{d\lambda}=\text{const}\equiv L, \tag{10}$$

$$\frac{d}{d\lambda}\left(2B\frac{dr}{d\lambda}\right)+A'\left(\frac{dt}{d\lambda}\right)^2-B'\left(\frac{dr}{d\lambda}\right)^2-(Cr^2)'\left(\frac{d\phi}{d\lambda}\right)^2=0, \tag{11}$$

where E and L are proportional to the conserved energy and angular momentum of the photon at infinity, and where prime denotes a derivative with respect to r. Substituting Eqs. (9) and (10) into (11), defining $u\equiv 1/r$, and converting from λ to ϕ as the independent variable using Eq. (10), we obtain the second-order differential equation for u,

$$\frac{d^2u}{d\phi^2}+\left(\frac{C}{B}\right)u=-\frac{1}{2}u^2\frac{d}{du}\left(\frac{C}{B}\right)+\frac{1}{2b^2}\frac{d}{du}\left(\frac{C^2}{AB}\right), \tag{12}$$

where $b \equiv L/E$ is the "impact parameter." The condition that the world line be null ($ds^2=0$) can be cast in the form

$$\left(\frac{du}{d\phi}\right)^2 = \frac{C}{B}\left(\frac{C}{Ab^2} - u^2\right). \tag{13}$$

Equation (13) can be shown to be equivalent to (12) by differentiating it with respect to ϕ. The minimum of r, or the maximum of u, denoted u_m, is the turning point of the motion. This occurs where $du/d\phi = 0$, i.e., when

$$b^2 = C(u_m)/A(u_m)u_m^2. \tag{14}$$

III. SOLUTIONS FOR THE SECOND-ORDER DEFLECTION

A. Schwarzschild coordinates

In Schwarzschild coordinates, $A = B^{-1} = 1-2Mu$, $C=1$, so Eq. (12) becomes

$$\frac{d^2u}{d\phi^2} + u = 3Mu^2. \tag{15}$$

The homogeneous solution is $u = u_0 \cos\phi$, which is a straight line with turning point at $\phi = 0$, and $u \to 0$ or $r \to \infty$ at $\phi = \pm\pi/2$. We need to find a solution to second order in the dimensionless parameter $Mu_0 = M/r_0$, which we assume to be small; to do so, we write

$$u = u_0[\cos\phi + Mu_0\delta u_1 + (Mu_0)^2\delta u_2 + \cdots], \tag{16}$$

substitute into (15), and collect terms of equal powers of Mu_0, to obtain the sequence of equations

$$\begin{aligned} &\frac{d^2\delta u_1}{d\phi^2} + \delta u_1 = 3\cos^2\phi, \\ &\frac{d^2\delta u_2}{d\phi^2} + \delta u_2 = 6\,\delta u_1 \cos\phi. \end{aligned} \tag{17}$$

Recalling the standard inhomogeneous solutions to the differential equation $d^2y/dx^2 + y = \cos(nx)$, namely $y = -\cos(nx)/(n^2-1)$, for $n \neq 1$, and $y = (\phi/2)\sin\phi$, for $n=1$, we obtain the solution

$$\begin{aligned} u/u_0 = &\cos\phi + (Mu_0)(3-\cos 2\phi)/2 \\ &+ 3(Mu_0)^2(20\phi\sin\phi + \cos 3\phi)/16. \end{aligned} \tag{18}$$

The maximum of u can be shown by differentiation to occur at $\phi = 0$, whereupon $u_m = u_0[1 + Mu_0 + 3(Mu_0)^2/16]$. Assuming that $u \to 0$ at $\phi = \pi/2 + \delta$ we can solve for δ to obtain $\delta = 2Mu_0 + 15\pi(Mu_0)^2/8 + O(Mu_0)^3$. Because the solution is symmetric about $\phi = 0$, the total deflection angle is twice this. Converting from u_0 to u_m, we obtain

$$\Delta\phi = 4Mu_m + (Mu_m)^2\left(\frac{15\pi}{4} - 4\right). \tag{19}$$

Substituting $u_m = 1/d_S$, we obtain the first equation of Eq. (6).

