Link http://geographiclib.sourceforge.net/geod.html
E-print arXiv:1109.4448

arXiv:1109.4448v2 [physics.geo-ph] 28 Mar 2012

Algorithms for geodesics

Charles F. F. Karney*

SRI International, 201 Washington Rd, Princeton, NJ 08543-5300, USA

(Dated: September 21, 2011; revised March 28, 2012)

Algorithms for the computation of geodesics on an ellipsoid of revolution are given. These provide accurate, robust, and fast solutions to the direct and inverse geodesic problems and they allow differential and integral properties of geodesics to be computed.

Keywords: geometrical geodesy, geodesics, polygonal areas, gnomonic projection, numerical methods

1. INTRODUCTION

The shortest path between two points on the earth, customarily treated as an ellipsoid of revolution, is called a *geodesic*. Two geodesic problems are usually considered: the *direct* problem of finding the end point of a geodesic given its starting point, initial azimuth, and length; and the *inverse* problem of finding the shortest path between two given points. Referring to Fig. 1, it can be seen that each problem is equivalent to solving the geodesic triangle NAB given two sides and their included angle (the azimuth at the first point, α_1, in the case of the direct problem and the longitude difference, λ_{12}, in the case of the inverse problem). The framework for solving these problems was laid down by Legendre (1806), Oriani (1806, 1808, 1810), Bessel (1825), and Helmert (1880). Based on these works, Vincenty (1975a) devised algorithms for solving the geodesic problems suitable for early programmable desk calculators; these algorithms are in widespread use today. A good summary of Vincenty's algorithms and the earlier work in the field is given by Rapp (1993, Chap. 1).

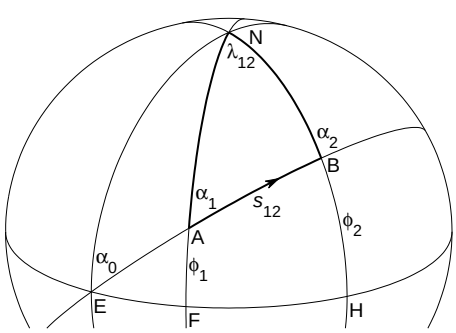

FIG. 1 The ellipsoidal triangle NAB. N is the north pole, NAF and NBH are meridians, and AB is a geodesic of length s_{12}. The longitude of B relative to A is λ_{12}; the latitudes of A and B are ϕ_1 and ϕ_2. EFH is the equator with E also lying on the extension of the geodesic AB; and α_0, α_1, and α_2 are the azimuths (in the forward direction) of the geodesic at E, A, and B.

The goal of this paper is to adapt the geodesic methods of Helmert (1880) and his predecessors to modern computers. The current work goes beyond Vincenty in three ways: (1) The accuracy is increased to match the standard precision of most computers. This is a relatively straightforward task of retaining sufficient terms in the series expansions and can be achieved at little computational cost. (2) A solution of the inverse problem is given which converges for all pairs of points. (Vincenty's method fails to converge for nearly antipodal points.) (3) Differential and integral properties of the geodesics are computed. The differential properties allow the behavior of nearby geodesics to be determined, which enables the scales of geodesic projections to be computed without resorting to numerical differentiation; crucially, one of the differential quantities is also used in the solution of the inverse problem. The integral properties provide a method for finding the area of a geodesic polygon, extending the work of Danielsen (1989).

Section 2 reviews the classical solution of geodesic problem by means of the auxiliary sphere and provides expansions of the resulting integrals accurate to $O(f^6)$ (where f is the flattening of the ellipsoid). These expansions can be inserted into the solution for the direct geodesic problem presented by, for example, Rapp (1993) to provide accuracy to machine precision. Section 3 gives the differential properties of geodesics reviewing the results of Helmert (1880) for the reduced length and geodesic scale and give the key properties of these quantities and appropriate series expansions to allow them to be calculated accurately. Knowledge of the reduced length enables the solution of the inverse problem by Newton's method which is described in Sect. 4. Newton's method requires a good starting guess and, in the case of nearly antipodal points, this is provided by an approximate solution of the inverse problem by Helmert (1880), as given in Sect. 5. The computation of area between a geodesic and the equator is formulated in Sect. 6, extending the work of Danielsen (1989). Some details of the implementation and present accuracy and timing data are discussed in Sect. 7. As an illustration of the

*Electronic address: charles.karney@sri.com

use of these algorithms, Sect. 8 gives an ellipsoidal gnomonic projection in which geodesics are very nearly straight. This provides a convenient way of solving several geodesic problems.

For the purposes of this paper, it is useful to generalize the definition of a geodesic. The geodesic curvature, κ, of an arbitrary curve at a point P on a surface is defined as the curvature of the projection of the curve onto a plane tangent to the surface at P. All shortest paths on a surface are *straight*, defined as $\kappa = 0$ at every point on the path. In the rest of this paper, I use straightness as the defining property of geodesics; this allows geodesic lines to be extended indefinitely (beyond the point at which they cease to be shortest paths).

Several of the results reported here appeared earlier in a technical report, Karney (2011).

2. BASIC EQUATIONS AND DIRECT PROBLEM

I consider an ellipsoid of revolution with equatorial radius a, and polar semi-axis b, flattening f, third flattening n, eccentricity e, and second eccentricity e' given by

$$f = (a-b)/a = 1 - \sqrt{1-e^2}, \tag{1}$$
$$n = (a-b)/(a+b) = f/(2-f), \tag{2}$$
$$e^2 = (a^2-b^2)/a^2 = f(2-f), \tag{3}$$
$$e'^2 = (a^2-b^2)/b^2 = e^2/(1-e^2). \tag{4}$$

As a consequence of the rotational symmetry of the ellipsoid, geodesics obey a relation found by Clairaut (1735), namely

$$\sin\alpha_0 = \sin\alpha_1\cos\beta_1 = \sin\alpha_2\cos\beta_2, \tag{5}$$

where β is the reduced latitude (sometimes called the parametric latitude), given by

$$\tan\beta = (1-f)\tan\phi. \tag{6}$$

The geodesic problems are most easily solved by using an *auxiliary sphere* which allows an exact correspondence to be made between a geodesic and a great circle on a sphere. On the sphere, the latitude ϕ is replaced by the reduced latitude β, and azimuths α are preserved. From Fig. 2, it is clear that Clairaut's equation, $\sin\alpha_0 = \sin\alpha\cos\beta$, is just the sine rule applied to the sides NE and NP of the triangle NEP and their opposite angles. The third side, the spherical arc length σ, and its opposite angle, the spherical longitude ω, are related to the equivalent quantities on the ellipsoid, the distance s and longitude λ, by (Rapp, 1993, Eqs. (1.28) and (1.170))

$$\frac{s}{b} = \int_0^\sigma \sqrt{1+k^2\sin^2\sigma'}\,\mathrm{d}\sigma' = I_1(\sigma), \tag{7}$$
$$\lambda = \omega - f\sin\alpha_0\int_0^\sigma \frac{2-f}{1+(1-f)\sqrt{1+k^2\sin^2\sigma'}}\,\mathrm{d}\sigma'$$
$$= \omega - f\sin\alpha_0\, I_3(\sigma), \tag{8}$$

where

$$k = e'\cos\alpha_0. \tag{9}$$

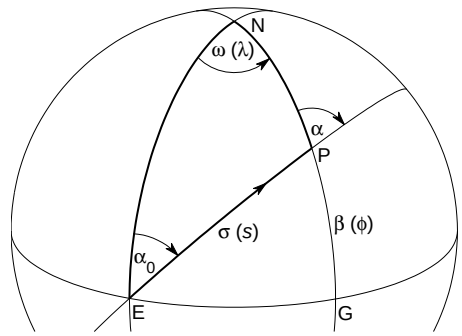

FIG. 2 The elementary ellipsoidal triangle NEP mapped to the auxiliary sphere. NE and NPG are meridians; EG is the equator; and EP is the great circle (i.e., the geodesic). The corresponding ellipsoidal variables are shown in parentheses. Here P represents an arbitrary point on the geodesic EAB in Fig. 1.

See also Eqs. (5.4.9) and (5.8.8) of Helmert (1880). The origin for s, σ, λ, and ω is the point E, at which the geodesic crosses the equator in the northward direction, with azimuth α_0. The point P can stand for either end of the geodesic AB in Fig. 1, with the quantities β, α, σ, ω, s, and λ acquiring a subscript 1 or 2. I also define $s_{12} = s_2 - s_1$ as the length of AB, with λ_{12}, σ_{12}, and ω_{12} defined similarly. (In this paper, α_2 is the forward azimuth at B. Several authors use the back azimuth instead; this is given by $\alpha_2 \pm \pi$.)

Because Eqs. (7) and (8) depend on α_0, the mapping between the ellipsoid and the auxiliary sphere is not a global mapping of one surface to another; rather the auxiliary sphere should merely be regarded as a useful mathematical technique for solving geodesic problems. Similarly, because the origin for λ depends on the geodesic, only longitude differences, e.g., λ_{12}, should be used in converting between longitudes relative to the prime meridian and λ.

