IOP Publishing — Classical and Quantum Gravity

Class. Quantum Grav. **27** (2010) 145013 (27pp) doi:10.1088/0264-9381/27/14/145013

Accurate light-time correction due to a gravitating mass

Neil Ashby[1,3] **and Bruno Bertotti**[2]

[1] Department of Physics, University of Colorado, Boulder, CO, USA
[2] Dipartimento di Fisica Nucleare e Teorica, Università di Pavia, Italy

E-mail: ashby@boulder.nist.gov

Received 14 December 2009, in final form 8 April 2010
Published 7 June 2010
Online at stacks.iop.org/CQG/27/145013

Abstract

This technical paper of mathematical physics arose as an aftermath of the 2002 Cassini experiment (Bertotti *et al* 2003 *Nature* **425** 374–6), in which the PPN parameter γ was measured with an accuracy $\sigma_\gamma = 2.3 \times 10^{-5}$ and found consistent with the prediction $\gamma = 1$ of general relativity. The Orbit Determination Program (ODP) of NASA's Jet Propulsion Laboratory, which was used in the data analysis, is based on an expression (8) for the gravitational delay Δt that differs from the standard formula (2); this difference is of second order in powers of m—the gravitational radius of the Sun—but in Cassini's case it was much larger than the expected order of magnitude m^2/b, where b is the distance of the closest approach of the ray. Since the ODP does not take into account any other second-order terms, it is necessary, also in view of future more accurate experiments, to revisit the whole problem, to systematically evaluate higher order corrections and to determine which terms, and why, are larger than the expected value. We note that light propagation in a static spacetime is equivalent to a problem in ordinary geometrical optics; Fermat's action functional at its minimum is just the light-time between the two end points A and B. A new and powerful formulation is thus obtained. This method is closely connected with the much more general approach of Le Poncin-Lafitte *et al* (2004 *Class. Quantum Grav.* **21** 4463–83), which is based on Synge's world function. Asymptotic power series are necessary to provide a safe and automatic way of selecting which terms to keep at each order. Higher order approximations to the required quantities, in particular the delay and the deflection, are easily obtained. We also show that in a close superior conjunction, when b is much smaller than the distances of A and B from the Sun, say of order R, the second-order correction has an *enhanced* part of order m^2R/b^2, which corresponds just to the second-order terms introduced

[3] Affiliation of the National Institute of Standards and Technology.

in the ODP. Gravitational deflection of the image of a far away source when observed from a finite distance from the mass is obtained up to $O(m^2)$.

PACS numbers: 04.20.Fy, 04.25.Nx, 95.30.Sf

List of symbols

A	event or point where the photon starts
$A(r)$	metric coefficient
B	event or point where the photon is detected
$B(r)$	metric coefficient
b	closest approach in isotropic variable
$C(r)$	metric coefficient
h	closest approach in Moyers's variable
b_0	Euclidian approximation of the same
ℓ	Euclidian arc length
m	gravitational radius
$N(r) = \sqrt{\frac{B(r)}{A(r)}}$	index of refraction
ODP	Orbit Determination Program
$p_\odot$	perihelion distance
$R = \frac{2r_A r_B}{r_A + r_B}$	harmonic mean of the distances
r	isotropic radial coordinate
$R_\odot$	radius of the Sun
$\mathbf{r}(\ell)$	photon trajectory
S	Fermat's action
$\mathfrak{S}(x^\mu)$	eikonal function
t	time in the rest frame of the mass
t_A	starting time of photon
t_B	arrival time of photon
γ	relativistic PPN coefficient
Δt	gravitational delay
Δ_s	expansion coefficients of delay (12)
λ	undefined parameter along the light path
$\rho = rN(r)$	Moyer's radial coordinate
ϕ	longitude
Φ_{AB}	elongation angle

1. Introduction

In the framework of metric theories of gravity and the PPN formalism, the main violations of general relativity—those linear in the masses—are described by a single dimensionless parameter γ. The question, at what level and how general relativity is violated, in particular how much γ differs from unity, Einstein's value, is still moot. No definite and consistent predictions about it are available, except for the inequality $\gamma < 1$, which must be fulfilled in a scalar–tensor theory, in particular those arising as the low-energy limit of certain string theories. To date, the best measurement of γ has been obtained with Cassini's experiment, which has provided the fit (at 1-σ)

$$\gamma - 1 = (2.1 \pm 2.3) \times 10^{-5}. \tag{1}$$

Einstein's prediction is still acceptable, but more accurate experiments are needed and planned.

While γ also controls other relativistic effects, in particular those related to gravitomagnetism, it mainly affects electromagnetic propagation. The differential displacement of the stellar images near the Sun historically was the first experimental effect to be investigated and is now of great importance in accurate astrometry. The bending of a light ray also increases the light-time between two points, an important effect usually named after its discoverer Shapiro [27]. Several experiments to measure this delay have been successfully carried out, using *wide-band* microwave signals passing near the Sun and transponded back, either passively by planets, or actively by space probes (see [24, 31]).

Cassini's 2002 experiment has implemented a third way to measure γ [4], in which *coherent* microwave trains sent from the ground station to the spacecraft (at that time about 7 AU far away) were transponded back continuously. The use of high-frequency carriers (in K_a band, 34 and 32 GHz) and the combination with standard X-band carriers (about 8 GHz) allowed successful elimination of the main hindrance, dispersive effects due to the solar corona traversed by the beam [3]. The tracking was carried out around the 2002 superior conjunction; the minimum value of the impact parameter of the beam was 1.6 $R_\odot$ ($R_\odot$ is the Sun's radius), but in effect only 18 passages have been used, with a minimum impact parameter of $\approx$6 $R_\odot$. The two-way total amount of phase difference between the time of emission and the time of arrival has been continuously measured in each passage. In effect, however, NASA's Deep Space Network provides the phase count in a given integration time τ. Mathematically, in the limit $\tau \to 0$ this would give the received frequency, in which Doppler effects and gravitational frequency shift are mixed up (section 3). Cassini's observable, therefore, can also be assessed in terms of the predicted change in frequency, as in [4]; however, in practice, taking τ small would introduce unacceptable high-frequency noise. The change in light-time in a given integration time is the correct, theoretically available observable.

In the standard formulation for a superior conjunction, and taking the Sun at rest, the (one-way) light-time from an event A to an event B is

$$t_B - t_A = r_{AB} + \Delta t = r_{AB} + (1+\gamma)m \ln \frac{r_A + r_B + r_{AB}}{r_A + r_B - r_{AB}}, \tag{2}$$

where $m = 1.48$ km is the gravitational radius of the Sun, r_A, r_B are, in Euclidian geometry (see figure 1, left), the distances of A and B from the Sun and r_{AB} is their distance. The velocity of light c is unity. Δt, the increase of the light-time over r_{AB}, is the *gravitational delay*.

In a *close superior conjunction* A and B are on the opposite sides of the mass and the Euclidian distance b_0 of the straight line AB from the mass fulfils, say, $b_0 \ll (r_A, r_B) = O(R)$. In this approximation equation (2) reduces to

$$t_B - t_A = r_{AB} + \Delta t = r_{AB} + (1+\gamma)m \ln \left(\frac{4 r_A r_B}{b_0^2}\right), \tag{3}$$

with a logarithmic enhancement over the formal order of magnitude $\Delta t = O(m)$.[4] Taking the logarithm equal to 10, this provides an estimate of the timing accuracy in terms of the error in γ:

$$\sigma_{\Delta t} = 1.43\sigma_\gamma \times 10^6 \text{ cm}, \tag{4}$$

[4] As stated in the supplementary material of [6], in equation (2) the two terms on the right-hand side should obviously be multiplied by a factor 2. This error, of course, had no consequence on the computer fit.

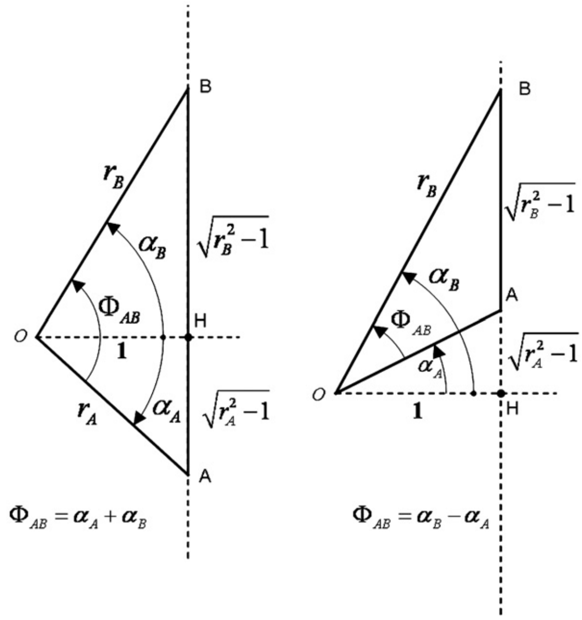

Figure 1. The background Euclidian geometry. The mass at O and the end points at A and B define a triangle AOB; the distance $OH = b_0$ from the straight line AB to the mass at the origin is taken as the unit of length in this figure. The angles α are taken positive. The internal angle Φ_{AB} can be obtuse (left) or acute (right); in the first, more interesting case, when, in addition, $r_A \geqslant r_B \gg b_0$, we have the most important case of a close superior conjunction, in which the deflection is large. Elementary trigonometry gives the relation $b_0\sqrt{r_A^2 + r_B^2 - 2r_Ar_B\cos\Phi_{AB}} = r_Ar_B\sin\Phi_{AB}$.

corresponding, in Cassini's case, to 30 cm. Equation (3) also embodies the one-way frequency change $\Delta\nu$ induced by gravity between A and B. Their motion makes b_0 (and the distances) change with time, so that, for a one-way experiment:

$$\frac{\Delta\nu}{\nu} = \frac{\mathrm{d}\Delta t}{\mathrm{d}t} = -2(\gamma+1)\frac{m}{b_0}\frac{\mathrm{d}b_0}{\mathrm{d}t}. \tag{5}$$

The basic geometric setup is straightforward: a point mass m at rest at the origin in an asymptotically flat space generates a line element with rotational symmetry. An invariant Killing time t is defined; events on each t = constant surface are 'simultaneous' and the metric components are constant. The proper time $\mathrm{d}s = \sqrt{g_{00}(r)}\,\mathrm{d}t$ of a static observer differs from $\mathrm{d}t$ by the red-shift factor $\sqrt{g_{00}(r)}$. A null geodesic runs from the event A (with radial coordinate r_A and time t_A) to the event B (with radial coordinate r_B and time t_B); it stays on a plane, taken here as the equatorial plane $\theta = \pi/2$. The (invariant) longitude difference $\Phi_{AB} = \phi_B - \phi_A$ completes the setup. In the PPN formalism and isotropic coordinates the metric reads

$$\mathrm{d}s^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}\ell^2$$
$$= \left(1 - \frac{2m}{r} + 2\beta\frac{m^2}{r^2} + \cdots\right)\mathrm{d}t^2 - \left(1 + \gamma\frac{2m}{r} + \frac{3\epsilon}{2}\frac{m^2}{r^2} + \cdots\right)\mathrm{d}\ell^2, \tag{6}$$

where

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) = \mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2$$

is the Euclidian line element. The parameters γ, β and ϵ are equal to 1 in general relativity; while γ and β are accurately known, currently no information is available about ϵ.