B. Isotropic coordinates

Isotropic coordinates have the property that the spatial part of the metric is proportional to the flat space metric, $dr^2 + r^2 d\Omega^2 = dx^2 + dy^2 + dz^2$. In these coordinates, $A = [(1 - Mu/2)/(1 + Mu/2)]^2$, $B = C = (1 + Mu/2)^4$, so Eq. (12) becomes

$$\begin{aligned} \frac{d^2u}{d\phi^2} + u &= \frac{2M}{b^2}\frac{(1+Mu/2)^5}{(1-Mu/2)^3}(1 - Mu/4) \\ &= \frac{2M}{b^2}\left(1 + \frac{15}{4}Mu + O(Mu)^2\right). \end{aligned} \tag{20}$$

Substituting Eq. (16), we obtain the sequence

$$\begin{aligned} &\frac{d^2\delta u_1}{d\phi^2} + \delta u_1 = 2(u_0 b)^{-2}, \\ &\frac{d^2\delta u_2}{d\phi^2} + \delta u_2 = \frac{15}{2}(u_0 b)^{-2}\cos\phi, \end{aligned} \tag{21}$$

with the solution

$$u/u_0 = \cos\phi + 2Mu_0/(u_0 b)^2 + 15(Mu_0)^2\phi\sin\phi/4(u_0 b)^2. \tag{22}$$

The maximum u_m is again at $\phi = 0$, with $u_m = u_0[1 + 2Mu_0/(u_0 b)^2]$. Also, from Eq. (14), we have $b^2 = u_m^{-2}(1 + Mu_m/2)^6/(1 - Mu_m/2)^2 \approx u_m^{-2}(1 + 4Mu_m)$. In this case, $u \to 0$ at $\phi = \pi/2 + \delta$ where $\delta = [2Mu_0 + 15\pi(Mu_0)^2/8 + O(Mu_0)^3]/(u_0 b)^2$. Eliminating u_0 and b in favor of u_m, and doubling the angle, we obtain

$$\Delta\phi = 4Mu_m + (Mu_m)^2\left(\frac{15\pi}{4} - 8\right). \tag{23}$$

Substituting $u_m = 1/d_I$, we obtain the second equation of Eq. (6).

C. Harmonic coordinates

In harmonic coordinates, $A = B^{-1} = (1-Mu)/(1+Mu)$, $C = (1+Mu)^2$. These coordinates have the property that, when first transformed to Cartesian coordinates via the normal transformations $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, and $z = r\cos\theta$, the resulting metric satisfies the differential equations

$$\frac{\partial}{\partial x^\nu}(\sqrt{-g}\,g^{\mu\nu}) = 0, \tag{24}$$

where g is the determinant of the metric, $g^{\mu\nu}$ is the inverse of the metric, and a summation over the four values (t,x,y,z) of the index ν is assumed. These coordinates are used mostly in analyzing the weak field limit of general relativity and the generation of gravitational radiation.

Equation (12) then becomes

$$\begin{aligned} \frac{d^2u}{d\phi^2} + u &= \frac{2M}{b^2}(1+Mu)^3 + 2M^2u^3 \\ &= \frac{2M}{b^2}(1 + 3Mu) + 2M^2u^3 + O(M^3u^4). \end{aligned} \tag{25}$$

Substituting Eq. (16), we obtain the sequence

$$\begin{aligned} &\frac{d^2\delta u_1}{d\phi^2} + \delta u_1 = 2(u_0 b)^{-2}, \\ &\frac{d^2\delta u_2}{d\phi^2} + \delta u_2 = 6(u_0 b)^{-2}\cos\phi + \frac{1}{2}(\cos 3\phi + 3\cos\phi), \end{aligned} \tag{26}$$

with the solution

$$u/u_0=\cos\phi+2(Mu_0)/(u_0b)^2+(Mu_0)^2 \times\{12(1+4/(u_0b)^2)\phi\sin\phi-\cos 3\phi\}/16. \quad (27)$$

The maximum u_m is again at $\phi=0$, with $u_m\approx u_0[1+2Mu_0/(u_0b)^2]$. Also, from Eq. (14), we have $b^2=u_m^{-2}(1+Mu_m)^3/(1-Mu_m)\approx u_m^{-2}(1+4Mu_m)$. The angle δ at which $u=0$ is given by $\delta=[2Mu_0+3\pi(Mu_0)^2(4+(u_0b)^{-2})/8+O(Mu_0)^3]/(u_0b)^2$. Eliminating u_0 and b in favor of u_m, and doubling the angle, we obtain

$$\Delta\phi=4Mu_m+(Mu_m)^2\left(\frac{15\pi}{4}-8\right). \quad (28)$$

Substituting $u_m=1/d_H$, we obtain the third equation of Eq. (6).