In solving the spherical trigonometrical problems, the following equations relating the sides and angles of NEP are useful,

$$\alpha_0 = \mathrm{ph}(|\cos\alpha + i\sin\alpha\sin\beta| + i\sin\alpha\cos\beta), \tag{10}$$
$$\sigma = \mathrm{ph}(\cos\alpha\cos\beta + i\sin\beta), \tag{11}$$
$$\omega = \mathrm{ph}(\cos\sigma + i\sin\alpha_0\sin\sigma), \tag{12}$$
$$\beta = \mathrm{ph}(|\cos\alpha_0\cos\sigma + i\sin\alpha_0| + i\cos\alpha_0\sin\sigma), \tag{13}$$
$$\alpha = \mathrm{ph}(\cos\alpha_0\cos\sigma + i\sin\alpha_0), \tag{14}$$

where $i = \sqrt{-1}$ and $\mathrm{ph}(x+iy)$ is the phase of a complex number (Olver *et al.*, 2010, §1.9(i)), typically given by the library function $\mathrm{atan2}(y,x)$. Equation (10) merely recasts Eq. (5) in a form that allows it to be evaluated accurately when α_0 is close to $\frac{1}{2}\pi$. The other relations are obtained by applying Napier's rules of circular parts to NEP.

The distance integral, Eq. (7), can be expanded in a Fourier series

$$I_1(\sigma) = A_1\Big(\sigma + \sum_{l=1}^{\infty} C_{1l}\sin 2l\sigma\Big), \tag{15}$$

with the coefficients determined by expanding the integral in a Taylor series. It is advantageous to follow Bessel (1825, §5) and Helmert (1880, Eq. (5.5.1)) and use ϵ, defined by

$$\epsilon = \frac{\sqrt{1+k^2}-1}{\sqrt{1+k^2}+1} \quad \text{or} \quad k = \frac{2\sqrt{\epsilon}}{1-\epsilon}, \tag{16}$$

as the expansion parameter. This leads to expansions with half as many terms as the corresponding ones in k^2. The expansion can be conveniently carried out to arbitrary order by a computer algebra system such as Maxima (2009) which yields

$$\begin{aligned}
A_1 &= (1-\epsilon)^{-1}\big(1+\tfrac{1}{4}\epsilon^2+\tfrac{1}{64}\epsilon^4+\tfrac{1}{256}\epsilon^6+\cdots\big), && (17)\\
C_{11} &= -\tfrac{1}{2}\epsilon+\tfrac{3}{16}\epsilon^3-\tfrac{1}{32}\epsilon^5+\cdots,\\
C_{12} &= -\tfrac{1}{16}\epsilon^2+\tfrac{1}{32}\epsilon^4-\tfrac{9}{2048}\epsilon^6+\cdots,\\
C_{13} &= -\tfrac{1}{48}\epsilon^3+\tfrac{3}{256}\epsilon^5+\cdots,\\
C_{14} &= -\tfrac{5}{512}\epsilon^4+\tfrac{3}{512}\epsilon^6+\cdots,\\
C_{15} &= -\tfrac{7}{1280}\epsilon^5+\cdots,\\
C_{16} &= -\tfrac{7}{2048}\epsilon^6+\cdots. && (18)
\end{aligned}$$

This extends Eq. (5.5.7) of Helmert (1880) to higher order. These coefficients may be inserted into Eq. (1.40) of Rapp (1993) using

$$\mathcal{B}_j = \begin{cases} A_1, & \text{for } j=0,\\ 2A_1C_{1l}, & \text{for } j=2l, \text{ with } l>0, \end{cases} \tag{19}$$

where here, and subsequently in Eqs. (22) and (26), a script letter, e.g., $\mathcal{B}$, is used to stand for Rapp's coefficients.

In the course of solving the direct geodesic problem (where s_{12} is given), it is necessary to determine σ given s. Vincenty solves for σ iteratively. However, it is simpler to follow Helmert (1880, §5.6) and substitute $s = bA_1\tau$ into Eqs. (7) and (15), to obtain $\tau = \sigma + \sum_l C_{1l}\sin 2l\sigma$; this may be inverted, for example, using Lagrange reversion, to give

$$\sigma = \tau + \sum_{l=1}^{\infty} C'_{1l}\sin 2l\tau, \tag{20}$$

where

$$\begin{aligned}
C'_{11} &= \tfrac{1}{2}\epsilon-\tfrac{9}{32}\epsilon^3+\tfrac{205}{1536}\epsilon^5+\cdots,\\
C'_{12} &= \tfrac{5}{16}\epsilon^2-\tfrac{37}{96}\epsilon^4+\tfrac{1335}{4096}\epsilon^6+\cdots,\\
C'_{13} &= \tfrac{29}{96}\epsilon^3-\tfrac{75}{128}\epsilon^5+\cdots,\\
C'_{14} &= \tfrac{539}{1536}\epsilon^4-\tfrac{2391}{2560}\epsilon^6+\cdots,\\
C'_{15} &= \tfrac{3467}{7680}\epsilon^5+\cdots,\\
C'_{16} &= \tfrac{38081}{61440}\epsilon^6+\cdots. && (21)
\end{aligned}$$

This extends Eq. (5.6.8) of Helmert (1880) to higher order. These coefficients may be used in Eq. (1.142) of Rapp (1993) using

$$\mathcal{D}_j = 2C'_{1l}, \quad \text{for } j=2l, \text{ with } l>0. \tag{22}$$

TABLE 1 The parameters for the WGS84 ellipsoid used in the examples. The column labeled "Eq." lists the equations used to compute the corresponding quantities.

Qty.	Value	Eq.
a	6 378 137 m	given
f	1/298.257 223 563	given
b	6 356 752.314 245 m	(1)
c	6 371 007.180 918 m	(60)
n	0.001 679 220 386 383 70	(2)
e^2	0.006 694 379 990 141 32	(3)
e'^2	0.006 739 496 742 276 43	(4)

Similarly, the integral appearing in the longitude equation, Eq. (8), can be written as a Fourier series

$$I_3(\sigma) = A_3\Big(\sigma + \sum_{l=1}^{\infty} C_{3l}\sin 2l\sigma\Big). \tag{23}$$

Following Helmert (1880), I expand jointly in n and ϵ, both of which are $O(f)$, to give

$$\begin{aligned}
A_3 &= 1-\big(\tfrac{1}{2}-\tfrac{1}{2}n\big)\epsilon-\big(\tfrac{1}{4}+\tfrac{1}{8}n-\tfrac{3}{8}n^2\big)\epsilon^2\\
&\quad-\big(\tfrac{1}{16}+\tfrac{3}{16}n+\tfrac{1}{16}n^2\big)\epsilon^3-\big(\tfrac{3}{64}+\tfrac{1}{32}n\big)\epsilon^4\\
&\quad-\tfrac{3}{128}\epsilon^5+\cdots, && (24)\\
C_{31} &= \big(\tfrac{1}{4}-\tfrac{1}{4}n\big)\epsilon+\big(\tfrac{1}{8}-\tfrac{1}{8}n^2\big)\epsilon^2+\big(\tfrac{3}{64}+\tfrac{3}{64}n-\tfrac{1}{64}n^2\big)\epsilon^3\\
&\quad+\big(\tfrac{5}{128}+\tfrac{1}{64}n\big)\epsilon^4+\tfrac{3}{128}\epsilon^5+\cdots,\\
C_{32} &= \big(\tfrac{1}{16}-\tfrac{3}{32}n+\tfrac{1}{32}n^2\big)\epsilon^2+\big(\tfrac{3}{64}-\tfrac{1}{32}n-\tfrac{3}{64}n^2\big)\epsilon^3\\
&\quad+\big(\tfrac{3}{128}+\tfrac{1}{128}n\big)\epsilon^4+\tfrac{5}{256}\epsilon^5+\cdots,\\
C_{33} &= \big(\tfrac{5}{192}-\tfrac{3}{64}n+\tfrac{5}{192}n^2\big)\epsilon^3+\big(\tfrac{3}{128}-\tfrac{5}{192}n\big)\epsilon^4\\
&\quad+\tfrac{7}{512}\epsilon^5+\cdots,\\
C_{34} &= \big(\tfrac{7}{512}-\tfrac{7}{256}n\big)\epsilon^4+\tfrac{7}{512}\epsilon^5+\cdots,\\
C_{35} &= \tfrac{21}{2560}\epsilon^5+\cdots. && (25)
\end{aligned}$$

This extends Eq. (5.8.14) of Helmert (1880) to higher order. These coefficients may be inserted into Eq. (1.56) of Rapp (1993) using

$$\mathcal{A}_j = \begin{cases} A_3, & \text{for } j=0,\\ 2A_3C_{3l}, & \text{for } j=2l, \text{ with } l>0. \end{cases} \tag{26}$$

The equations given in this section allow the direct geodesic problem to be solved. Given ϕ_1 (and hence β_1) and α_1 solve the spherical triangle NEA to give α_0, σ_1, and ω_1 using Eqs. (10), (11), and (12). Find s_1 and λ_1 from Eqs. (7) and (8) together with Eqs. (15) and (23). (Recall that the origin for λ is E in Fig. 1.) Determine $s_2 = s_1 + s_{12}$ and hence σ_2 using Eq. (20). Now solve the spherical triangle NEB to give α_2, β_2 (and hence ϕ_2), and ω_2, using Eqs. (14), (13), and (12). Finally, determine λ_2 (and λ_{12}) from Eqs. (8) and (23). A numerical example of the solution of the direct problem is given in Table 2 using the parameters of Table 1.

TABLE 2 A sample direct calculation specified by $\phi_1 = 40°$, $\alpha_1 = 30°$, and $s_{12} = 10\,000\,\mathrm{km}$. For equatorial geodesics ($\phi_1 = 0$ and $\alpha_1 = \frac{1}{2}\pi$), Eq. (11) is indeterminate; in this case, take $\sigma_1 = 0$.