In our case the best mathematical tool to deal with electromagnetic propagation is not null geodesics, but the theory of the eikonal. It is known (e.g. [19]) that in this problem Fermat's Principle holds, corresponding to the refractive index

$$N(r) = \sqrt{\frac{B(r)}{A(r)}}; \tag{7}$$

we develop *ab initio* the eikonal and solve for it by separation of variables (section 3). The radial part provides Fermat's action as a radial integral containing $N(r)$ and the impact parameter h; when computed at the true value h_{true}, such action is just the required light-time. The solution can be obtained recursively, using appropriate expansions in powers of m: the expansion for h begins with $h_0 = b_0$, the distance of the straight line AB from the origin. In this way the variational nature of the problem brings about a great conceptual and algebraic simplification. At the linear approximation in m one would expect that the light-time contains h_1, the correction in the impact parameter linear in the mass; as one can see from (2), this is not the case. This property is generally true: the correction to the light-time $O(m)^k$ *does not contain* h_k (section 5).

Cassini's and many other space experiments have been analysed by the use of NASA's Orbit Determination Program (ODP), developed by NASA at Jet Propulsion Laboratory in the 1960s and steadily improved since; a new version called MONTE is under development. The ODP, whose theoretical formulation is due to Moyer [21], integrates the equations of motion of the relevant bodies and provides their trajectories in the *ephemeris time*. This task is carried out in a reference system, called BCRS (barycentric coordinate reference system) [28], in which the centre of gravity of the solar system is at rest and the Sun moves around with a velocity $v_\odot \approx 10\,\mathrm{m\,s^{-1}} = 3 \times 10^{-8}\,c$. As discussed in [2], the light-time in this frame differs from the rest frame of the Sun essentially due to Lorentz time dilatation; being of order $v_\odot$, this difference is quite below the sensitivity of the Cassini experiment [17]. We do not discuss this point any more; t is just Killing time.

The ODP uses a fictitious Euclidian space $S_3(x, y, z)$, which corresponds to the isotropic coordinates of (6). This space is just a computational convenience and should not be considered as a physical background in which gravity acts. For example, replacing r, the Euclidian distance from the origin, with $r + km$, where k is an arbitrary constant, is fully legitimate in a covariant theory, but it destroys the conformal flatness of space, introduces a gravitational potential $-km^2/r^2$ and adds a second-order term to the delay Δt. Strictly speaking, the word 'delay' is inappropriate: we just have a light-time and there is nothing with respect to which a delay can be reckoned. The object of the measurement is the time change of the delay. The arbitrariness of the radial coordinate also affects gravitational bending: its second-order approximation up to $O(m/b)^2$ depends on which radial coordinate is used (see [7, 10, 13, 25]).

It should also be noted that the spacetime coordinates of the end events are not directly provided in the experimental setup and depend on the gravitational delay Δt, the very quantity one sets out to measure. The trajectories $\mathbf{r}_A(t)$ and $\mathbf{r}_B(t)$ are given by the numerical code; the starting time t_A is just a label of the ray, but the arrival time t_B is greater than $t_A + r_{AB}$. The way out is to take for the end point

$$\mathbf{r}_B(t_B) = \mathbf{r}_B(t_A + r_{AB}) + \Delta t\,\mathbf{u}_B(t_A + r_{AB}),$$

where $\mathbf{u}_B = \mathrm{d}\mathbf{r}_B/\mathrm{d}t$. For a typical velocity $10^{-4}\,c$ the correction is of order $20 \times 1.4 \times 10^5 \times 10^{-4} = 300$ cm, and the *a priori* accuracy in Δt is sufficient.

Since for electromagnetic propagation dt and $d\ell$ in (6) are almost equal, (2) is the correct approximation to the delay to $O(m)$; one would expect this to be the first term in an expansion in powers of m/b_0, so that the next term should be

$$\approx m\frac{m}{b_0} = m\frac{m}{R_\odot}\frac{R_\odot}{b_0} = 0.3\frac{R_\odot}{b_0}\text{ cm},$$

quite below Cassini's sensitivity. The present paper arose because the ODP (equation (8–54) of [21]), in fact does not use (2), but, in our notation,

$$t_B - t_A = r_{AB} + \Delta t = r_{AB} + (1+\gamma)m\ln\left(\frac{r_A + r_B + r_{AB} + (1+\gamma)m}{r_A + r_B - r_{AB} + (1+\gamma)m}\right). \quad (8)$$

We have not been able to fully reconstruct Moyer's derivation of this expression. It introduces nonlinear corrections arising from nonlinear effects of linear metric terms, but not quadratic metric terms. However, the difference between the two expressions of the delay is much larger than the estimate above; this arises because in Cassini's case, in (2) the denominator $r_A + r_B - r_{AB}$ is much smaller than the numerator $\approx 2r_{AB}$. Indeed,

$$\Delta t - (\Delta t)_{\text{ODP}} = -2(1+\gamma)^2\frac{m^2}{b_0^2}\frac{r_A r_B}{r_A + r_B} = -(1+\gamma)^2\frac{m^2 R}{b_0^2}, \quad (9)$$

where we have introduced the harmonic mean of the distances

$$\frac{2}{R} = \frac{1}{r_A} + \frac{1}{r_B} = \frac{r_A + r_B}{r_A r_B}. \quad (10)$$

If, as in Cassini's experiment, $r_B \gg r_A = 1\text{ AU} = 200\,R_\odot$, $R = 400\,R_\odot$ the correction (9) is about

$$1600m\frac{m}{R_\odot}\left(\frac{R_\odot}{b_0}\right)^2 = 500\left(\frac{R_\odot}{b_0}\right)^2\text{ cm}.$$

Even at $\approx 6\,R_\odot$ this correction is somewhat below sensitivity (4) and it should not affect the result. However, it cannot be excluded that neglected nonlinear terms relevant for Cassini's experiment affect fit (1). One could say, (8) is not justified; a full clarification of the problem is needed.

Empirically dropping or keeping 'small' terms may lead to inconsistencies and therefore does not work; the rigourous method of *asymptotic perturbation theory* (see, e.g. [11, 16]) must be used. We briefly sketch it now at a practical level. One begins with a wise choice of a dimensionless 'smallness' parameter, and expands every function in the corresponding power series. Our main choice will be m/b_0, but convenience may suggest using other lengths, such as in m/r. An asymptotic series

$$G = \sum_s G_s\left(\frac{m}{b_0}\right)^s$$

is a formal object assigned just by the sequence of its coefficients G_s; arithmetic and calculus follow the obvious rules for sum, multiplication and differentiation. Equality between two asymptotic series just means that the coefficients of the same order are equal. The value of $G(m)$ as a function of m plays no role, and even the convergence of the series is irrelevant; what matters is only the truncated value at any order k:

$$G_{(k)} = \sum_{s=0}^{k} G_s\left(\frac{m}{b_0}\right)^s. \quad (11)$$

The remainder is of $O((m/b_0)^{k+1})$. The parameter m/b_0 should not be understood as a fixed number, but as a variable that tends to zero. The symbol $O(.)$ means *order of infinitesimal*;

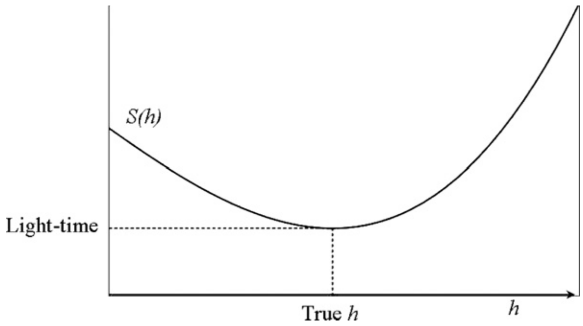

Figure 2. The minimum of the reduced action (51) is equal to the light-time at the true value h_{true}. See equation (52).

it states how fast the remainder tends to zero as the parameter diminishes. An asymptotic series can be constructed from an ordinary arbitrary function $G(m)$; however, a whole class of functions give rise to the same series; for example, if G_s is the sequence generated by $G(m)$, the same sequence is also generated by

$$[1 + P\exp(-Qb_0/m)]G(m) \qquad (Q > 0).$$

In this way any recursive iteration then proceeds automatically and safely, even in the most complex situations.

In our case light-time will be provided as an asymptotic power series

$$t_B - t_A = r_{AB} + m\sum_{s=1}\Delta_s\left(\frac{r_A}{b_0}, \frac{r_B}{b_0}\right)\left(\frac{m}{b_0}\right)^{s-1}, \tag{12}$$

with dimensionless coefficients Δ_s. Δ_1 provides the lowest standard approximation to Δt (see (2)). In principle, asymptotic analysis provides no numerical estimate of the remainder in a given situation; this is a physical, not a mathematical question. But when the problem, properly formulated, does not contain small dimensionless quantities other than the smallness parameter itself, one can expect the mathematical operations leading to the result to maintain the order of magnitude and to lead to expansions whose coefficients are numerically of the same order. This is the case of deflection, the angle between the asymptotes of the ray. There is only one length in the problem, the distance b of the point of closest approach, or, equivalently, the impact parameter $h = bN(b)$ (see figure 3); hence, in the expansion

$$\delta = \sum_s \delta_s\left(\frac{m}{h}\right)^s \tag{13}$$

the coefficients δ_s are dimensionless numbers, determined solely by the PPN parameters, and must be of order unity (see section 8). But in the delay problem the coefficients Δ_s depend on the geometrical configuration. They are of order unity in the generic (but scarcely interesting) case in which r_A, r_B and b_0 are of the same order; however, in a close superior conjunction—of crucial relevance in experimental gravitation—when $b_0 \ll (r_A, r_B) = O(R)$, besides m/b_0, there is another smallness parameter, namely b_0/R, and there is no reason to exclude that the Δ_s increase with R/b_0 beyond the expected order of magnitude unity. This we call *enhancement*. We already saw in (3) that Δ_1 is enhanced, albeit only logarithmically; the ODP correction (9), formally of second order, is enhanced by R/b_0. This could place serious limitations on the method and even invalidate the iteration itself. This would occur, for instance, when $mR \approx b_0^2$; if $b_0 = R_\odot = 1/200\,\text{AU}$, this corresponds to $R = 2000$ AU.

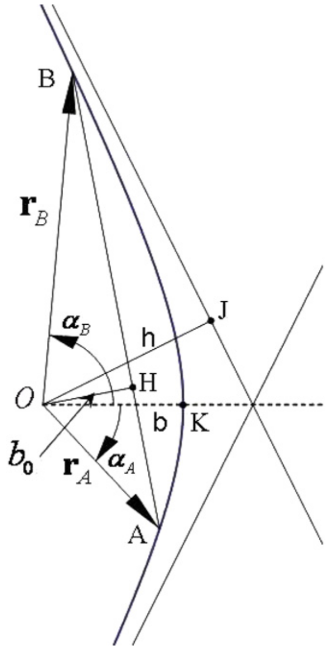

Figure 3. Three ways to define the separation of the ray from the origin: the distance $b_0 = h_0 = OH$ (in this paper often taken as unit of length) of the straight line AB; the distance $b = OK$ the point of closest approach; the impact parameter $h = bN(b) = OJ$.