IV. DISCUSSION

In Sec. I, we showed explicitly how the seemingly different second-order expressions could be seen to be equivalent when expressed in terms of the proper circumferential radius of the circle of closest approach. This radius is obtained by calculating the proper distance around a circle of constant coordinate radius r and fixed time $(dt=0)$, say with $\theta=\pi/2$, and dividing by 2π. From the general form of the metric, Eq. (8), we see that, around this circle, $ds=rC(r)^{1/2}d\phi$, so that $r_C=rC(r)^{1/2}$. For the three coordinate systems, we obtain $r_C=r_S=r_I(1+M/2r_I)^2=r_H(1+M/r_H)$. In terms of d_C then, all three formulas for the deflection collapse to Eq. (7).

An alternative coordinate-independent variable is the impact parameter $b=L/E$; it is coordinate independent because L and E are the physically measurable angular momentum and energy of the photon by an observer at rest far from the solar system. From Eq. (14) and the expressions in Sec. III, we see that, to first order in M/d, $b=d_S(1+M/d_S)=d_I(1+2M/d_I)=d_H(1+2M/d_H)$. Converting each expression in Eq. (6) to b, we find the common, coordinate independent result

$$\Delta\phi=\frac{4M}{b}+\frac{15\pi}{4}\left(\frac{M}{b}\right)^2. \quad (29)$$

An alternative way to demonstrate explicitly that the deflection angle must ultimately be independent of the choice of radial coordinate is via Eq. (13); inverting and integrating with respect to u, we obtain an exact expression for the deflection

$$\Delta\phi=2\int_0^{u_m}\frac{\sqrt{B/C}}{\sqrt{C/Ab^2-u^2}}du-\pi, \quad (30)$$

where u_m is the value where $u^2=C/Ab^2$. Any change of integration variable $u=f(v)$ cannot alter the value of $\Delta\phi$, only its explicit expression in terms of u_m. Note that, because the integral depends only on the impact parameter b, and on M (which appears in the functions A, B, and C), there must be a unique answer for $\Delta\phi$ in terms of b, independent of coordinate system. Equation (29) is that answer to second order.

Although the circumferential radius d_C and the impact parameter b are formally observable quantities, neither is very practical for real-world measurements of the deflection of light, for obvious reasons. Instead, one combines the equation of motion for the light signal with an equation of motion for the observer receiving the signal (such as a telescope on Earth), and calculates the angle of the received signal (usually relative to a similarly calculated angle from a reference source nearby in the sky) as a function of proper (atomic) time measured at the receiver. In observations that use radio interferometry, this angle can be directly related to a phase difference in the radio signal between the two telescopes, so that the measurable, coordinate-independent quantity is $\Phi(\tau)$, the phase difference as a function of proper time. The measured $\Phi(\tau)$ can then be compared with the predicted $\Phi(\tau)$, using least-squares or other estimation techniques, to see how well theory matches the observations. These kinds of analyses are standard in Very Long Baseline Interferometry (VBLI).

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant No. PHY 00-96522. One of us (J.B.) is grateful to the McNair Scholarship Program of the University of Wisconsin, River Falls, for financial support.

[a]Electronic mail: cmw@wuphys.wustl.edu

[1]C. M. Will, "The Confrontation between General Relativity and Experiment," Living Rev. Relativ. **4**, 2001–2004 (2001), on line at http://www.livingreviews.org/Articles/Volume4/2001-4will/.

[2]Space Interferometry Mission, Project Website at http://planetquest.jpl.nasa.gov/SIM/sim_index.html.

[3]Galactic Census Project GAIA, Project Website at http://astro.estec.esa.nl/GAIA.

[4]R. Epstein and I. I. Shapiro, "Post-post-Newtonian deflection of light by the Sun," Phys. Rev. D **22**, 2947–2949 (1980).

[5]E. Fischbach and B. S. Freeman, "Second-order contribution to the gravitational deflection of light," Phys. Rev. D **22**, 2950–2952 (1980).

[6]G. W. Richter and R. A. Matzner, "Second-order contributions to gravitational deflection of light in the parametrized post-Newtonian formalism," Phys. Rev. D **26**, 1219–1224 (1982).

[7]J.-M. Gérard and S. Pireaux, "The observable light deflection angle," unpublished (gr-qc/9907034).