Qty.	Value	Eq.
ϕ_1	$40°$	given
α_1	$30°$	given
s_{12}	$10\,000\,000\,\mathrm{m}$	given
Solve triangle NEA		
β_1	$39.905\,277\,146\,01°$	(6)
α_0	$22.553\,940\,202\,62°$	(10)
σ_1	$43.999\,153\,645\,00°$	(11)
ω_1	$20.323\,718\,278\,37°$	(12)
Determine σ_2		
k^2	$0.005\,748\,029\,628\,57$	(9)
ϵ	$0.001\,432\,892\,204\,16$	(16)
A_1	$1.001\,435\,462\,362\,07$	(17)
$I_1(\sigma_1)$	$0.768\,315\,388\,864\,12$	(15)
s_1	$4\,883\,990.626\,232\,\mathrm{m}$	(7)
s_2	$14\,883\,990.626\,232\,\mathrm{m}$	$s_1 + s_{12}$
τ_2	$133.962\,660\,502\,08°$	$s_2/(bA_1)$
σ_2	$133.921\,640\,830\,38°$	(20)
Solve triangle NEB		
α_2	$149.090\,169\,318\,07°$	(14)
β_2	$41.697\,718\,092\,50°$	(13)
ω_2	$158.284\,121\,471\,12°$	(12)
Determine λ_{12}		
A_3	$0.999\,284\,243\,06$	(24)
$I_3(\sigma_1)$	$0.767\,737\,860\,69$	(23)
$I_3(\sigma_2)$	$2.335\,343\,221\,70$	(23)
λ_1	$20.267\,150\,380\,16°$	(8)
λ_2	$158.112\,050\,423\,93°$	(8)
λ_{12}	$137.844\,900\,043\,77°$	$\lambda_2 - \lambda_1$
Solution		
ϕ_2	$41.793\,310\,205\,06°$	(6)
λ_{12}	$137.844\,900\,043\,77°$	
α_2	$149.090\,169\,318\,07°$	

3. DIFFERENTIAL QUANTITIES

Before turning to the inverse problem, I present Gauss' solution for the differential behavior of geodesics. One differential quantity, the reduced length m_{12}, is needed in the solution of the inverse problem by Newton's method (Sect. 4) and an expression for this quantity is given at the end of this section. However, because this and other differential quantities aid in the solution of many geodesic problems, I also discuss their derivation and present some of their properties.

Consider a reference geodesic parametrized by distance s and a nearby geodesic separated from the reference by infinitesimal distance $t(s)$. Gauss (1828) showed that $t(s)$ satisfies the differential equation

$$\frac{\mathrm{d}^2 t(s)}{\mathrm{d}s^2} + K(s)\,t(s) = 0, \tag{27}$$

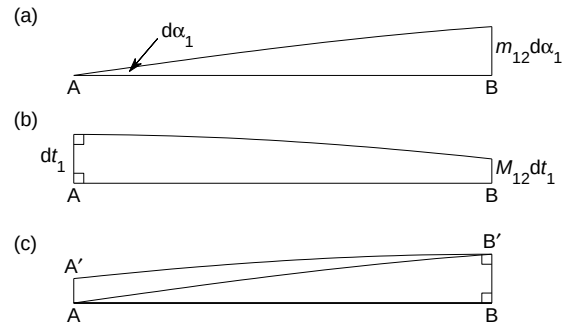

FIG. 3 The definitions of m_{12} and M_{12} are illustrated in (a) and (b). A geometric proof of Eq. (29) is shown in (c); here AB and $A'B'$ are parallel at B and B', $BAB' = \mathrm{d}\alpha_1$, $BB' = m_{12}\,\mathrm{d}\alpha_1$, $AA' = M_{21} m_{12}\,\mathrm{d}\alpha_1$, and finally $AB'A' = M_{21}\,\mathrm{d}\alpha_1$, from which Eq. (29) follows.

where $K(s)$ is the Gaussian curvature of the surface. As a second order, linear, homogeneous differential equation, its solution can be written as

$$t(s) = A t_A(s) + B t_B(s),$$

where A and B are (infinitesimal) constants and t_A and t_B are independent solutions. When considering the geodesic segment spanning $s_1 \le s \le s_2$, it is convenient to specify

$$t_A(s_1) = 0, \quad \left.\frac{\mathrm{d}t_A(s)}{\mathrm{d}s}\right|_{s=s_1} = 1,$$
$$t_B(s_1) = 1, \quad \left.\frac{\mathrm{d}t_B(s)}{\mathrm{d}s}\right|_{s=s_1} = 0,$$

and to write

$$m_{12} = t_A(s_2), \quad M_{12} = t_B(s_2).$$

The quantity m_{12} is the *reduced length* of the geodesic (Christoffel, 1868). Consider two geodesics which cross at $s = s_1$ at a small angle $\mathrm{d}\alpha_1$, Fig 3(a); at $s = s_2$, they will be separated by a distance $m_{12}\,\mathrm{d}\alpha_1$. Similarly I call M_{12} the *geodesic scale*. Consider two geodesics which are parallel at $s = s_1$ and separated by a small distance $\mathrm{d}t_1$, Fig 3(b); at $s = s_2$, they will be separated by a distance $M_{12}\,\mathrm{d}t_1$.

Several relations between m_{12} and M_{12} follow from the defining equation, Eq. (27). The reduced length obeys a reciprocity relation (Christoffel, 1868, §9), $m_{21} + m_{12} = 0$; the Wronskian is given by

$$W(M_{12}, m_{12})(s_2) = M_{12}\frac{\mathrm{d}m_{12}}{\mathrm{d}s_2} - m_{12}\frac{\mathrm{d}M_{12}}{\mathrm{d}s_2} = 1; \tag{28}$$

and the derivatives are

$$\frac{\mathrm{d}m_{12}}{\mathrm{d}s_2} = M_{21}, \tag{29}$$
$$\frac{\mathrm{d}M_{12}}{\mathrm{d}s_2} = -\frac{1 - M_{12}M_{21}}{m_{12}}. \tag{30}$$

The constancy of the Wronskian follows by noting that its derivative with respect to s_2 vanishes; its value is found by evaluating it at $s_2 = s_1$. A geometric proof of Eq. (29) is given in Fig 3(c) and Eq. (30) then follows from Eq. (28). With knowledge of the derivatives, addition rules for m_{12} and M_{12} are easily found,

$$m_{13} = m_{12}M_{23} + m_{23}M_{21}, \tag{31}$$

$$M_{13} = M_{12}M_{23} - (1 - M_{12}M_{21})\frac{m_{23}}{m_{12}}, \tag{32}$$

$$M_{31} = M_{32}M_{21} - (1 - M_{23}M_{32})\frac{m_{12}}{m_{23}}, \tag{33}$$

where points 1, 2, and 3 all lie on the same geodesic.

Geodesics allow concepts from plane geometry to be generalized to apply to a curved surface. In particular, a geodesic circle may be defined as the curve which is a constant geodesic distance from a fixed point. Similarly, a geodesic parallel to a reference curve is the curve which is a constant geodesic distance from that curve. (Thus a circle is a special case of a parallel obtained in the limit when the reference curve degenerates to a point.) Parallels occur naturally when considering, for example, the "12-mile limit" for territorial waters which is the boundary of points lying within 12 nautical miles of a coastal state.

The geodesic curvature of a parallel can be expressed in terms of m_{12} and M_{12}. Let point 1 be an arbitrary point on the reference curve with geodesic curvature κ_1. Point 2 is the corresponding point on the parallel, a fixed distance s_{12} away. The geodesic curvature of the parallel at that point is found from Eqs. (29) and (30),

$$\kappa_2 = \frac{M_{21}\kappa_1 - (1 - M_{12}M_{21})/m_{12}}{m_{12}\kappa_1 + M_{12}}. \tag{34}$$

The curvature of a circle is given by the limit $\kappa_1 \to \infty$,

$$\kappa_2 = M_{21}/m_{12}. \tag{35}$$

If the reference curve is a geodesic ($\kappa_1 \to 0$), then the curvature of its parallel is

$$\kappa_2 = -(1 - M_{12}M_{21})/(M_{12}m_{12}). \tag{36}$$

If the reference curve is indented, then the parallel intersects itself at a sufficiently large distance from the reference curve. This begins to happen when $\kappa_2 \to \infty$ in Eq. (34).

The results above apply to general surfaces. For a geodesic on an ellipsoid of revolution, the Gaussian curvature of the surface is given by

$$K = \frac{(1 - e^2\sin^2\phi)^2}{b^2} = \frac{1}{b^2(1 + k^2\sin^2\sigma)^2}. \tag{37}$$

Helmert (1880, Eq. (6.5.1)) solves Eq. (27) in this case to give

$$\begin{aligned} m_{12}/b = {} & \sqrt{1 + k^2\sin^2\sigma_2}\,\cos\sigma_1\sin\sigma_2 \\ & - \sqrt{1 + k^2\sin^2\sigma_1}\,\sin\sigma_1\cos\sigma_2 \\ & - \cos\sigma_1\cos\sigma_2\big(J(\sigma_2) - J(\sigma_1)\big), \end{aligned} \tag{38}$$

$$\begin{aligned} M_{12} = {} & \cos\sigma_1\cos\sigma_2 + \frac{\sqrt{1 + k^2\sin^2\sigma_2}}{\sqrt{1 + k^2\sin^2\sigma_1}}\sin\sigma_1\sin\sigma_2 \\ & - \frac{\sin\sigma_1\cos\sigma_2\big(J(\sigma_2) - J(\sigma_1)\big)}{\sqrt{1 + k^2\sin^2\sigma_1}}, \end{aligned} \tag{39}$$

where

$$\begin{aligned} J(\sigma) &= \int_0^\sigma \frac{k^2\sin^2\sigma'}{\sqrt{1 + k^2\sin^2\sigma'}}\,\mathrm{d}\sigma' \\ &= \frac{s}{b} - \int_0^\sigma \frac{1}{\sqrt{1 + k^2\sin^2\sigma'}}\,\mathrm{d}\sigma' \\ &= I_1(\sigma) - I_2(\sigma). \end{aligned} \tag{40}$$