The enhancement, which has never been discussed in the literature, has been fully understood and tamed in the present paper (section 7). We have found, indeed, that *the second-order terms embodied in the ODP expression (8) that was used in Cassini's experiment are just the enhanced second-order terms;* Cassini's result (1) is still safe.

The problem can be reduced to one of ordinary optics; due to its variational nature, the eikonal function can be easily solved in an expansion in powers of m/h. The second-order expression of the light-time for a static spacetime has been obtained; extension to third order is also easy. This approach should be compared with the much more general work of [18], who consider Synge's world function $\Omega(x_A, x_B)$ in a generic spacetime for a generic geodesic (not necessarily null) between two events A and B. On the basis of Hamiltonian theory, they develop a method to solve for $\Omega(x_A, x_B)$ in a formal power series with respect to the gravitational constant G and compute it up to the second order. In the null case the world function vanishes on the solution and becomes the eikonal function. Our method, limited of course to the spherically symmetric case, exploits directly the variational nature of the problem and leads to the second-order expression of the light-time, which agrees with the expression of [18]; extension to third order is also easy.

For a realistic observation of a distant source, from a point B at a finite distance r_B, (13) must be generalized to an expansion of the type

$$\delta_B = \sum_s \delta_{Bs} \left(\frac{r_B}{h}\right) \left(\frac{m}{h}\right)^s, \tag{14}$$

where h is the impact parameter. The linear term has been evaluated in [19], section 40.3; the quadratic correction will be obtained in section 8.

2. Hyperbolic Newtonian dynamics

Newtonian dynamics of a test particle attracted by a point mass M, an exactly soluble problem, illustrates these issues. We consider a motion in the equatorial plane $\theta = \pi/2$, with the radial coordinate r and the azimuthal longitude ϕ. The Lagrangian function

$$\mathcal{L}_{\text{Newton}} = \frac{1}{2}\left[\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}t}\right)^2\right] + \frac{GM}{r} \tag{15}$$

keeps the total energy $v_\infty^2/2$ constant; v_∞, the ultimate speed of the particle at a large distance, plays a role analogous to the speed of light and we can define a new time coordinate by replacing $v_\infty t$ by t. Energy conservation gives

$$\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}t}\right)^2 - 2\frac{m}{r} = 1, \tag{16}$$

where $m = GM/v_\infty^2$ is the gravitational radius. ϕ is an ignorable coordinate, so that the angular momentum

$$\frac{\partial\mathcal{L}_{\text{Newton}}}{\partial(\mathrm{d}\phi/\mathrm{d}t)} = r^2\frac{\mathrm{d}\phi}{\mathrm{d}t} = h \tag{17}$$

is constant. Since the trajectory at infinity is straight, h is the impact parameter. Eliminating $\mathrm{d}t$ we get

$$r\frac{\mathrm{d}\phi}{\mathrm{d}r} = \pm\frac{h}{\sqrt{r(r+2m) - h^2}}; \tag{18}$$

hence,

$$h = \sqrt{b(b+2m)} \tag{19}$$

determines b, the distance of the closest approach where $\mathrm{d}r/\mathrm{d}\phi = 0$. The sign depends upon whether the ray is ingoing or outgoing. Integrating (18) we get the true anomaly α:

$$\alpha = \arccos\left(\frac{b^2 + 2mb - mr}{r(b+m)}\right). \tag{20}$$

Alternatively, the motion can be expressed in terms of the semi-major axis $a = m$ and the hyperbolic eccentricity $e = 1 + b/m$:

$$r = \frac{a(e^2 - 1)}{1 + e\cos f}. \tag{21}$$

The acute angle δ between the asymptotes is given by

$$\sin\delta = \sin\left(2\arccos\left(-\frac{1}{e}\right)\right) = \frac{2m}{b+m}\sqrt{1 - \frac{m^2}{(b+m)^2}}. \tag{22}$$

This angle has a regular expansion in powers of m/b, with no enhancement.

Consider, however, the hyperbola determined by the two points A and B on the opposite sides of the vertex (left side in figure 1). As in space navigation—in particular in the ODP—the end points are provided in terms of the initial and final position vectors, or equivalently, in terms of the initial and final distances r_A and r_B and the elongation angle Φ_{AB}; the 'unperturbed distance of the closest approach' may then be calculated from elementary geometry:

$$b_0 = \frac{r_A r_B \sin\Phi_{AB}}{\sqrt{r_A^2 + r_B^2 - 2r_A r_B\cos\Phi_{AB}}}. \tag{23}$$

Choosing the angles α_A, α_B and Φ_{AB} positive, we can express b in terms of b_0 with the condition

$$\Phi_{AB} = \alpha_A + \alpha_B = \arccos\frac{b_0}{r_A} + \arccos\frac{b_0}{r_B}$$
$$= \arccos\left(\frac{b(b+2m)}{r_A(b+m)} - \frac{m}{b+m}\right) + \arccos\left(\frac{b(b+2m)}{r_B(b+m)} - \frac{m}{b+m}\right), \tag{24}$$

from (20). The symmetric case $r_A = r_B = R$ is sufficient to exhibit the problem. The condition (24) reads

$$\frac{b}{R}\frac{b+2m}{b+m} - \frac{m}{b+m} - \frac{b_0}{R} = 0, \tag{25}$$

or

$$b^2 + (2m - b_0)b - m(R + b_0) = 0, \tag{26}$$

with the solution

$$\frac{2b}{b_0} = \frac{b_0 - 2m}{b_0} + \sqrt{1 + 4\frac{m}{b_0}\frac{R+m}{b_0}}. \tag{27}$$

Expansion in powers of m gives

$$\frac{b}{b_0} = 1 + \left(\frac{R}{b_0} - 1\right)\frac{m}{b_0} - \left(\left(\frac{R}{b_0}\right)^2 - 1\right)\left(\frac{m}{b_0}\right)^2 + O\left(\frac{mR}{b_0^2}\right)^3. \tag{28}$$

The enhancement is clear: when $R = O(b_0)$ the truncation error at order k is $O(m/b_0)^{k+1}$, with a coefficient of order unity, as naïvely expected; however, when—as in a close superior conjunction—$R \gg b_0$, the error is larger, $O\left(mR/b_0^2\right)^{k+1}$. Formally this requires introducing another smallness parameter b_0/R and expanding every coefficient of the primary m-expansion in descending powers of R/b_0. Of course, the condition

$$\frac{mR}{b_0^2} \ll 1 \tag{29}$$

must be fulfilled, lest the whole procedure breaks down. One could say, anchoring the trajectory at far away end points has a lever effect, so that an increase in the mass produces a large increase in the closest approach.

The quantity (29) gives, in order of magnitude, the ratio between the deflection $\approx m/b_0$ and the angle b_0/R which separates the central mass and a distant star, as seen from a distance R. Hence the limiting constraint above implies that the geometry of astronomical deflection is the same as in the classical case (see figure 4): sources in the sky near the Sun are displaced outwards by an amount inversely proportional to the angular distance. The transition through the milestone $mR = b_0^2$ marks the passage to the gravitational lensing regime, in which the image can appear on both sides.

In section 7 the light-time enhancement is dealt with in the general case and it is shown that the dimensionless coefficients Δ_s in (12) are $O(R/b_0)^{s-1}$.

3. The radial gauge

The metric of a spherical body at rest has the general form

$$ds^2 = A(r)\,dt^2 - B(r)\,dr^2 - C(r)r^2\,d\Omega^2, \tag{30}$$

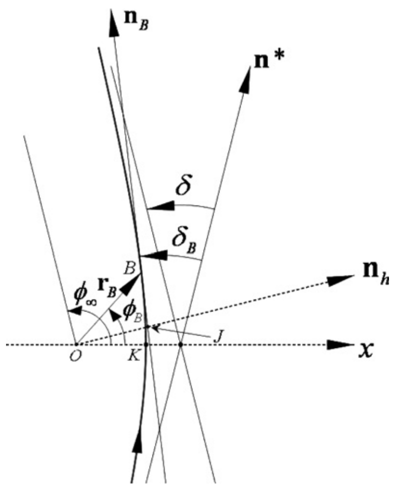

Figure 4. Deflection measured from a finite distance. A ray from a faraway source arriving along the direction $\mathbf{n}^\star$ is deflected, and arrives at the observation point B from a different direction, with unit vector $\mathbf{n}_B$ (equation (111)) tangent to the ray. The deflection angle δ_B is smaller than the asymptotic deflection δ, the angle between the asymptotes. The origin of longitudes is taken on the axis OK through the closest approach, so that $\phi_\infty = \pi/2 + \delta/2$. The figure also illustrates the meaning of the parameter h. The point J is the intersection of the tangent through B with a line through O perpendicular to the asymptote. It is easily seen that the distance $OJ = rh/\sqrt{r^2N(r)^2 - h^2}$, so that at great distance this distance becomes h, which is therefore just the *impact parameter*. In the Newtonian dynamical model h (17) is a constant of the motion, with the same meaning.

where $A(r)$, $B(r)$, $C(r)$ are the power series of the form

$$A(r) = \sum_s A_s \left(\frac{m}{r}\right)^s. \tag{31}$$

It is asymptotically flat, so that $A_0 = B_0 = C_0 = 1$. The radial coordinate is otherwise arbitrary; this is the *gauge freedom* at our disposal. For consistency, however, any change $r \to \bar{r} = g(r)$ must become an identity at infinity and have a similar expansion:

$$g(r) = r + g_1 m + g_2 \frac{m^2}{r} + \cdots; \tag{32}$$

the coefficients A_s, B_s, C_s are not gauge invariant. Two gauges are common. In the *isotropic* form—the canonical choice in space physics—$C(r) = B(r)$, so that

$$\mathrm{d}s^2 = A(r)\,\mathrm{d}t^2 - B(r)(\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2) = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}\ell^2; \tag{33}$$

the space part of the metric is conformally flat. We define

$$N(r) = \sqrt{\frac{B(r)}{A(r)}} = \sum_s N_s \left(\frac{m}{r}\right)^s = 1 + N_1\frac{m}{r} + N_2\left(\frac{m}{r}\right)^2 + O\left(\frac{m}{r}\right)^3. \tag{34}$$

In the PPN scheme (e.g. [31])

$$N_1 = \gamma + 1, \qquad N_2 = \frac{6 - 4\beta + 3\epsilon + 4\gamma - 2\gamma^2}{4}. \tag{35}$$

In 'Schwarzschild' gauge $\bar{C}(\bar{r}) = 1$ and

$$\mathrm{d}s^2 = \bar{A}(\bar{r})\,\mathrm{d}t^2 - \bar{B}(\bar{r})\,\mathrm{d}\bar{r}^2 - \bar{r}^2\,\mathrm{d}\Omega^2;$$

the area of a sphere of radius $\bar{r}$ is just the Euclidian expression $4\pi\bar{r}^2$, which defines $\bar{r}$ in an invariant way. In the original Schwarzschild solution $\bar{A}(\bar{r}) = 1/\bar{B}(\bar{r}) = 1 - 2m/\bar{r}$. To get the isotropic form one requires

$$g^2(r) = \bar{B}(g(r))\left(\frac{\mathrm{d}g}{\mathrm{d}r}\right)^2 r^2; \tag{36}$$

to first order

$$\bar{r} = r + \gamma m + \cdots. \tag{37}$$

In the present paper a third radial coordinate

$$\rho = rN(r) = r\sqrt{\frac{B(r)}{A(r)}} = r + mN_1 + m^2\frac{N_2}{r} + \cdots \tag{38}$$

plays an important role. It is a monotonic function of r and ensures that $A(\rho) = C(\rho)$. In the linear approximation it was introduced by Moyer in [21] (equation (8)–(23)), and boils down to just adding to r a constant term, equal to 2.95 km for the Sun.

4. Geometrical optics

It is convenient to reduce the problem to geometrical optics by use of the eikonal function $\mathfrak{S}$. In a generic spacetime $\mathfrak{S}$ fulfils the eikonal equation

$$g^{\mu\nu}\partial_\mu\mathfrak{S}\partial_\nu\mathfrak{S} = 0; \tag{39}$$

its characteristics are the null rays (see, e.g. [1]). $\mathfrak{S}$ is the phase of the electromagnetic wave. Let $r_A^\mu = r^\mu(s_A), r_B^\mu = r^\mu(s_B)$ be the trajectories of the end points of the light path, given as functions of their proper times s_A, s_B; let

$$v_A^\mu = \frac{\mathrm{d}r^\mu}{\mathrm{d}s_A}, \qquad v_B^\mu = \frac{\mathrm{d}r^\mu}{\mathrm{d}s_B}$$

be the corresponding four-velocities. Clocks associated with them measure the proper angular frequencies

$$\omega_A = -v_A^\mu\partial_\mu\mathfrak{S} = -\frac{\mathrm{d}\mathfrak{S}}{\mathrm{d}s_A}, \qquad \omega_B = -v_B^\mu\partial_\mu\mathfrak{S} = -\frac{\mathrm{d}\mathfrak{S}}{\mathrm{d}s_B}. \tag{40}$$

In the simple case in which the end points are far away from the source, where the metric corrections can be neglected, the contribution to the frequency difference corresponds to the ordinary Doppler effect, and can be evaluated with a slow motion expansion; the change in $\mathfrak{S}$ between A and B is determined by the accumulated gravitational effect along the ray and mainly comes from the region near the mass:

$$g^{\mu\nu}\partial_\mu\mathfrak{S}\partial_\nu\mathfrak{S} = 0 = N^2(\mathbf{r})(\partial_t\mathfrak{S})^2 - \nabla\mathfrak{S}\cdot\nabla\mathfrak{S}, \tag{41}$$

where ∇ is the Euclidian gradient operator. We are really interested only in the spherically symmetric case, but the reasoning of this section holds also for an arbitrary $N(\mathbf{r})$.

$\mathfrak{S}$ is the phase; propagation occurs keeping it constant. Separating space and time variables with

$$\mathfrak{S} = \mathfrak{S}_t(t) + \overline{\mathfrak{S}}_{\mathbf{r}}(\mathbf{r}),$$

leads to the class of solutions

$$\mathfrak{S} = \omega_0(\overline{\mathfrak{S}}(\mathbf{r}) - t), \tag{42}$$

where $\omega_0\overline{\mathfrak{S}}(\mathbf{r})$ is the spatial part of the phase and ω_0 is a constant frequency. $\overline{\mathfrak{S}}$ has the dimension of time and satisfies

$$\nabla\overline{\mathfrak{S}} \cdot \nabla\overline{\mathfrak{S}} = N^2(\mathbf{r}). \tag{43}$$

If a clock is at rest relative to the mass, $v^\mu = (1, \mathbf{0})/\sqrt{A(r)}$ is its four-velocity and the measured proper frequency $\omega_0/\sqrt{A(r)}$ includes the appropriate gravitational shift away from the asymptotic value ω_0. This is enough to reduce the problem to geometrical optics (see, e.g. [8], chapter III). A ray $\mathbf{r}(\ell)$, as a function of the Euclidian arc length ℓ, is orthogonal to the eikonal surfaces $\overline{\mathfrak{S}}(\mathbf{r}) = \text{const}$ and fulfils

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\left(N(\mathbf{r})\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\ell}\right) = \nabla N(\mathbf{r}). \tag{44}$$

The index of refraction is the rate of increase of the spatial phase along the ray:

$$\frac{\mathrm{d}\overline{\mathfrak{S}}}{\mathrm{d}\ell} = N(\mathbf{r}).$$

Consider now Fermat's action functional

$$S[\mathbf{r}(\lambda)] = \int_{\lambda_A}^{\lambda_B} \mathrm{d}\lambda\, N(\mathbf{r})\sqrt{\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\lambda}\cdot\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\lambda}} = \int_{\lambda_A}^{\lambda_B} \mathrm{d}\lambda\, \mathcal{L}_F, \tag{45}$$

where the trajectory, any path joining the end points, is expressed in terms of a generic parameter λ:

$$\mathbf{r}(\lambda_A) = \mathbf{r}_A, \qquad \mathbf{r}(\lambda_B) = \mathbf{r}_B. \tag{46}$$

Because the action is, in fact, independent of the choice of λ, no generality is lost if $\mathrm{d}\lambda = \mathrm{d}\ell$, the Euclidean line element. The Euler–Lagrange equation for the action (45) reduces to (44). The actual elapsed time

$$t_B - t_A = S(A, B) = \int_{\ell_A}^{\ell_B} \mathrm{d}\ell\, N(r) = \overline{\mathfrak{S}}_B - \overline{\mathfrak{S}}_A \tag{47}$$

is just the value of $S[.]$ computed at a local minimum—the actual ray (*Fermat's Principle*). One should keep in mind the distinction between the action functional, with its argument in square brackets, and the action computed at the extremum, an ordinary function of the end points denoted with $S(A, B)$. In $S(A, B)$, but not in $S[.]$, it is allowed to replace the generic independent variable λ with a more convenient one related to the solution, such as r. For simplicity, the different functions denoted by the symbol S are distinguished by their arguments; below, the quantity $S(r_A, r_B; b) = S(h)$ will be introduced to denote the action corresponding to a ray anchored at r_A and r_B, but with arbitrary b (or h).

5. The solution

The eikonal function provides a deep simplification in the evaluation of the light-time. Having already separated out the time, the three-dimensional eikonal equation (43) in spherical symmetry and in the equatorial plane can be solved by separating out the longitude ϕ: setting

$$\overline{\mathfrak{S}}(r, \phi) = \overline{\mathfrak{S}}_r(r) + \overline{\mathfrak{S}}_\phi(\phi).$$

It satisfies[5]

$$r^2(\overline{\mathfrak{S}}_r')^2 + (\overline{\mathfrak{S}}_\phi')^2 = r^2N^2(r),$$

[5] For a function of a single variable, a prime indicates the derivative.

so that $\overline{\mathfrak{S}}'_\phi$ is a constant. Setting $\overline{\mathfrak{S}}_\phi = h\phi$, the eikonal equation reduces to

$$(\overline{\mathfrak{S}}'_r)^2 = \frac{1}{r^2}(r^2N^2(r) - h^2),$$

with the primitive

$$\overline{\mathfrak{S}}(r) = \pm \int^r \frac{\mathrm{d}r}{r}\sqrt{r^2N^2(r) - h^2}.$$

The + and the − signs correspond, respectively, to outgoing and incoming photons. The radial coordinate of the closest approach b, where $\overline{\mathfrak{S}}'_r = 0$, is the solution of

$$bN(b) = h; \tag{48}$$

since $r \geqslant b$, $\mathfrak{S}$ is a real function. In section 8 it will be shown that h, just as in the Newtonian case, is the impact parameter (figure 4). The total phase is therefore

$$\mathfrak{S} = \omega_0 \left(h\phi \pm \int^r \frac{\mathrm{d}r}{r}\sqrt{r^2N^2(r) - h^2} - t \right). \tag{49}$$

A wavefront propagates keeping $\mathfrak{S}$ constant, so that the time along the ray is

$$t = \pm \int^r \frac{\mathrm{d}r}{r}\sqrt{r^2N^2(r) - h^2} + h\phi. \tag{50}$$

In the usual case (see figure 1), in which the point of the closest approach is within the angle $\widehat{AOB}$, the ray has two branches, both taken with the positive sign: an incoming one from r_A to b and an outgoing one from b to r_B. In the inferior case b is never reached and we have just an outgoing ray from r_A to r_B. In both cases, ingoing from A to B the longitude increases by $\phi_B - \phi_A = \Phi_{AB}$. The quantity

$$S(h) = \int_b^{r_B} \frac{\mathrm{d}r}{r}\sqrt{r^2N^2(r) - h^2} \pm \int_b^{r_A} \frac{\mathrm{d}r}{r}\sqrt{r^2N^2(r) - h^2} + h\Phi_{AB} \tag{51}$$

gives the phase change, hence the light-time, between the end points, but the quantity h is still arbitrary. The upper (lower) sign corresponds to the case of superior (inferior) conjunction; in the latter case the two integrals combine in a single one from r_A to r_B, and b disappears as a lower limit. Equation (51) is what Fermat's action functional becomes when its variability is restricted to h and the longitude constraint is not imposed; it shall be called *reduced action*. At the true value it satisfies

$$S'(h_{\mathrm{true}}) = 0, \tag{52}$$

keeping the end points fixed. This is illustrated in figure 2.

The present work aims at providing the theoretical foundation for the time delay in all configurations; the sign freedom allows dealing with both cases at the same time, but applications will be given mainly for a conjunction, with +. The origin of longitudes is arbitrary. This general approach is relevant, for example, for a spacecraft on an almost parabolic orbit, as in the Solar Probe concept; with a perihelion as low as 4 $R_\odot$, it can have a strong enhancement of the light-time even in the inferior configuration.

In the derivative $S'(h)$ there are no contributions from the lower limits; then (52) provides h as an implicit function of the total elongation Φ_{AB}:

$$\Phi_{AB} + \int_b^{r_B} \frac{\mathrm{d}r}{r}\frac{-h}{\sqrt{(rN(r))^2 - h^2}} \pm \int_b^{r_A} \frac{\mathrm{d}r}{r}\frac{-h}{\sqrt{(rN(r))^2 - h^2}} = 0. \tag{53}$$

Hence (51) reads[6]

$$S(h) = \int_b^{r_B} \mathrm{d}r\, N(r) \frac{rN(r)}{\sqrt{(rN(r))^2 - h^2}} \pm \int_b^{r_A} \mathrm{d}r\, N(r) \frac{rN(r)}{\sqrt{(rN(r))^2 - h^2}}. \tag{54}$$

Both integrals are convergent (and in the inferior case the singularity at $rN(r) = h$ is not even reached). Equation (51) suggests the introduction of the function

$$G(r, h) = \int_b^r \frac{\mathrm{d}r}{r} \sqrt{(rN(r))^2 - h^2}\,, \tag{55}$$

in terms of which

$$S(h) = G(r_B, h) \pm G(r_A, h) + h\Phi_{AB}. \tag{56}$$

Equation (53) reads[7]

$$G_{,h}(r_B, h) \pm G_{,h}(r_A, h) + \Phi_{AB} = 0. \tag{57}$$

While in (51) h is an independent parameter, in (54) it is fixed by (53).