Equation (39) may be obtained from Eq. (6.9.7) of Helmert (1880), which gives $\mathrm{d}m_{12}/\mathrm{d}s_2$; M_{12} may then be found from Eq. (29) with an interchange of indices. In the spherical limit, $f \to 0$, Eqs. (38) and (39) reduce to

$$\begin{aligned} m_{12} &= a\sin\sigma_{12} = a\sin(s_{12}/a), \\ M_{12} &= \cos\sigma_{12} = \cos(s_{12}/a). \end{aligned}$$

The integral $I_2(\sigma)$ in Eq. (40) may be expanded in a Fourier series in similar fashion to $I_1(\sigma)$, Eq. (15),

$$I_2(\sigma) = A_2\Big(\sigma + \sum_{l=1}^{\infty} C_{2l}\sin 2l\sigma\Big), \tag{41}$$

where

$$A_2 = (1 - \epsilon)\big(1 + \tfrac{1}{4}\epsilon^2 + \tfrac{9}{64}\epsilon^4 + \tfrac{25}{256}\epsilon^6 + \cdots\big), \tag{42}$$

$$\begin{aligned} C_{21} &= \tfrac{1}{2}\epsilon + \tfrac{1}{16}\epsilon^3 + \tfrac{1}{32}\epsilon^5 + \cdots, \\ C_{22} &= \tfrac{3}{16}\epsilon^2 + \tfrac{1}{32}\epsilon^4 + \tfrac{35}{2048}\epsilon^6 + \cdots, \\ C_{23} &= \tfrac{5}{48}\epsilon^3 + \tfrac{5}{256}\epsilon^5 + \cdots, \\ C_{24} &= \tfrac{35}{512}\epsilon^4 + \tfrac{7}{512}\epsilon^6 + \cdots, \\ C_{25} &= \tfrac{63}{1280}\epsilon^5 + \cdots, \\ C_{26} &= \tfrac{77}{2048}\epsilon^6. \end{aligned} \tag{43}$$

4. INVERSE PROBLEM

The inverse problem is intrinsically more complicated than the direct problem because the given included angle, λ_{12} in Fig. 1, is related to the corresponding angle on the auxiliary sphere ω_{12} via an unknown equatorial azimuth α_0. Thus, the inverse problem inevitably becomes a root-finding exercise.

I tackle this problem as follows. Assume that α_1 is known. Solve the *hybrid* geodesic problem: given ϕ_1, ϕ_2, and α_1, find λ_{12} corresponding to the first intersection of the geodesic with the circle of latitude ϕ_2. The resulting λ_{12} differs, in general, from the given λ_{12}; so adjust α_1 using Newton's method until the correct λ_{12} is obtained.

I begin by putting the points in a canonical configuration,

$$\phi_1 \le 0, \quad \phi_1 \le \phi_2 \le -\phi_1, \quad 0 \le \lambda_{12} \le \pi. \tag{44}$$

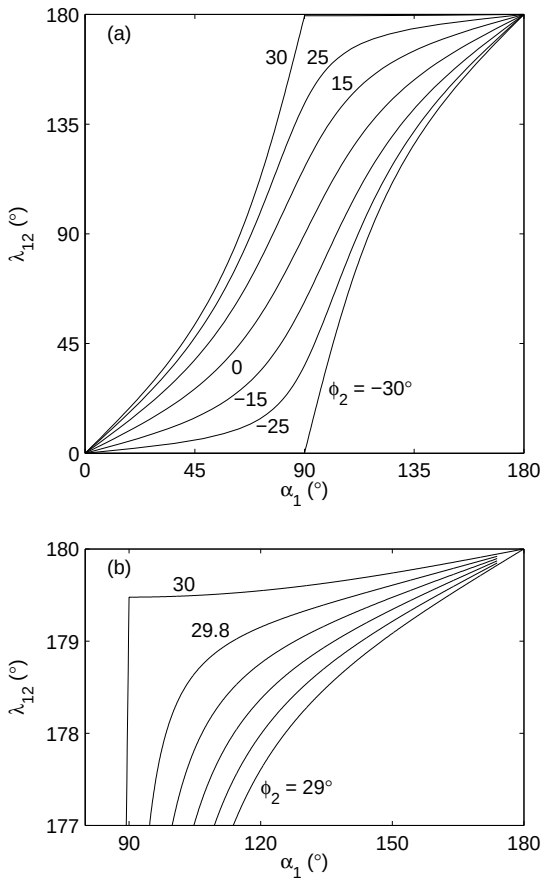

FIG. 4 The variation of λ_{12} as a function of α_1 for $\phi_1 = -30°$, various ϕ_2, and the WGS84 ellipsoid. Part (a) shows λ_{12} for $\phi_2 = 0°, \pm 15°, \pm 25°$, and $\pm 30°$. For $|\phi_2| < |\phi_1|$, the curves are strictly increasing, while for $\phi_2 = \pm\phi_1$, the curves are non-decreasing with discontinuities in the slopes at $\alpha_1 = 90°$. An enlargement of the top right corner of (a) is shown in (b) with $\phi_2 \in [29°, 30°]$ at intervals of $0.2°$.

This may be accomplished swapping the end points and the signs of the coordinates if necessary, and the solution may similarly be transformed to apply to the original points. All geodesics with $\alpha_1 \in [0, \pi]$ intersect latitude ϕ_2 with $\lambda_{12} \in [0, \pi]$. Furthermore, the search for solutions can be restricted to $\alpha_2 \in [0, \frac{1}{2}\pi]$, because this corresponds to the first intersection with latitude ϕ_2.

Meridional ($\lambda_{12} = 0$ or π) and equatorial ($\phi_1 = \phi_2 = 0$, with $\lambda_{12} \le (1 - f)\pi$) geodesics are treated as special cases, since the azimuth is then known: $\alpha_1 = \lambda_{12}$ and $\alpha_1 = \frac{1}{2}\pi$ respectively. The general case is solved by Newton's method as outlined above.

The solution of the hybrid geodesic problem is straightforward. Find β_1 and β_2 from Eq. (6), solve for α_0 and α_2 from Eq. (5), taking $\cos\alpha_0 > 0$ and $\cos\alpha_2 \ge 0$. In order to compute α_2 accurately, use

$$\cos\alpha_2 = \frac{+\sqrt{\cos^2\alpha_1\cos^2\beta_1 + (\cos^2\beta_2 - \cos^2\beta_1)}}{\cos\beta_2}, \quad (45)$$

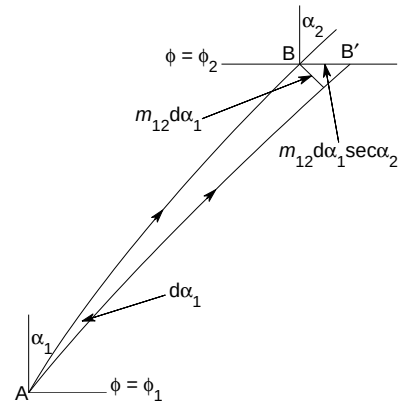

FIG. 5 Finding $\mathrm{d}\lambda_{12}/\mathrm{d}\alpha_1$ with ϕ_1 and ϕ_2 held fixed.

in addition to Eq. (5). Compute σ_1, ω_1, σ_2, and ω_2 using Eqs. (11) and (12). Finally, determine λ_{12} (and, once convergence is achieved, s_{12}) as in the solution to the direct problem. The behavior of λ_{12} as a function of α_1 is shown in Fig. 4.

To apply Newton's method, an expression for $\mathrm{d}\lambda_{12}/\mathrm{d}\alpha_1$ is needed. Consider a geodesic with initial azimuth α_1. If the azimuth is increased to $\alpha_1 + \mathrm{d}\alpha_1$ with its length held fixed, then the other end of the geodesic moves by $m_{12}\,\mathrm{d}\alpha_1$ in a direction $\frac{1}{2}\pi + \alpha_2$. If the geodesic is extended to intersect the parallel ϕ_2 once more, the point of intersection moves by $m_{12}\,\mathrm{d}\alpha_1/\cos\alpha_2$; see Fig. 5. The radius of this parallel is $a\cos\beta_2$; thus the rate of change of the longitude difference is

$$\frac{\mathrm{d}\lambda_{12}}{\mathrm{d}\alpha_1} = \frac{m_{12}}{a}\frac{1}{\cos\alpha_2\cos\beta_2}. \quad (46)$$

This equation can also be obtained from Eq. (6.9.8b) of Helmert (1880). Equation (46) becomes indeterminate when $\beta_2 = \pm\beta_1$ and $\alpha_1 = \frac{1}{2}\pi$, because m_{12} and $\cos\alpha_2$ both vanish. In this case, it is necessary to let $\alpha_1 = \frac{1}{2}\pi + \delta$ and to take the limit $\delta \to \pm 0$, which gives

$$\frac{\mathrm{d}\lambda_{12}}{\mathrm{d}\alpha_1} = -\frac{\sqrt{1 - e^2\cos^2\beta_1}}{\sin\beta_1}\big(1 \mp \mathrm{sign}(\cos\alpha_1)\big), \quad (47)$$

where $\mathrm{sign}(\cos\alpha_1) = -\mathrm{sign}(\delta)$. A numerical example of solving the inverse geodesic problem by this method is given at the end of the next section.

Vincenty (1975a), who uses the iterative method of Helmert (1880, §5.13) to solve the inverse problem, was aware of its failure to converge for nearly antipodal points. In an unpublished report (Vincenty, 1975b), he gives a modification of his method which deals with this case. Unfortunately, this sometimes requires many thousands of iterations to converge, whereas Newton's method as described here only requires a few iterations.