This expression for h can also be derived directly from Fermat's Principle, thus providing its significance. Fermat's action (47), expressed as a function of r, has the Lagrange functional

$$\mathcal{L}_F[\phi(r)] = N(r)\sqrt{1 + r^2(\mathrm{d}\phi/\mathrm{d}r)^2}, \tag{58}$$

with the (positive) constant of the motion

$$\frac{\partial \mathcal{L}_F}{\partial(\mathrm{d}\phi/\mathrm{d}r)} = \pm \frac{r^2 N(r)}{\sqrt{1 + r^2(\mathrm{d}\phi/\mathrm{d}r)^2}} \frac{\mathrm{d}\phi}{\mathrm{d}r} = h. \tag{59}$$

The upper (lower) holds for the outgoing (incoming) branch. Thus,

$$r\frac{\mathrm{d}\phi}{\mathrm{d}r} = \pm \frac{h}{\sqrt{r^2 N^2(r) - h^2}} = \pm \frac{h}{\sqrt{\rho^2 - h^2}}, \tag{60}$$

and then integrating, (53) is recovered. Comparison with (18) shows that in the Newtonian case the *exact* index of refraction is

$$N_{\mathrm{Newton}}(r) = \sqrt{1 + 2\frac{m}{r}}, \tag{61}$$

corresponding, as expected, to $\gamma = 0$, $N_1 = 1$ and $N_2 = -N_3 = -1/2$, etc.

6. A variational argument

At this point one could proceed as follows: using power series, solve (53) for h in terms of Φ_{AB}, a known quantity. The value of h, inserted into (54), provides the required light-time. The stationary character of action (52), however, brings about a deep and important simplification. This is already tacitly applied in the usual derivation of the gravitational delay (2). To first order, the integral of $\mathrm{d}t = N(r)\,\mathrm{d}\ell$ in (33) reads

$$t_B - t_A = \int_{\ell_A}^{\ell_B} \mathrm{d}\ell + (\gamma + 1) \int_{\ell_A}^{\ell_B} \mathrm{d}\ell \frac{m}{r};$$

the second integral can be carried out along the straight path from A to B, leading to the characteristic logarithmic term. In principle, however, the first integral should take into account the (first-order) deflection; we should understand $\int \mathrm{d}\ell = \ell_{AB}$ as the Euclidian length

[6] In a slightly inconsistent notation, we often use h to denote both an independent and variable quantity, and the fixed value h_{true} determined by the elongation. The context should be sufficient to clear the ambiguity.

[7] The suffix $_{,h}$ indicates partial derivative.

of the bent arc between A and B. But the length r_{AB} of the straight segment AB is a minimum in the set of all curves joining A and B, so that $\ell_{AB} - r_{AB}$ vanishes to $O(m)$.[8] *Ray bending is irrelevant here.*

In order to exploit the variational nature of the problem it is convenient to apply power expansions *before* imposing the extremum condition (52). We just need the value of the reduced action (51) $S(h) = \sum_s m^s S_s(h)$ at the value which fulfils (52), namely $0 = \sum_s m^s S_s'(h)$. Setting $h = h_0 + mh_1 + m^2 h_2$ and expanding $\Sigma_s m^s S_s(h_0 + mh_1 + m^2 h_2)$, the solution to second order is obtained iteratively:

$$S_0'(h_0) = 0, \tag{62}$$

$$h_1 S_0''(h_0) + S_1'(h_0) = 0, \tag{63}$$

$$h_2 S_0''(h_0) + \frac{h_1^2}{2} S_0'''(h_0) + h_1 S_1''(h_0) + S_2'(h_0) = 0. \tag{64}$$

In the expression

$$S(h) = S_0(h_0) + m(h_1 S_0'(h_0) + S_1(h_0)) + m^2 \left(h_2 S_0'(h_0) + \frac{h_1^2}{2} S_0''(h_0) + h_1 S_1'(h_0) + S_2(h_0) \right), \tag{65}$$

the effect of the extremum property is clear: because $S_0'(h_0) = 0$, the first-order term does not contain h_1, and the second-order term does not contain h_2; in general, the term in $S(h)$ of order m^k does not depend on h_k. This important result is reflected in the general approach of [18]. Referring to the equation numbering of [18], their world function $\Omega(x_A, x_B)$ fulfils the Hamilton–Jacobi equation (30). In the null case $\Omega = 0$, (30) becomes the eikonal equation. Their theorem 2 proves that the nth-order Ω^n can be expressed in terms of integrals along the Minkowskian path of lowest order. In our case this *variational lemma* (namely that the term in $S(h)$ of order m^k does not depend on h_k) clarifies the matter and produces considerable simplifications. Using (63), the light-time to second order reads

$$S(h) = S_0(h_0) + mS_1(h_0) + m^2 \left(\frac{h_1^2}{2} S_0''(h_0) + h_1 S_1'(h_0) + S_2(h_0) \right), \tag{66}$$

where h_1 is given by (63). The delay coefficients (12) read

$$\Delta_1 = S_1(h_0), \qquad \Delta_2/h_0 = \frac{h_1^2}{2} S_0''(h_0) + h_1 S_1'(h_0) + S_2(h_0). \tag{67}$$

The second-order correction in the impact parameter h_2, given by (64), is needed only at third and higher orders. For the record, note the third-order contribution to the light-time:

$$\Delta_3/h_0^2 = h_1 h_2 S_0'' + \frac{h_1^3}{6} S_0''' + h_2 S_1' + \frac{h_1^2}{2} S_1'' + h_1 S_2' + S_3, \tag{68}$$

where, for simplicity, the arguments h_0 have been understood, and h_2 is provided by (64).

7. Power series

We now proceed to apply this simple and general lemma to the light-time. To lowest order, in (55) we use $b = b_0 = h_0$ and $N(r) = 1$, so that

$$S_0(h) = \sqrt{r_B^2 - h^2} \pm \sqrt{r_A^2 - h^2} - h \left(\arccos \frac{h}{r_B} \pm \arccos \frac{h}{r_A} - \Phi_{AB} \right). \tag{69}$$

[8] A didactical remark is in order here. This minimum property, crucial to the argument, is often omitted in the usual derivation. See, e.g. [19] p 1107, [9] p 125; in equation (17.59) of [5], p 581, the minimum is not mentioned, and a factor of 4 is missing in the argument of the logarithm.

The condition $S_0'(h_0) = 0$ determines h_0 with the trigonometric relation (see figure 1)

$$\Phi_{AB} = \arccos\frac{h_0}{r_B} \pm \arccos\frac{h_0}{r_A}, \tag{70}$$

or

$$h_0 = \frac{r_A r_B}{r_{AB}} \sin\Phi_{AB}. \tag{71}$$

Therefore,

$$S_0(h_0) = \sqrt{r_B^2 - h_0^2} \pm \sqrt{r_A^2 - h_0^2} = r_{AB} = \sqrt{r_A^2 + r_B^2 - 2r_A r_B \cos\Phi_{AB}}, \tag{72}$$

is the geometric distance AB. h_0 is now fixed and could be taken equal to unity without loss of generality. Because of the variational lemma, at the next order we can retain $h = h_0$. $S_1(h)$ is

$$S_1(h) = N_1\left[\ln\left(\frac{r_B + \sqrt{r_B^2 - h^2}}{h}\right) \pm \ln\left(\frac{r_A + \sqrt{r_A^2 - h^2}}{h}\right)\right]. \tag{73}$$

Then (51), with $N(r) = 1 + mN_1/r$, reads

$$S_0(h_0) + mS_1(h_0) = r_{AB} + mN_1\left[\ln\left(\frac{r_B + \sqrt{r_B^2 - h_0^2}}{h}\right) \pm \ln\left(\frac{r_A + \sqrt{r_A^2 - h_0^2}}{h}\right)\right]. \tag{74}$$

In the case of superior conjunction (with the + sign) the logarithm has the argument

$$\left(r_B + \sqrt{r_B^2 - h_0^2}\right)\left(r_A + \sqrt{r_A^2 - h_0^2}\right) = \frac{r_A + r_B + r_{AB}}{r_A + r_B - r_{AB}}, \tag{75}$$

as easily checked by cross multiplication using (72); then the standard expression (2) is properly recovered. (See the appendix for the confusion that can arise due to the gauge freedom and the difference between the closest approach and the distance b_0.) In the inferior case, instead,

$$t_B - t_A = r_{AB} + mN_1 \ln\left(\frac{r_B + \sqrt{r_B^2 - h_0^2}}{r_A + \sqrt{r_A^2 - h_0^2}}\right).$$

Before proceeding to the next order we need to evaluate h_1 with (63). Differentiating (69) twice we easily get

$$h_1\left[\frac{1}{\sqrt{r_B^2 - h_0^2}} \pm \frac{1}{\sqrt{r_A^2 - h_0^2}}\right] = \frac{N_1}{h_0}\left[\frac{r_B}{\sqrt{r_B^2 - h_0^2}} \pm \frac{r_A}{\sqrt{r_A^2 - h_0^2}}\right]. \tag{76}$$

In section 8 the superior case in which $r_A \to \infty$ will be considered; it simply gives

$$h_1 = \frac{N_1 r_B}{h_0}. \tag{77}$$

Considerable simplification may be achieved with the aid of the identities:

$$\sqrt{r_B^2 - h_0^2} = r_B(r_B - r_A\cos\Phi_{AB})/r_{AB}; \tag{78}$$

$$\sqrt{r_A^2 - h_0^2} = \pm r_A(r_A - r_B\cos\Phi_{AB})/r_{AB}. \tag{79}$$

In both cases the expression for h_1 becomes

$$h_1 = N_1\left(\frac{r_A + r_B}{r_{AB}}\right)\left(\frac{1 - \cos\Phi_{AB}}{\sin\Phi_{AB}}\right). \tag{80}$$

It is useful to record the value of $b_1 = h_1 - N_1$:

$$b_1 \left[\frac{1}{\sqrt{r_B^2 - h_0^2}} \pm \frac{1}{\sqrt{r_A^2 - h_0^2}} \right] = \frac{N_1}{h_0} \left[\sqrt{\frac{r_B - h_0}{r_B + h_0}} \pm \sqrt{\frac{r_A - h_0}{r_A + h_0}} \right]. \tag{81}$$

Enhancement is in progress: in the superior case, with the + sign, the elongation comes close to π and h_1 becomes large, as discussed in the following section. In the inferior case h_1 remains of order unity.

An expansion of $G(r, h)$ in powers of m generates a corresponding expansion of $S(h)$ (see equation (87) below). At the next order (see (65)), we need $G_2(r, h)$, $G_1(r, h)$ and its first derivative with respect to h, and $G_0(r, h)$ with its first and second derivatives (see (55)). At order s we need $G_0(r, h)$ with its first s derivatives. If these differentiations are carried out *before* the integration, a technical difficulty arises. h appears both in the lower limit and in the square root. In the superior case, already at the second order each of the two contributions diverges; the second derivative of the integrand, for instance, has a non-integrable term $\propto \left(r^2 - h_0^2\right)^{-3/2}$; it turns out, however, that this divergence is compensated by the lower limit contribution. At higher orders the complexity increases. In the inferior case the singular point is not within the integration domain and no hindrance arises. This suggests that the integration is best carried out first, leading to a finite result whose differentiation is straightforward.

The hindrance arises because as $m \to 0$ the singular point at $r = b$ moves. The integration variable

$$u(r) = \frac{rN(r)}{bN(b)} = \frac{\rho}{h} \tag{82}$$

keeps the singularity fixed at $u = 1$ and cures the problem. Note the appearance of Moyer's radial coordinate

$$\rho = rN(r) = r + mN_1 + m^2 \frac{N_2}{r}. \tag{83}$$

Then (55) reads

$$G(r, h) = h \int_1^{u(r)} \mathrm{d}u \frac{\mathrm{d} \ln r(u)}{\mathrm{d}u} \sqrt{u^2 - 1}. \tag{84}$$

$r(u)$, the inverse of $u(r)$, is itself a power series, so that

$$\frac{\mathrm{d} \ln r(u)}{\mathrm{d}u} = \sum_{s=0} \left(\frac{m}{h}\right)^s \frac{C_s}{u^{s+1}} = \frac{1}{u} + \frac{m}{h} \sum_{r=0} \left(\frac{m}{h}\right)^r \frac{C_{r+1}}{u^{r+2}} = \frac{1}{u} + \frac{m}{h} q(u). \tag{85}$$

We have split out the main part $1/u$ from the correction $O(m/h)$. C_s are the numbers $O(m^0)$ constructed with the set $\{N_k\}$:

$$C_0 = 1, \quad C_1 = N_1, \quad C_2 = N_1^2 + 2N_2, \quad C_3 = N_1^3 + 6N_1N_2 + 3N_3, \ldots. \tag{86}$$

Hence

$$G(r, h) = h \sum_s \left(\frac{m}{h}\right)^s C_s J_s(u) = \sum_s m^s G_s(r, h), \tag{87}$$

where

$$J_s(u) = \int_1^u \mathrm{d}u \frac{\sqrt{u^2 - 1}}{u^{s+1}} \tag{88}$$

are the elementary functions. Except for constant contributions, their power expansions for large u are odd (even) and for s even (odd). As implied in equation (87), h is not expanded in the functions G_s.

With this general formalism we can draw an interesting conclusion about enhancement, which corresponds to the limit $(u_A, u_B) \gg 1$. When $u \gg 1$ the functions $J_s(u)$ converge to a finite limit of order unity, except for $J_0(u) \to u$ and $J_1(u) \to \ln u$; hence, when h is fixed, at higher order no enhanced terms arise in $G(r, h)$ and in the reduced action. *Enhancement occurs only when h itself is expanded and expressed in terms of the geometric distances of the end points*, just as it happens in the case of Newtonian hyperbolic motion.

Using the universal integration variable u, the second-order contribution to the light-time in (67) has been calculated out with the aid of a computer algebra code. We have

$$S_2(h) = \frac{N_1^2}{2}\left(\frac{1}{\sqrt{r_B^2 - h^2}} \pm \frac{1}{\sqrt{r_A^2 - h^2}}\right) + \frac{1}{2h}\left(N_1^2 + 2N_2\right)\left(\arccos\frac{h}{r_B} \pm \arccos\frac{h}{r_A}\right). \tag{89}$$

With the help of (70), (78) and (79), this expression reduces to

$$S_2(h_0) = \frac{N_1^2 r_{AB}^3}{2r_A r_B(r_B - r_A\cos\Phi_{AB})(r_A - r_B\cos\Phi_{AB})} + \frac{(N_1^2 + 2N_2)}{2h_0}\Phi_{AB}. \tag{90}$$

The last line in (65) requires the derivatives $S_1'(h_0)$ and $S_0''(h_0)$:

$$\begin{aligned} S_1'(h_0) &= -\frac{N_1}{h_0}\left(\frac{r_B}{\sqrt{r_B^2 - h_0^2}} \pm \frac{r_A}{\sqrt{r_A^2 - h_0^2}}\right) \\ &= -\frac{N_1 r_{AB}(r_A + r_B)(1 - \cos\Phi_{AB})}{h_0(r_B - r_A\cos\Phi_{AB})(r_A - r_B\cos\Phi_{AB})}; \end{aligned} \tag{91}$$

$$\begin{aligned} S_0''(h_0) &= \left(\frac{1}{\sqrt{r_B^2 - h_0^2}} \pm \frac{1}{\sqrt{r_A^2 - h_0^2}}\right) \\ &= \frac{r_{AB}^3}{r_A r_B(r_B - r_A\cos\Phi_{AB})(r_A - r_B\cos\Phi_{AB})}. \end{aligned} \tag{92}$$

The second term in parentheses in (65) is therefore

$$-\frac{1}{2}\frac{S_1'^2(h_0)}{S_0''(h_0)} = -\frac{N_1^2 r_{AB}(r_A + r_B)^2(1 - \cos\Phi_{AB})^2}{2r_B r_A\sin^2\Phi_{AB}(r_B - r_A\cos\Phi_{AB})(r_A - r_B\cos\Phi_{AB})}. \tag{93}$$

Combining this with the last term in parentheses in (65), we obtain

$$\begin{aligned} S_2(h_0) - \frac{1}{2}\frac{S_1'^2(h_0)}{S_0''(h_0)} &= -\frac{N_1^2 r_{AB}}{r_A r_B(1 + \cos\Phi_{AB})} + \frac{N_1^2 + 2N_2}{2h_0}\Phi_{AB} \\ &= -\frac{N_1^2 r_{AB}}{r_A r_B(1 + \cos\Phi_{AB})} + \frac{r_{AB}(8 - 4\beta + 8\gamma - 3\epsilon)\Phi_{AB}}{4r_B r_A\sin\Phi_{AB}}. \end{aligned}$$

The light-time to second order (65) is therefore

$$\begin{aligned} t_B - t_A &= r_{AB} + mN_1\ln[(r_B + r_A + r_{AB})/(r_B + r_A - r_{AB})] \\ &\quad + m^2\frac{r_{AB}}{r_A r_B}\left(\frac{N_1^2 + 2N_2}{2}\frac{\Phi_{AB}}{\sin\Phi_{AB}} - \frac{N_1^2}{1 + \cos\Phi_{AB}}\right); \end{aligned} \tag{94}$$

$$\begin{aligned} t_B - t_A &= r_{AB} + mN_1\ln\left[\left(r_B + \sqrt{r_B^2 - h_0^2}\right)/\left(r_A + \sqrt{r_A^2 - h_0^2}\right)\right] \\ &\quad + m^2\frac{r_{AB}}{r_A r_B}\left(\frac{N_1^2 + 2N_2}{2}\frac{\Phi_{AB}}{\sin\Phi_{AB}} - \frac{N_1^2}{1 + \cos\Phi_{AB}}\right), \end{aligned} \tag{95}$$

in the superior and inferior cases, respectively. Remarkably, Δ_2 has the same expression. This agrees with the result obtained in [29].

With the same technique, using (68), we have also computed the reduced action at the third order. For good measure, here is the result:

$$S_3(h) = \frac{1}{6h^2 r_B\left(r_B^2 - h^2\right)^{3/2}}\left(2\left(N_1^3 + 6N_1N_2 + 3N_3\right)r_B^4 - 3h^2\left(N_1^3 + 6N_1N_2 + 4N_3\right)r_B^2 + 6h^4(N_1N_2 + N_3)\right) \pm \frac{1}{6h^2 r_A\left(r_A^2 - h^2\right)^{3/2}}\left(2\left(N_1^3 + 6N_1N_2 + 3N_3\right)r_A^4 - 3h^2\left(N_1^3 + 6N_1N_2 + 4N_3\right)r_A^2 + 6h^4(N_1N_2 + N_3)\right). \quad (96)$$

8. Enhancement

Enhancement occurs in the superior case when $r_A \sim r_B \sim R$ are both much larger than $b_0 = h_0$, so that (76) reduces to

$$h_1\left(\frac{1}{r_A} + \frac{1}{r_B}\right) = h_1\frac{2}{R} = \frac{2N_1}{h_0}. \quad (97)$$

As hinted in section 1 for the Newtonian case, it is appropriate to formally introduce another infinitesimal parameter h_0/R, where R is the harmonic mean of the distances (10). When the ratio r_A/r_B is $O(R^0)$, as we assume, the nth-order harmonic average $1\big/r_A^n + 1\big/r_B^n$ is $O(1/R^n)$. The intermediate case $b_0 \approx r_A \ll r_B$, not discussed here, also shows enhancement. For instance, it occurs in a nearly parabolic orbit with a small perihelion distance $p_\odot$, as in the case of a solar probe, for which even $p_\odot = 4\,R_\odot$ has been envisaged. The expansion of

$$h_1 = \frac{RN_1}{h_0} + O(h_0/R) \quad (98)$$

has only odd terms. One should also note that, as can be seen from figure 1, the angle Φ_{AB} is fixed by the Euclidean experimental setup and should be considered independent of m. In the approximation $h_0 \ll R$,

$$\Phi_{AB} = \pi - 2h_0/R + O(h_0/R)^3$$

is slightly less than π; the law of cosines has been used here.

We now proceed to discuss enhancement at the second and third order. It is convenient to first review the behaviour of the function $G(r, h)$ (87) in the limit $r/h = O(R) \gg 1$. Replacing ρ with its expression (38), using (87), and expanding, one gets

$$G_0(r,h) = hJ_0(r/h), \qquad G_1(r,h) = N_1[J_0'(r/h) + J_1(r/h)],$$
$$G_2(r,h) = \frac{N_2}{r}J_0'(r/h) + \frac{N_1^2}{2h}[J_0''(r/h) + 2J_1'(r/h)] + \frac{C_2}{h}J_2(r/h).$$

Now, when $u \gg 1$

$$J_0(u) = -\frac{\pi}{2} + u + \frac{1}{2u} + \frac{1}{24u^3} + \cdots,$$
$$J_1(u) = -1 + \ln(2u) + \cdots, \qquad J_2(u) = \frac{\pi}{4} - \frac{1}{u} + \cdots;$$

setting $u = r/h$,

$$G_0(r,h) = r - \frac{\pi h}{2} + \frac{h^2}{2r} + \frac{h^4}{24r^3} + \cdots,$$
$$G_1(r,h) = N_1\left[-\frac{h^2}{4r^2} + \ln\frac{2r}{h} + \cdots\right],$$
$$G_2(r,h) = N_1^2\left(\frac{\pi}{4h} + \frac{h^2}{6r^3}\right) + N_2\left(\frac{\pi}{2h} - \frac{h^2}{6r^3} - \frac{1}{r}\right) + \cdots.$$

We need

$$G_{0,hh}(r,h) = \frac{1}{r} + \frac{h^2}{2r^3} \to \frac{1}{r}, \qquad G_{0,hhh}(r,h) \to \frac{h}{r^3},$$
$$G_{1,h}(r,h) = -N_1\left(\frac{1}{h} + \frac{h}{2r^2}\right) \to -\frac{N_1}{h},$$
$$G_{1,hh}(r,h) = N_1\left(-\frac{1}{2r^2} + \frac{1}{h^2}\right) \to \frac{N_1}{h^2},$$
$$G_{2,h}(r,h) = N_1^2\left(-\frac{\pi}{4h^2} + \frac{h}{3r^3}\right) + N_2\left(-\frac{\pi}{2h^2} - \frac{h}{3r^3}\right)$$
$$\to -\left(N_1^2 + 2N_2\right)\frac{\pi}{4h^2}.$$