5. STARTING POINT FOR NEWTON'S METHOD

To complete the solution of the inverse problem a good starting guess for α_1 is needed. In most cases, this is provided

TABLE 3 First sample inverse calculation specified by $\phi_1 = -30.123\,45°$, $\phi_2 = -30.123\,44°$, and $\lambda_{12} = 0.000\,05°$. Because the points are not nearly antipodal, an initial guess for α_1 is found assuming $\omega_{12} = \lambda_{12}/\bar{w}$. However, in this case, the line is short enough that the error in ω_{12} is negligible at the precision given and the solution of the inverse problem is completed by using $s_{12} = a\bar{w}\sigma_{12}$. More generally, the value of α_1 would be refined using Newton's method.

Qty.	Value	Eq.
ϕ_1	$-30.123\,45°$	given
ϕ_2	$-30.123\,44°$	given
λ_{12}	$0.000\,05°$	given
Determine ω_{12}		
β_1	$-30.039\,990\,838\,21°$	(6)
β_2	$-30.039\,980\,854\,91°$	(6)
$\bar{w}$	$0.997\,488\,477\,44$	(48)
ω_{12}	$0.000\,050\,125\,89°$	$\lambda_{12}/\bar{w}$
σ_{12}	$0.000\,044\,526\,41°$	(51)
Solution		
α_1	$77.043\,533\,542\,37°$	(49)
α_2	$77.043\,508\,449\,13°$	(50)
s_{12}	$4.944\,208$ m	$a\bar{w}\sigma_{12}$

by assuming that $\omega_{12} = \lambda_{12}/\bar{w}$, where

$$\bar{w} = \sqrt{1 - e^2\big((\cos\beta_1 + \cos\beta_2)/2\big)^2} \tag{48}$$

and solving for the great circle on the auxiliary sphere, using (Vincenty, 1975a)

$$\begin{aligned} z_1 &= \cos\beta_1 \sin\beta_2 - \sin\beta_1 \cos\beta_2 \cos\omega_{12} \\ &\quad + i \cos\beta_2 \sin\omega_{12}, \\ z_2 &= -\sin\beta_1 \cos\beta_2 + \cos\beta_1 \sin\beta_2 \cos\omega_{12} \\ &\quad + i \cos\beta_1 \sin\omega_{12}, \\ \alpha_1 &= \operatorname{ph} z_1, \qquad (49) \\ \alpha_2 &= \operatorname{ph} z_2, \qquad (50) \\ \sigma_{12} &= \operatorname{ph}(\sin\beta_1 \sin\beta_2 + \cos\beta_1 \cos\beta_2 \cos\omega_{12} + i\,|z_1|). \qquad (51) \end{aligned}$$

An example of the solution of the inverse problem by this method is given in Table 3.

This procedure is inadequate for nearly antipodal points because both the real and imaginary components of z_1 are small and α_1 depends very sensitively on ω_{12}. In the corresponding situation on the sphere, it is possible to determine α_1 by noting that all great circles emanating from A meet at O, the point antipodal to A. Thus α_1 may be determined as the supplement of the azimuth of the great circle BO at O; in addition, because B and O are close, it is possible to approximate the sphere, locally, as a plane.

The situation for an ellipsoid is slightly different because the geodesics emanating from A, instead of meeting at a point, form an envelope, centered at O, in the shape of an *astroid*

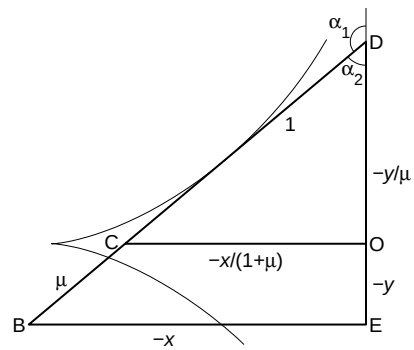

FIG. 6 The solution of the astroid equations by similar triangles. The scaled coordinates of B are (x, y); O is the point antipodal to A. The line BCD, which is given by Eq. (54), is the continuation of the geodesic from AB with C being its intersection with the circle $\beta = -\beta_1$ and D its intersection with the meridian $\lambda = \lambda_1 + \pi$. The envelope of lines satisfying $CD = 1$ gives the astroid, a portion of which is shown by the curves.

whose extent is $O(f)$ (Jacobi, 1891, Eqs. (16)–(17)). The position at which a particular geodesic touches this envelope is given by the condition $m_{12} = 0$. However elementary methods can be used to determine the envelope. Consider a geodesic leaving A (with $\beta_1 \le 0$) with azimuth $\alpha_1 \in [\frac{1}{2}\pi, \pi]$. This first intersects the circle of opposite latitude, $\beta_2 = -\beta_1$, with $\sigma_{12} = \omega_{12} = \pi$ and $\alpha_2 = \pi - \alpha_1$. Equation (8) then gives

$$\lambda_{12} = \pi - f\pi \cos\beta_1 \sin\alpha_1 + O(f^2). \tag{52}$$

Define a plane coordinate system (x, y) centered on the antipodal point where $\Delta = f a \pi \cos^2\beta_1$ is the unit of length, i.e.,

$$\lambda_{12} = \pi + \frac{\Delta}{a\cos\beta_1} x, \quad \beta_2 = -\beta_1 + \frac{\Delta}{a} y. \tag{53}$$

In this coordinate system, Eq. (52) corresponds to the point $x = -\sin\alpha_1$, $y = 0$ and the slope of the geodesic is $-\cot\alpha_1$. Thus, in the neighborhood of the antipodal point, the geodesic may be approximated by

$$\frac{x}{\sin\alpha_1} + \frac{y}{\cos\alpha_1} + 1 = 0, \tag{54}$$

where terms of order f^2 have been neglected. Allowing α_1 to vary, Eq. (54) defines a family of lines approximating the geodesics emanating from A. Differentiating this equation with respect to α_1 and solving the resulting pair of equations for x and y gives the parametric equations for the astroid, $x = -\sin^3\alpha_1$, and $y = -\cos^3\alpha_1$. Note that, for the ordering of points given by Eq. (44), $x \le 0$ and $y \le 0$.

Given x and y (i.e., the position of point B), Eq. (54) may be solved to obtain a first approximation to α_1. This prescription is given by Helmert (1880, Eq. (7.3.7)) who notes that this results in a quartic which may be found using the construction given in Fig. 6. Here COD and BED are similar triangles; if the (signed) length BC is μ, then an equation for μ can be

TABLE 4 Second sample inverse calculation specified by $\phi_1 = -30°$, $\phi_2 = 29.9°$, and $\lambda_{12} = 179.8°$. Because the points are nearly antipodal, an initial guess for α_1 is found by solving the astroid problem. Here μ is the positive root of Eq. (55). If $y = 0$, then α_1 is given by Eq. (57). The value of α_1 is used in Table 5.

Qty.	Value	Eq.
ϕ_1	$-30°$	given
ϕ_2	$29.9°$	given
λ_{12}	$179.8°$	given
Solve the astroid problem		
x	$-0.382\,344$	(53)
y	$-0.220\,189$	(53)
μ	$0.231\,633$	(55)
Initial guess for α_1		
α_1	$161.914°$	(56)

found by applying Pythagoras' theorem to COD,

$$\frac{x^2}{(1+\mu)^2} + \frac{y^2}{\mu^2} = 1,$$

which can be expanded to give a 4th-order polynomial in μ,

$$\mu^4 + 2\mu^3 + (1 - x^2 - y^2)\mu^2 - 2y^2\mu - y^2 = 0. \qquad (55)$$

Descartes' rule of signs shows that, for $y \neq 0$, there is one positive root (Olver *et al.*, 2010, §1.11(ii)) and this is the solution corresponding to the shortest path. This root can be found by standard methods (Olver *et al.*, 2010, §1.11(iii)). Equation (55) arises in converting from geocentric to geodetic coordinates, and I use the solution to that problem given by Vermeille (2002). The azimuth can then be determined from the triangle COD in Fig. 6,

$$\alpha_1 = \mathrm{ph}\big(y/\mu - ix/(1+\mu)\big). \qquad (56)$$

If $y = 0$, the solution is found by taking the limit $y \to 0$,

$$\alpha_1 = \mathrm{ph}\big(\pm\sqrt{\max(0, 1 - x^2)} - ix\big). \qquad (57)$$

Tables 4–6 together illustrate the complete solution of the inverse problem for nearly antipodal points.

6. AREA

The last geodesic algorithm I present is for geodesic areas. Here, I extend the method of Danielsen (1989) to higher order so that the result is accurate to round-off, and I recast his series into a simple trigonometric sum.

Let S_{12} be the area of the geodesic quadrilateral $AFHB$ in Fig. 1. Following Danielsen (1989), this can be expressed as the sum of a spherical term and an integral giving the ellipsoidal correction,

$$S_{12} = S(\sigma_2) - S(\sigma_1), \qquad (58)$$

TABLE 5 Second sample inverse calculation, continued. Here $\lambda_{12}^{(0)}$ denotes the desired value of the longitude difference; Newton's method is used to adjust α_1 so that $\lambda_{12} = \lambda_{12}^{(0)}$. The final value of α_1 is used in Table 6.