In the expression (67) of Δ_2 the last term is constructed with $G_2(r,h)$ and is not enhanced. The second term comes from $G_{1,h}(r,h_0) = -N_1/h_0$ and, when summed over the end points, contributes to the light-time with $-2N_1h_1/h_0 = -2N_1^2R\big/h_0$. Lastly, the first term gives $h_0h_1^2\big/R = N_1^2R\big/h_0$. Therefore, the enhanced part of the second-order contribution to the light-time is

$$\Delta_{2\,\mathrm{enh}} = -N_1^2R\big/h_0 + O(R^0), \tag{99}$$

in agreement with (9). *The second-order terms in the ODP are just the enhanced ones.*

In a similar way, we get the enhanced third-order terms. For this we need the enhanced part of h_2, to be extracted from (64); its terms are constructed, respectively, with $G_{0,hh}$, $G_{0,hhh}$, $G_{1,hh}$ and $G_{2,h}$. Using their asymptotic expressions above one gets the relation

$$\frac{2}{R}h_2 + \frac{h_1^2}{2}\left(\frac{h_0}{r_A^3} + \frac{h_0}{r_B^3}\right) + \frac{2h_1N_1}{h_0^2} - \left(N_1^2 + 2N_2\right)\left(\frac{2\pi}{4h_0^2}\right) = 0.$$

The third term prevails, and

$$h_2 = -h_1N_1R\big/h_0^2 = -N_1^2R^2\big/h_0^3 + O\left(R\big/h_0^2\right), \tag{100}$$

in agreement with the Newtonian case (which corresponds to $N_1 = 1$).

In the expression (68) for Δ_3,

$$\Delta_3/h_0^2 = \frac{2h_1h_2}{R} + \frac{h_1^3}{6}\left(\frac{h_0}{r_A^3} + \frac{h_0}{r_B^3}\right) - \frac{2h_2N_1}{h_0} + \frac{h_1^2N_1}{h_0}$$
$$- \frac{\pi h_1}{2h_0^2}\left(N_1^2 + 2N_2\right) + \frac{1}{3h_0^2}\left(2N_1^3 + 6N_1N_2 + 3N_3\right); \tag{101}$$

the first, third and fourth terms are enhanced so that finally

$$\Delta_{3\,\mathrm{enh}} = N_1^3R^2\big/h_0^2 + O(R/h_0). \tag{102}$$

Similarly, it turns out that $\Delta_{4\,\mathrm{enh}} \propto N_1^4R^3\big/h_0^3 + O(R/h_0)^2$.

To summarize, expansion (12) reads (for the Sun)

$$\frac{\Delta t}{m} = \Delta_1 + 2 \times 10^{-6} \frac{R_\odot}{h_0} \Delta_2 + 4 \times 10^{-12} \left(\frac{R_\odot}{h_0}\right)^2 \Delta_3 + \cdots. \tag{103}$$

In the superior case, when $R \gg h_0$, Δ_s is a descending power of R/h_0, beginning with $(R/h_0)^{s-1}$. This is the main enhanced term. It depends only on the single PPN parameter N_1: one could say, enhancement arises due to the long-range component $\propto 1/r$ of the index of refraction. Δ_1, typically $\approx 10\, N_1$, is the (logarithmically enhanced) term of (3);

$$\Delta_2 = -N_1^2 \left(\frac{R}{h_0} + O(1)\right), \qquad \Delta_3 = N_1^3 \left[\left(\frac{R}{h_0}\right)^2 + O\left(\frac{R}{h_0}\right)\right]$$

single out the main enhanced contribution. For a given R, the strongest possible enhancement occurs when $h_0 = R_\odot$; numerically

$$\frac{\Delta t}{m} = 10 N_1 - 2 \times 10^{-6} \frac{R}{R_\odot} N_1^2 + 4 \times 10^{-12} N_1^3 \left(\frac{R}{R_\odot}\right)^2 + \cdots. \tag{104}$$

In a typical configuration, with one station on the Earth, $r_A = 1AU \ll r_B$, so that $R = 2\,AU = 400\,R\odot$. The three terms in the expression above are about 20, 3.2×10^{-3}, and 6.4×10^{-7}. For a given accuracy in N_1 (or Δ_1), this shows how many terms are needed in the expansion in this extreme case.

9. Deflection

In the standard theory, the deflection of the image of a faraway source is the acute angle δ between the asymptotes of the ray. Taking the origin of longitudes on the symmetry axis OK through the closest approach (figure 4) and using (53), the longitude of the outgoing asymptote reads (with $\rho = uh$)

$$\phi_\infty = \frac{\pi + \delta}{2} = h \int_b^\infty \frac{\mathrm{d}r}{r\sqrt{\rho^2 - h^2}} = \int_1^\infty \mathrm{d}u \frac{\mathrm{d}\ln r(u)}{\mathrm{d}u} \frac{1}{\sqrt{u^2 - 1}}. \tag{105}$$

Expanding in powers of m/h, using (86) and separating out the main part, we obtain

$$\phi_\infty = \sum_{s=0} C_s I_s \left(\frac{m}{h}\right)^s = \frac{\pi}{2} + \frac{m}{h} \sum_{s=1} C_s I_s \left(\frac{m}{h}\right)^s, \tag{106}$$

where

$$I_s = \int_1^\infty \frac{\mathrm{d}u}{u^{s+1}\sqrt{u^2 - 1}}$$

are the numerical constants and $\mathrm{d}(\log(r(u))/\mathrm{d}u$ has been defined in equation (85). The total deflection is explicitly

$$\delta = 2N_1 \frac{m}{h} + \pi \frac{N_1^2 + 2N_2}{2} \left(\frac{m}{h}\right)^2 + \frac{4\left(N_1^3 + 6N_1N_2 + 3N_3\right)}{3} \left(\frac{m}{h}\right)^3 + \cdots.$$

In the more common isotropic gauge (83)

$$h = bN(b) = b + N_1 m + N_2 \frac{m^2}{b} + N_3 \frac{m^3}{b^2} + \cdots,$$

and so

$$\delta = 2N_1 \frac{m}{b} + \frac{\pi(N_1^2 + 2N_2) - 4N_1^2}{2} \left(\frac{m}{b}\right)^2 + \\ + \frac{10N_1^3 + 18N_1N_2 + 12N_3 - 3\pi N_1^3 - 6\pi N_1 N_2}{3} \left(\frac{m}{b}\right)^3 + \cdots. \tag{107}$$

In terms of the PPN coefficients and using the expansion of h, to second order we have

$$\delta = \frac{2m(\gamma+1)}{h} + \frac{\pi m^2}{4}(8 - 4\beta + 3\epsilon + 8\gamma), \tag{108}$$

which agrees with [13] and [26]; in general relativity, and using the closest approach b,

$$\delta = 4\frac{m}{b} + (15\pi - 32)\frac{m^2}{4b^2} + \frac{(155 - 45\pi)m^3}{3b^3}, \tag{109}$$

in agreement to second order with [7].

This standard approach, however, is not adequate for astrometric observations, which are carried out from a point B at a finite distance r_B. In the linear approximation this problem has been solved in [19], section 40.3; here we give a general formulation and derive the quadratic term. Referring to figure 4, we need the unit tangent vector $\mathbf{n}(\phi)$ in the counterclockwise direction (increasing ϕ) at a generic point $(r\cos\phi, r\sin\phi)$ on the ray (for simplicity, on the outgoing branch), expressed in terms of the function $r(\phi)$:

$$\mathbf{n}(\phi) = \frac{(r'\cos\phi - r\sin\phi, r'\sin\phi + r\cos\phi)}{\sqrt{r'^2 + r^2}}. \tag{110}$$

From (60)

$$r'(\phi) = r\frac{\sqrt{\rho^2 - h^2}}{h},$$

so that at B, the tangent vector is

$$\begin{aligned}\mathbf{n}_B &= \frac{1}{\rho_B}\left(\sqrt{\rho_B^2 - h^2}\cos\phi_B - h\sin\phi_B, \sqrt{\rho_B^2 - h^2}\sin\phi_B + h\cos\phi_B\right)\\ &= (n_{Bx}, n_{By}).\end{aligned} \tag{111}$$

With

$$\cos\chi_B = h/\rho_B = 1/u_B, \qquad \sin\chi_B = \sqrt{\rho_B^2 - h^2}\big/\rho_B = \sqrt{1 - 1/u_B^2},$$

it is convenient to introduce the quantity χ_B, a function on the ray; in the limit $m \to 0$, since $\rho \to r$ and $h \to 1$, it reduces to α_B (figure 1). Then

$$\mathbf{n}_B = (\sin(\chi_B - \phi_B), \cos(\chi_B - \phi_B)). \tag{112}$$

The deflection δ_B is provided by the vector product

$$|\mathbf{n}^\star \times \mathbf{n}_B| = \sin\delta_B,$$

where $\mathbf{n}^\star = (\sin(\delta/2), \cos(\delta/2))$ is a unit vector along the asymptote of the incoming ray. Hence we obtain the exact expression

$$\delta_B = \phi_B - \chi_B + \frac{\delta}{2}. \tag{113}$$

Two effects contribute in (113): a local term χ_B due to the change in the tangent, and a change in the orientation of the outgoing asymptote relative to OA. In the case of GAIA and other space astrometric projects, no images can be obtained near the Sun, so that $r_B = 1$ AU $\approx h$ and there is little enhancement. The data analysis will be truly global, with subtle statistics. The expected angular measurement error $\approx 5 \times 10^{-11}$ is quite below the first-order deflection $\approx 4 \times 10^{-8}$ and much larger than the second-order term $\approx 10^{-16}$; however, the fractional difference between δ and δ_B is not small. With our powerful formalism the derivation of the second-order approximation to δ_B is straightforward.

Two limits are noteworthy. When $m \to 0$, ϕ_B tends to α_B and, of course, there is no deflection. To recover the standard expression when B goes to infinity, note that, by use of (85),

$$\phi_B = \int_1^{u_B} \mathrm{d}u \frac{\mathrm{d}\ln r(u)}{\mathrm{d}u} \frac{1}{\sqrt{u^2-1}} = \chi_B + \frac{m}{h}\int_1^{u_B} \mathrm{d}u \frac{q(u)}{\sqrt{u^2-1}};$$

therefore, to second order,

$$\phi_B = \left(1 + \frac{m^2 C_2}{2h^2}\right)\chi_B + \frac{mC_1\sqrt{u_B^2-1}}{hu_B} + \frac{m^2 C_2\sqrt{u_B^2-1}}{2h^2 u_B^2}. \tag{114}$$

Thus, the deflection reads

$$\delta_B = \frac{\delta}{2} + \left(\frac{m^2 C_2}{2h^2}\right)\chi_B + \frac{mC_1\sqrt{u_B^2-1}}{hu_B} + \frac{m^2 C_2\sqrt{u_B^2-1}}{2h^2 u_B^2}. \tag{115}$$

In the limit $u_B \to \infty$, this agrees with equation (107). In terms of r_B and h_0, this is

$$\delta_B = \frac{\delta}{2} + \sqrt{r_B^2 - h_0^2}\frac{m}{h_0 r_B}\left(C_1 + \frac{mC_2}{2r_B}\right) + \frac{m^2 C_2 \chi_B}{2h_0^2} + \frac{m^2 C_1}{\sqrt{r_B^2 - h_0^2}}\left(\frac{N_1 h_0}{r_B^2} - \frac{h_1 r_B}{h_0^2}\right). \tag{116}$$

For u_B finite, this result agrees to first order with [19], equation (40.11).

10. Conclusion

With the implementation of optical lasers in deep space, experimental gravity will undergo a big leap. The planned mission ASTROD ([22, 23] and other papers) will consist of a fleet of three drag-free spacecraft in a triangular configuration with semi-major axes of about 1 AU. Although no detailed error analysis is available, ranging accuracies of 3×10^{-3} cm or better are expected; with closest approach less than 1 AU, this error is comparable with, or smaller than, the second-order gravitational delay.

Optical interferometry in space will make huge improvements in phase measurements possible. The GAME (Gamma Astrometric Measurements Experiment) project (see [15]) consists of a Fizeau interferometer in the focal plane of a space telescope to measure the angular separation of stars in a narrow field of view near the Sun. The expected accuracy in γ of 10^{-7} will require second-order corrections in the gravitational delay.

LISA—a planned mission for low-frequency gravitational wave detection ([12, 14] and many other papers, in particular [20])—will fly three drag-free spacecraft orbiting at 1 AU at the vertices of an equilateral triangle with sides $L = 5 \times 10^{11}$ cm; this fleet will rotate around its centre with the period of a year. Three optical interferometers with baseline L will operate simultaneously, with an expected sensitivity $\sigma_L/L \approx 10^{-21}$ or better. The change in light-time difference between two arms due to the solar gravitational delay has the period of 6 months, in a frequency band overwhelmed by the acceleration noise, but it is interesting to evaluate the effect. For two vertices A and B, $r_B - r_A = \delta r \approx L \ll (r_A, r_B) = 1$ AU. In the (now generic) inferior case the reduced action (51) (with the $-$ sign!) is of order

$$r_{AB} + mN_1 \frac{\delta r}{\sqrt{r_A^2 - h_0^2}} \approx 5 \times 10^{11}\mathrm{cm} + 10^4\ \mathrm{cm}.$$

With the approximation $\delta r \ll 1$ AU the action reads

$$S(h) = h\Phi_{AB} + \frac{\delta r}{r_A}\sqrt{r_A^2 N^2\left(r_A^2\right) - h^2}$$
$$= h\Phi_{AB} + \frac{\delta r}{r_A}\left[\sqrt{r_A^2 - h^2} + m\frac{N_1 r_A}{\sqrt{r_A^2 - h^2}} + \frac{m^2}{2}\left(\frac{N_1^2 + 2N_2}{\sqrt{r_A^2 - h^2}} - \frac{N_1^2 r_A^2}{\left(r_A^2 - h^2\right)^{3/2}}\right)\right], \tag{117}$$

an expression that can be used directly to obtain all relevant quantities. For an estimate, however, it suffices to remark that in the above m-expansion each term is smaller than the previous one by $O(m/r_A) = 10^{-8}$; hence for LISA the first-, second- and third-order corrections to the light-time are, respectively, of order 10^4 cm, 10^{-4} cm and 10^{-12} cm, corresponding to gravitational wave signals of order

$$2 \times 10^{-8}, \qquad 2 \times 10^{-16}, \qquad 2 \times 10^{-24}.$$

We did not investigate the consequences of this large, but low-frequency signal on the performance of the instrument.

The puzzle of the ODP expression for the gravitational delay has been understood. It must be considered in the framework of an expansion in powers of m/b; of all second-order terms so arising in a close conjunction some are enhanced. They can be rigorously singled out with a further expansion in diminishing powers of R/b_0; those that appear in the ODP are just those of order $m(m/b_0)(R/b_0)$. With the powerful tool of geometrical optics, we have provided a procedure to extend the calculation to higher order and have obtained the full correct second-order term of the delay.

A methodological reflection is a fit conclusion. The evaluation of the gravitational delay, a conceptually simple and straightforward problem, faces subtle mathematical difficulties and a great algebraic complexity. Our approach is based upon two unusual mathematical levels of description: light propagation with the eikonal theory, rather than null geodesics, and asymptotic power series, an abstract mathematical tool. The latter, in which ordinary functions are set aside and an abstract mathematical tool is employed, seemingly runs against physical intuition. As shown, both are essential to directly attain, and take advantage of, the crucial features of the problem: the light-time as the minimum of Fermat's action, and a safe and automatic procedure to select and estimate different terms. This is another example of the tenet that *every physical problem has an appropriate, often not intuitive, level of mathematical description*, and severe penalties are in store for its neglect.

Appendix

The radial gauge freedom and the difference between the closest approach and b_0 can cause some confusion. For example, the textbook [30] presents (equation (8.7.4)) the light-time between the closest approach and a generic point; it is expressed in Schwarzschild's gauge $\bar{r}$ and reads

$$t(\bar{r}, \bar{b}) = \sqrt{\bar{r}^2 - \bar{b}^2} + (1+\gamma)m \ln\frac{\bar{r} + \sqrt{\bar{r}^2 - \bar{b}^2}}{\bar{b}} + m\sqrt{\frac{\bar{r} - \bar{b}}{\bar{r} + \bar{b}}},$$

quite different from (2). In a real case two such terms are needed, one for each branch. But, contrary to what stated in the textbook, the sum of the two square roots (first term) *is not the distance AB*. The isotropic gauge and the distance b_0, not the closest approach, should be used. First, if one sets (37) $\bar{r} = r + \gamma m$, the formula reads, to $O(m)$,

$$t(r, b) = \sqrt{r^2 - b^2} + (1+\gamma)m\left(\ln\frac{r + \sqrt{r^2 - b^2}}{b} + \sqrt{\frac{r - b}{r + b}}\right).$$

Both formulae are useless, however, because the closest approach $b = b_0 + mb_1$ is not known beforehand. The ray must be anchored to two known points and, with b_1, is determined by the unknown γ with (81). Because

$$\sqrt{r^2 - b^2} = \sqrt{r^2 - b_0^2} - m\frac{b_1}{\sqrt{r^2 - b_0^2}},$$

$$\sqrt{r_A^2 - b^2} + \sqrt{r_B^2 - b^2} = r_{AB} - mb_1\left(\frac{1}{\sqrt{r_A^2 - b_0^2}} + \frac{1}{\sqrt{r_B^2 - b_0^2}}\right)$$

$$= r_{AB} - m(1+\gamma)\left(\sqrt{\frac{r_A - b_0}{r_A + b_0}} + \sqrt{\frac{r_B - b_0}{r_B + b_0}}\right),$$

and the standard formula is recovered.

References

[1] Bel Ll and Martin J 1994 Fermat's principle in general relativity *Gen. Rel. Grav.* **26** 567–85
[2] Bertotti B, Ashby N and Iess L 2008 The effect of the motion of the Sun on the light-time in interplanetary relativity experiments *Class. Quantum Grav.* **25** 045013
[3] Bertotti B, Comoretto G and Iess L 1993 Doppler tracking of spacecraft with multi-frequency links *Astron. Astrophys.* **269** 608–16
[4] Bertotti B and Giampieri G 1992 Relativistic effects for Doppler measurements near solar conjunction *Class. Quantum Grav.* **9** 777
[5] Bertotti B, Farinella P and Vokrouhlický D 2003 *Physics of the Solar System* (Dordrecht: Kluwer)
[6] Bertotti B, Iess L and Tortora P 2003 A test of general relativity using radio links with the Cassini spacecraft *Nature* **425** 374–6
[7] Bodenner J and Will C M 2003 Deflection of light to second order: a tool for illustrating principles of general relativity *Am. J. Phys.* **71** 770
[8] Born M and Wolf E 1964 *Principles of Optics* (Oxford: Pergamon)
[9] Ciufolini I and Wheeler J A 1995 *Gravitation and Inertia* (Princeton, NJ: Princeton University Press)
[10] Epstein R and Shapiro I I 1980 Post-post Newtonian deflection of light by the Sun *Phys. Rev.* D **22** 2947–9
[11] Erdély A 1956 *Asymptotic Expansions* (New York: Dover)
[12] European Space Agency LISA 2000 Laser interferometer space antenna: a comerstone mission for the observation of gravitational waves ESA-SCI(2000) **11** 45–56
[13] Fischbach E and Freeman B S 1980 Second-order contribution to the gravitational deflection of light. *Phys. Rev.* D **22** 2950–2
[14] Folkner W M (ed) 1998 Laser Interferometer Space Antenna *2nd International LISA Symposium AIP Conf. Proc.* **456** 11–16
[15] Gai M, Lattanzi M G, Ligori S and Vecchiato A 2008 GAME: Gamma Astrometric Measurement Experiment *Proc. SPIE* **7010** 701027
[16] Hinch E J 1991 *Perturbation Methods* (Cambridge: Cambridge University Press)
[17] Kopeikin S M, Polnarev A G, Schäfer G and Vlasov I Yu 2007 Gravimagnetic effect of the barycentric motion of the Sun and determination of the post-Newtonian parameter γ in the Cassini experiment *Phys. Lett.* A **367** 276–80
[18] Le Poncin-Lafitte C, Linet B and Tyssandier P 2004 World function and time transfer: general post-Minkowskian expansions *Class. Quantum Grav.* **21** 4463–83
[19] Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (San Francisco: Freeman)
[20] Moore A T and Hellings R W 2001 Angular resolution of space-based gravitational wave detectors *Phys. Rev.* D **65** 062001
[21] Moyer T D 2003 *Formulation for Observed and Computed Values of Deep Space Network Data Types for Navigation* (Hoboken, NJ: Wiley) chapter 8
[22] Ni W-T *et al* 2006 ASTROD I: mission concept and Venus flyby *Acta Astronaut.* **59** 598–607
[23] Ni W-T 2007 ASTROD (Astrodynamical Space Test of Relativity using Optical Devices) and ASTROD I *Nucl. Phys. Proc. Suppl.* **166** 153–8

[24] Reasenberg R D *et al* 1979 Viking relativity experiment: verification of signal retardation by solar gravity *Astrophys. J.* **234** L219–21

[25] Richter G W and Matzner R A 1982 Second-order contributions to the gravitational deflection of light in the parametrized post-Newtonian formalism *Phys. Rev.* D **26** 1219–24

[26] Richter G W and Matzner R A 1984 Second-order contributions to relativistic time delay in the parametrized post-Newtonian formalism *Phys. Rev.* D **28** 3007–12

[27] Shapiro I I 1964 Fourth test of general relativity *Phys. Rev. Lett.* **13** 789–91

[28] Soffel M H *et al* 2003 The IAU 2000 resolutions for astrometry, celestial mechanics, and metrology in the relativistic framework: explanatory supplement *Astr. J.* **126** 2687–706

[29] Teyssandier P and Le Poncin-Lafitte C 2008 General post-Minkowskian expansion of time transfer functions *Class. Quantum Grav.* **25** 145020

[30] Weinberg S 1972 *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (New York: Wiley)

[31] Will C M 1993 *Theory and Experiment in Gravitational Physics* (Cambridge: Cambridge University Press)