Qty.	Value	Eq.
ϕ_1	$-30°$	given
ϕ_2	$29.9°$	given
α_1	$161.914°$	Table 4
$\lambda_{12}^{(0)}$	$179.8°$	given
Solve triangle NEA		
β_1	$-29.916\,747\,713\,24°$	(6)
α_0	$15.609\,397\,464\,14°$	(10)
σ_1	$-148.812\,535\,665\,96°$	(11)
ω_1	$-170.748\,966\,961\,28°$	(12)
Solve triangle NEB		
β_2	$29.816\,916\,421\,89°$	(6)
α_2	$18.067\,287\,962\,31°$	(5), (45)
σ_2	$31.082\,449\,768\,95°$	(11)
ω_2	$9.213\,457\,611\,10°$	(12)
Determine λ_{12}		
k^2	$0.006\,251\,537\,916\,62$	(9)
ϵ	$0.001\,558\,018\,267\,80$	(16)
λ_1	$-170.614\,835\,524\,58°$	(8)
λ_2	$9.185\,420\,098\,39°$	(8)
λ_{12}	$179.800\,255\,622\,97°$	$\lambda_2 - \lambda_1$
Update α_1		
$\delta\lambda_{12}$	$0.000\,255\,622\,97°$	$\lambda_{12} - \lambda_{12}^{(0)}$
$J(\sigma_1)$	$-0.009\,480\,409\,276\,40$	(40)
$J(\sigma_2)$	$0.000\,313\,491\,286\,30$	(40)
m_{12}	$57\,288.000\,110\,\mathrm{m}$	(38)
$\mathrm{d}\lambda_{12}/\mathrm{d}\alpha_1$	$0.010\,889\,317\,161\,15$	(46)
$\delta\alpha_1$	$-0.023\,474\,655\,19°$	$-\delta\lambda_{12}/(\mathrm{d}\lambda_{12}/\mathrm{d}\alpha_1)$
α_1	$161.890\,525\,344\,81°$	$\alpha_1 + \delta\alpha_1$
Next iteration		
$\delta\lambda_{12}$	$0.000\,000\,006\,63°$	
α_1	$161.890\,524\,736\,33°$	

$$S(\sigma) = c^2\alpha + e^2a^2\cos\alpha_0\sin\alpha_0\, I_4(\sigma), \qquad (59)$$

where

$$c^2 = \frac{a^2}{2} + \frac{b^2}{2}\frac{\tanh^{-1} e}{e} \qquad (60)$$

is the authalic radius,

$$I_4(\sigma) = -\int_{\pi/2}^{\sigma} \frac{t(e'^2) - t(k^2\sin^2\sigma')}{e'^2 - k^2\sin^2\sigma'}\frac{\sin\sigma'}{2}\,\mathrm{d}\sigma', \qquad (61)$$
$$t(x) = x + \sqrt{x^{-1}+1}\,\sinh^{-1}\sqrt{x}.$$

Expanding the integrand in powers of e'^2 and k^2 and performing the integral gives

$$I_4(\sigma) = \sum_{l=0}^{\infty} C_{4l}\cos\big((2l+1)\sigma\big), \qquad (62)$$

TABLE 6 Second sample inverse calculation, concluded. Here the hybrid problem (ϕ_1, ϕ_2, and α_1 given) is solved. The computed value of λ_{12} matches that given in the specification of the inverse problem in Table 4.

Qty.	Value	Eq.
ϕ_1	$-30°$	given
ϕ_2	$29.9°$	given
α_1	$161.890\,524\,736\,33°$	Table 5
Solve triangle NEA		
β_1	$-29.916\,747\,713\,24°$	(6)
α_0	$15.629\,479\,665\,37°$	(10)
σ_1	$-148.809\,136\,917\,76°$	(11)
ω_1	$-170.736\,343\,780\,66°$	(12)
Solve triangle NEB		
β_2	$29.816\,916\,421\,89°$	(6)
α_2	$18.090\,737\,245\,74°$	(5), (45)
σ_2	$31.085\,834\,470\,40°$	(11)
ω_2	$9.226\,028\,621\,10°$	(12)
Determine s_{12} and λ_{12}		
s_1	$-16\,539\,979.064\,227\,\mathrm{m}$	(7)
s_2	$3\,449\,853.763\,383\,\mathrm{m}$	(7)
s_{12}	$19\,989\,832.827\,610\,\mathrm{m}$	$s_2 - s_1$
λ_1	$-170.602\,047\,121\,48°$	(8)
λ_2	$9.197\,952\,878\,52°$	(8)
λ_{12}	$179.800\,000\,000\,00°$	$\lambda_2 - \lambda_1$
Solution		
α_1	$161.890\,524\,736\,33°$	
α_2	$18.090\,737\,245\,74°$	
s_{12}	$19\,989\,832.827\,610\,\mathrm{m}$	

where

$$
\begin{aligned}
C_{40} &= \left(\tfrac{2}{3} - \tfrac{1}{15}e'^2 + \tfrac{4}{105}e'^4 - \tfrac{8}{315}e'^6 + \tfrac{64}{3465}e'^8 - \tfrac{128}{9009}e'^{10}\right) \\
&\quad - \left(\tfrac{1}{20} - \tfrac{1}{35}e'^2 + \tfrac{2}{105}e'^4 - \tfrac{16}{1155}e'^6 + \tfrac{32}{3003}e'^8\right)k^2 \\
&\quad + \left(\tfrac{1}{42} - \tfrac{1}{63}e'^2 + \tfrac{8}{693}e'^4 - \tfrac{80}{9009}e'^6\right)k^4 \\
&\quad - \left(\tfrac{1}{72} - \tfrac{1}{99}e'^2 + \tfrac{10}{1287}e'^4\right)k^6 \\
&\quad + \left(\tfrac{1}{110} - \tfrac{1}{143}e'^2\right)k^8 - \tfrac{1}{156}k^{10} + \cdots, \\
C_{41} &= \left(\tfrac{1}{180} - \tfrac{1}{315}e'^2 + \tfrac{2}{945}e'^4 - \tfrac{16}{10395}e'^6 + \tfrac{32}{27027}e'^8\right)k^2 \\
&\quad - \left(\tfrac{1}{252} - \tfrac{1}{378}e'^2 + \tfrac{4}{2079}e'^4 - \tfrac{40}{27027}e'^6\right)k^4 \\
&\quad + \left(\tfrac{1}{360} - \tfrac{1}{495}e'^2 + \tfrac{2}{1287}e'^4\right)k^6 \\
&\quad - \left(\tfrac{1}{495} - \tfrac{2}{1287}e'^2\right)k^8 + \tfrac{5}{3276}k^{10} + \cdots, \\
C_{42} &= \left(\tfrac{1}{2100} - \tfrac{1}{3150}e'^2 + \tfrac{4}{17325}e'^4 - \tfrac{8}{45045}e'^6\right)k^4 \\
&\quad - \left(\tfrac{1}{1800} - \tfrac{1}{2475}e'^2 + \tfrac{2}{6435}e'^4\right)k^6 \\
&\quad + \left(\tfrac{1}{1925} - \tfrac{2}{5005}e'^2\right)k^8 - \tfrac{1}{2184}k^{10} + \cdots, \\
C_{43} &= \left(\tfrac{1}{17640} - \tfrac{1}{24255}e'^2 + \tfrac{2}{63063}e'^4\right)k^6 \\
&\quad - \left(\tfrac{1}{10780} - \tfrac{1}{14014}e'^2\right)k^8 + \tfrac{5}{45864}k^{10} + \cdots, \\
C_{44} &= \left(\tfrac{1}{124740} - \tfrac{1}{162162}e'^2\right)k^8 - \tfrac{1}{58968}k^{10} + \cdots, \\
C_{45} &= \tfrac{1}{792792}k^{10} + \cdots.
\end{aligned}
\tag{63}
$$

TABLE 7 The calculation of the area between the equator and the geodesic specified by $\phi_1 = 40°$, $\alpha_1 = 30°$, and $s_{12} = 10\,000\,\mathrm{km}$. This uses intermediate values computed in Table 2.

Qty.	Value	Eq.
α_0	$22.553\,940\,202\,62°$	Table 2
α_1	$30°$	Table 2
α_2	$149.090\,169\,318\,07°$	Table 2
σ_1	$43.999\,153\,645\,00°$	Table 2
σ_2	$133.921\,640\,830\,38°$	Table 2
k^2	$0.005\,748\,029\,628\,57$	Table 2
Compute area		
$I_4(\sigma_1)$	$0.479\,018\,145\,20$	(62)
$I_4(\sigma_2)$	$-0.461\,917\,119\,02$	(62)
$S(\sigma_1)$	$21\,298\,942.667\,15\,\mathrm{km}^2$	(59)
$S(\sigma_2)$	$105\,574\,566.089\,50\,\mathrm{km}^2$	(59)
S_{12}	$84\,275\,623.422\,35\,\mathrm{km}^2$	(58)

An example of the computation of S_{12} is given in Table 7.

Summing S_{12}, Eq. (58), over the edges of a geodesic polygon gives the area of the polygon provided that it does not encircle a pole; if it does, $2\pi c^2$ should be added to the result. The first term in Eq. (59) contributes $c^2(\alpha_2 - \alpha_1)$ to S_{12}. This is the area of the quadrilateral $AFHB$ on a sphere of radius c and it is proportional to its spherical excess, $\alpha_2 - \alpha_1$, the sum of its interior angles less 2π. It is important that this term be computed accurately when the edge is short (and α_1 and α_2 are nearly equal). A suitable identity for $\alpha_2 - \alpha_1$ is given by Bessel (1825, §11),

$$
\tan\frac{\alpha_2 - \alpha_1}{2} = \frac{\sin\frac{1}{2}(\beta_2 + \beta_1)}{\cos\frac{1}{2}(\beta_2 - \beta_1)}\tan\frac{\omega_{12}}{2}. \tag{64}
$$

7. IMPLEMENTATION

The algorithms described in the preceding sections can be readily converted into working code. The polynomial expansions, Eqs. (17), (18), (21), (24), (25), (42), (43), and (63), are such that the final results are accurate to $O(f^6)$ which means that, even for $f = \frac{1}{150}$, the truncation error is smaller than the round-off error when using IEEE double precision arithmetic (with the fraction of the floating point number represented by 53 bits). For speed and to minimize round-off errors, the polynomials should be evaluated with the Horner method. The parenthetical expressions in Eqs. (24), (25), and (63) depend only on the flattening of the ellipsoid and can be computed once this is known. When determining many points along a single geodesic, the polynomials need be evaluated just once. Clenshaw (1955) summation should be used to sum the Fourier series, Eqs. (15), (23), (41), and (62).

There are several other details to be dealt with in implementing the algorithms: where to apply the two rules for choosing starting points for Newton's method, a slight improvement to the starting guess Eq. (56), the convergence cri-

terion for Newton's method, how to minimize round-off errors in solving the trigonometry problems on the auxiliary sphere, rapidly computing intermediate points on a geodesic by using σ_{12} as the metric, etc. I refer the reader to the implementations of the algorithms in GeographicLib (Karney, 2012) for possible ways to address these issues. The C++ implementation has been tested against a large set of geodesics for the WGS84 ellipsoid; this was generated by continuing the series expansions to $O(f^{30})$ and by solving the direct problem using with high-precision arithmetic. The round-off errors in the direct and inverse methods are less than 15 nanometers and the error in the computation of the area S_{12} is about $0.1\,\mathrm{m}^2$. Typically, 2 to 4 iterations of Newton's method are required for convergence, although in a tiny fraction of cases up to 16 iterations are required. No convergence failures are observed. With the C++ implementation compiled with the g++ compiler, version 4.4.4, and running on a $2.66\,\mathrm{GHz}$ Intel processor, solving the direct geodesic problem takes $0.88\,\mu\mathrm{s}$, while the inverse problem takes $2.34\,\mu\mathrm{s}$ (on average). Several points along a geodesic can be computed at the rate of $0.37\,\mu\mathrm{s}$ per point. These times are comparable to those for Vincenty's algorithms implemented in C++ and run on the same architecture: $1.11\,\mu\mathrm{s}$ for the direct problem and $1.34\,\mu\mathrm{s}$ for the inverse problem. (But note that Vincenty's algorithms are less accurate than those given here and that his method for the inverse problem sometimes fails to converge.)

8. ELLIPSOIDAL GNOMONIC PROJECTION

As an application of the differential properties of geodesics, I derive a generalization of the gnomonic projection to the ellipsoid. The gnomonic projection of the sphere has the property that all geodesics on the sphere map to straight lines (Snyder, 1987, §22). Such a projection is impossible for an ellipsoid because it does not have constant Gaussian curvature (Beltrami, 1865, §18); nevertheless, a projection can be constructed in which geodesics are very nearly straight.

The spherical gnomonic projection is the limit of the doubly azimuthal projection of the sphere, wherein the bearings from two fixed points A and A' to B are preserved, as A' approaches A (Bugayevskiy and Snyder, 1995). The construction of the generalized gnomonic projection proceeds in the same way; see Fig. 7. Draw a geodesic $A'B'$ such that it is parallel to the geodesic AB at A. Its initial separation from AB is $\sin\gamma\,\mathrm{d}t$; at B', the point closest to B, the separation becomes $M_{12}\sin\gamma\,\mathrm{d}t$ (in the limit $\mathrm{d}t \to 0$). Thus the difference in the azimuths of the geodesics $A'B$ and $A'B'$ at A' is $(M_{12}/m_{12})\sin\gamma\,\mathrm{d}t$, which gives $\gamma + \gamma' = \pi - (M_{12}/m_{12})\sin\gamma\,\mathrm{d}t$. Now, solving the planar triangle problem with γ and γ' as the two base angles gives the distance AB on the projection plane as m_{12}/M_{12}.

This leads to the following specification for the generalized gnomonic projection. Let the center point be A; for an arbitrary point B, solve the inverse geodesic problem between A and B; then B projects to the point

$$x = \rho\sin\alpha_1, \quad y = \rho\cos\alpha_1, \quad \rho = m_{12}/M_{12}; \tag{65}$$

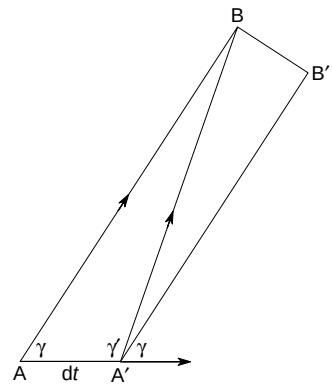

FIG. 7 The construction of the generalized gnomonic projection as the limit of a doubly azimuthal projection.

the projection is undefined if $M_{12} \le 0$. In the spherical limit, this becomes the standard gnomonic projection, $\rho = a\tan\sigma_{12}$ (Snyder, 1987, p. 165). The azimuthal scale is $1/M_{12}$ and the radial scale, found by taking the derivative $\mathrm{d}\rho/\mathrm{d}s_{12}$ and using Eq. (28), is $1/M_{12}^2$. The reverse projection is found by computing $\alpha_1 = \mathrm{ph}(y + ix)$, finding s_{12} using Newton's method with $\mathrm{d}\rho/\mathrm{d}s_{12} = 1/M_{12}^2$ (i.e., the radial scale), and solving the resulting direct geodesic problem.

In order to gauge the usefulness of the ellipsoidal gnomonic projection, consider two points on the earth B and C, map these points to the projection, and connect them with a straight line in this projection. If this line is mapped back onto the surface of the earth, it will deviate slightly from the geodesic BC. To lowest order, the maximum deviation $\mathbf{h}$ occurs at the midpoint of the line segment BC; empirically, I find

$$\mathbf{h} = \frac{l^2}{32}(\nabla K \cdot \mathbf{t})\mathbf{t}, \tag{66}$$

where l is the length of the geodesic, K is the Gaussian curvature, ∇K is evaluated at the center of the projection A, and $\mathbf{t}$ is the perpendicular vector from the center of projection to the geodesic. The deviation in the azimuths at the end points is about $4h/l$ and the length is greater than the geodesic distance by about $\frac{8}{3}h^2/l$. In the case of an ellipsoid of revolution, the curvature is given by differentiating Eq. (37) with respect to ϕ and dividing by the meridional radius of curvature to give

$$\nabla K = -\frac{4a}{b^4}e^2(1 - e^2\sin^2\phi)^{5/2}\cos\phi\sin\phi\,\hat{\boldsymbol{\phi}}, \tag{67}$$

where $\hat{\boldsymbol{\phi}}$ is a unit vector pointing north. Bounding the magnitude of $\mathbf{h}$, Eq. (66), over all the geodesics whose end points lie within a distance r of the center of projection, gives (in the limit that f and r are small)

$$\frac{h}{r} \le \frac{f}{8}\frac{r^3}{a^3}. \tag{68}$$

The maximum value is attained when the center of projection is at $\phi = \pm 45^\circ$ and the geodesic is running in an east-west

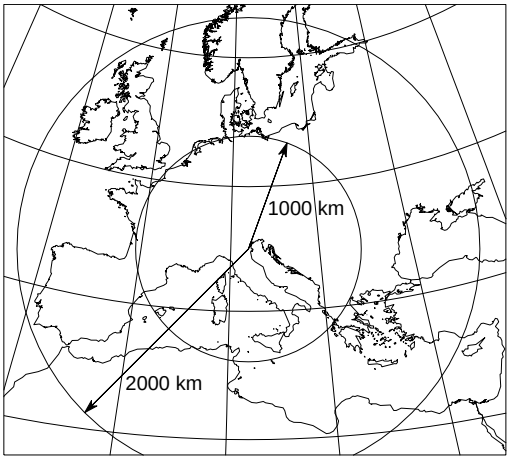

FIG. 8 The coast line of Europe and North Africa in the ellipsoidal gnomonic projection with center at $(45°\mathrm{N}, 12°\mathrm{E})$ near Venice. The graticule lines are shown at multiples of $10°$. The two circles are centered on the projection center with (geodesic) radii of $1000\,\mathrm{km}$ and $2000\,\mathrm{km}$. The data for the coast lines is taken from GMT (Wessel and Smith, 2010) at "low" resolution.

direction with the end points at bearings $\pm 45°$ or $\pm 135°$ from the center.

Others have proposed different generalizations of the gnomonic projection. Bowring (1997) and Williams (1997) give a projection in which great ellipses project to straight lines; Letoval'tsev (1963) suggests a projection in which normal sections through the center point map to straight lines. Empirically, I find that h/r is proportional to r/a and r^2/a^2 for these projections. Thus, neither does as well as the projection derived above (for which h/r is proportional to r^3/a^3) at preserving the straightness of geodesics.

As an illustration of the properties of the ellipsoidal gnomonic projection, Eq. (65), consider Fig. 8 in which a projection of Europe is shown. The two circles are geodesic circles of radii $1000\,\mathrm{km}$ and $2000\,\mathrm{km}$. If the geodesic between any two points within one of these circles is estimated by using a straight line on this figure, the deviation from the true geodesic is less than $1.7\,\mathrm{m}$ and $28\,\mathrm{m}$, respectively. The maximum errors in the end azimuths are $1.1''$ and $8.6''$ and the maximum errors in the lengths are only $5.4\,\mu\mathrm{m}$ and $730\,\mu\mathrm{m}$.

The gnomonic projection can be used to solve two geodesic problems accurately and rapidly. The first is the *intersection* problem: given two geodesics between A and B and between C and D, determine the point of intersection, O. This can be solved as follows. Guess an intersection point $O^{(0)}$ and use this as the center of the gnomonic projection; define $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ as the positions of A, B, C, D in the projection; find the intersection of AB and CD in the projection, i.e.,

$$\mathbf{o} = \frac{(\mathbf{c}\times\mathbf{d}\cdot\hat{\mathbf{z}})(\mathbf{b}-\mathbf{a}) - (\mathbf{a}\times\mathbf{b}\cdot\hat{\mathbf{z}})(\mathbf{d}-\mathbf{c})}{(\mathbf{b}-\mathbf{a})\times(\mathbf{d}-\mathbf{c})\cdot\hat{\mathbf{z}}}, \tag{69}$$

where $\hat{}$ indicates a unit vector ($\hat{\mathbf{a}} = \mathbf{a}/a$) and $\hat{\mathbf{z}} = \hat{\mathbf{x}}\times\hat{\mathbf{y}}$ is in the direction perpendicular to the projection plane. Project $\mathbf{o}$ back to geographic coordinates $O^{(1)}$ and use this as a new center of projection; iterate this process until $O^{(i)} = O^{(i-1)}$ which is then the desired intersection point.

The second problem is the *interception* problem: given a geodesic between A and B, find the point O on the geodesic which is closest to a given point C. The solution is similar to that for the intersection problem; however the interception point in the projection is

$$\mathbf{o} = \frac{\mathbf{c}\cdot(\mathbf{b}-\mathbf{a})(\mathbf{b}-\mathbf{a}) - (\mathbf{a}\times\mathbf{b}\cdot\hat{\mathbf{z}})\hat{\mathbf{z}}\times(\mathbf{b}-\mathbf{a})}{|\mathbf{b}-\mathbf{a}|^2}.$$

Provided the given points lie within about a quarter meridian of the intersection or interception points (so that the gnomonic projection is defined), these algorithms converge quadratically to the exact result.

9. CONCLUSIONS

The classical geodesic problems entail solving the ellipsoidal triangle NAB in Fig. 1, whose sides and angles are represented by ϕ_1, ϕ_2, s_{12} and α_1, α_2, λ_{12}. In the direct problem ϕ_1, α_1, and s_{12} are given, while in the inverse problem ϕ_1, λ_{12}, and ϕ_2 are specified; and the goal in each case is to solve for the remaining side and angles. The algorithms given here provide accurate, robust, and fast solutions to these problems; they also allow the differential and integral quantities m_{12}, M_{12}, M_{21}, and S_{12} to be computed.

Much of the work described here involves applying standard computational techniques to earlier work. However, at least two aspects are novel: (1) This paper presents the first complete solution to the inverse geodesic problem. (2) The ellipsoidal gnomonic projection is a new tool to solve various geometrical problems on the ellipsoid.

Furthermore, the packaging of these various geodesic capabilities into a single library is also new. This offers a straightforward solution of several interesting problems. Two geodesic projections, the azimuthal equidistant projection and the Cassini-Soldner projection, are simple to write and their domain of applicability is not artificially restricted, as would be the case, for example, if the series expansion for the Cassini-Soldner projection were used (Snyder, 1987, §13); the scales for these projections are simply given in terms of m_{12} and M_{12}. Several other problems can be readily tackled with this library, e.g., solving other ellipsoidal trigonometry problems and finding the median line and other maritime boundaries. These and other problems are explored in Karney (2011). The web page http://geographiclib.sf.net/geod.html provides additional information, including the Maxima (2009) code used to carry out the Taylor expansions and a JavaScript implementation which allows geodesic problems to be solved on many portable devices.

Acknowledgments

I would like to thank Rod Deakin, John Nolton, Peter Osborne, and the referees of this paper for their helpful comments.

References

E. Beltrami, 1865, *Risoluzione del problema: Riportare i punti di una superficie sopra un piano in modo che le linee geodetiche vengano rappresentate da linee rette*, Annali Mat. Pura App., **7**, 185–204, http://books.google.com/books?id=dfgEAAAAYAAJ&pg=PA185.

F. W. Bessel, 1825, *Über die Berechnung der geographischen Längen und Breiten aus geodätischen Vermessungen*, Astron. Nachr., **4**(86), 241–254, http://adsabs.harvard.edu/abs/1825AN.....4..241B, translated into English by C. F. F. Karney and R. E. Deakin as *The calculation of longitude and latitude from geodesic measurements*, Astron. Nachr. **331**(8), 852–861 (2010), doi:10.1002/asna.201011352, E-print arXiv:0908.1824.

B. R. Bowring, 1997, *The central projection of the spheroid and surface lines*, Survey Review, **34**(265), 163–173.

L. M. Bugayevskiy and J. P. Snyder, 1995, *Map Projections: A Reference Manual* (Taylor & Francis, London), http://www.worldcat.org/oclc/31737484.

E. B. Christoffel, 1868, *Allgemeine Theorie der geodätischen Dreiecke*, Math. Abhand. König. Akad. der Wiss. zu Berlin, **8**, 119–176, http://books.google.com/books?id=EEtFAAAAcAAJ&pg=PA119.

A. C. Clairaut, 1735, *Détermination géometrique de la perpendiculaire à la méridienne tracée par M. Cassini*, Mém. de l'Acad. Roy. des Sciences de Paris 1733, pp. 406–416, http://books.google.com/books?id=GOAEAAAAQAAJ&pg=PA406.

C. W. Clenshaw, 1955, *A note on the summation of Chebyshev series*, Math. Tables Aids Comput., **9**(51), 118–120, http://www.jstor.org/stable/2002068.

J. S. Danielsen, 1989, *The area under the geodesic*, Survey Review, **30**(232), 61–66.

C. F. Gauss, 1828, *Disquisitiones Generales circa Superficies Curvas* (Dieterich, Göttingen), http://books.google.com/books?id=bX0AAAAAMAAJ, translated into English by J. C. Morehead and A. M. Hiltebeitel as *General Investigations of Curved Surfaces of 1827 and 1825* (Princeton Univ. Lib., 1902), http://books.google.com/books?id=a1wTJR3kHwUC.

F. R. Helmert, 1880, *Die Mathematischen und Physikalischen Theorieen der Höheren Geodäsie*, volume 1 (Teubner, Leipzig), http://books.google.com/books?id=qt2CAAAAIAAJ, translated into English by Aeronautical Chart and Information Center (St. Louis, 1964) as *Mathematical and Physical Theories of Higher Geodesy, Part 1*, http://geographiclib.sf.net/geodesic-papers/helmert80-en.html.

C. G. J. Jacobi, 1891, *Über die Curve, welche alle von einem Punkte ausgehenden geodätischen Linien eines Rotationsellipsoides berührt*, in K. T. W. Weierstrass, editor, *Gesammelte Werke*, volume 7, pp. 72–87 (Reimer, Berlin), op. post., completed by F. H. A. Wangerin, http://books.google.com/books?id=_09tAAAAMAAJ&pg=PA72.

C. F. F. Karney, 2011, *Geodesics on an ellipsoid of revolution*, Technical report, SRI International, E-print arXiv:1102.1215v1.

—, 2012, *GeographicLib, version 1.20*, http://geographiclib.sf.net.

A. M. Legendre, 1806, *Analyse des triangles tracés sur la surface d'un sphéroïde*, Mém. de l'Inst. Nat. de France, 1st sem., pp. 130–161, http://books.google.com/books?id=-d0EAAAAQAAJ&pg=PA130-IA4.

I. G. Letoval'tsev, 1963, *Generalization of the gnomonic projection for a spheroid and the principal geodetic problems involved in the alignment of surface routes*, Geodesy and Aerophotography, **5**, 271–274, translation of Geodeziya i Aerofotos'emka **5**, 61–68 (1963).

Maxima, 2009, *A computer algebra system, version 5.20.1*, http://maxima.sf.net.

F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors, 2010, *NIST Handbook of Mathematical Functions* (Cambridge Univ. Press), http://dlmf.nist.gov.

B. Oriani, 1806, *Elementi di trigonometria sferoidica, Pt. 1*, Mem. dell'Ist. Naz. Ital., **1**(1), 118–198, http://books.google.com/books?id=SydFAAAAcAAJ&pg=PA118.

—, 1808, *Elementi di trigonometria sferoidica, Pt. 2*, Mem. dell'Ist. Naz. Ital., **2**(1), 1–58, http://www.archive.org/stream/memoriedellistit21isti#page/1.

—, 1810, *Elementi di trigonometria sferoidica, Pt. 3*, Mem. dell'Ist. Naz. Ital., **2**(2), 1–58, http://www.archive.org/stream/memoriedellistit22isti#page/1.

R. H. Rapp, 1993, *Geometric geodesy, part II*, Technical report, Ohio State Univ., http://hdl.handle.net/1811/24409.

J. P. Snyder, 1987, *Map projection—a working manual*, Professional Paper 1395, U.S. Geological Survey, http://pubs.er.usgs.gov/publication/pp1395.

H. Vermeille, 2002, *Direct transformation from geocentric coordinates to geodetic coordinates*, J. Geod., **76**(9), 451–454, doi:10.1007/s00190-002-0273-6.

T. Vincenty, 1975a, *Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations*, Survey Review, **23**(176), 88–93, addendum: Survey Review **23**(180), 294 (1976), http://www.ngs.noaa.gov/PUBS_LIB/inverse.pdf.

—, 1975b, *Geodetic inverse solution between antipodal points*, unpublished report dated Aug. 28, http://geographiclib.sf.net/geodesic-papers/vincenty75b.pdf.

P. Wessel and W. H. F. Smith, 2010, *Generic mapping tools, 4.5.5*, http://gmt.soest.hawaii.edu/.

R. Williams, 1997, *Gnomonic projection of the surface of an ellipsoid*, J. Nav., **50**(2), 314–320, doi:10.1017/S0373463300023936.