T. Vincenty

Vol. XXIII No. 176

SURVEY REVIEW

April, 1975

Published by the
Directorate of Overseas Surveys
of the
Ministry of Overseas Development
Kingston Road, Tolworth, Surrey

Annual Subscription £4 post free Single copies £1 post free

Survey Review XXII, 176, April 1975

DIRECT AND INVERSE SOLUTIONS OF GEODESICS ON THE ELLIPSOID WITH APPLICATION OF NESTED EQUATIONS

T. Vincenty

DMAAC Geodetic Survey Squadron,
F. E. Warren AFB, Wyoming 82001

ABSTRACT

This paper gives compact formulae for the direct and inverse solutions of geodesics of any length. Existing formulae have been recast for efficient programming to conserve space and reduce execution time. The main feature of the new formulae is the use of nested equations for elliptic terms. Both solutions are iterative.

1. INTRODUCTION

In selecting a formula for the solution of geodesics it is of primary importance to consider the length of the program, that is the amount of core which it will occupy in the computer along with trigonometric and other required functions. It is advantageous to have on the computer system only one direct and one inverse subroutine, both of which should give complete accuracy over lines of any length, from a few centimetres to nearly 20 000 km.

Experiments have shown that noniterative solutions, such as Bowring's inverse [2] for lines up to about 1500 km, Sodano's direct and inverse [5], or McCaw's direct as given by Rainsford [3], consume more space than the iterative solutions described in this paper and that some may even be slower in execution.

The recommended direct and inverse solutions were developed from Rainsford's inverse formula [3]. The direct solution was obtained by reversing the inverse, using the approach of Rapp [4]. Rainsford's terms in f^4 have been omitted as negligible but the most significant terms in u^8 have been retained. Certain closed equations were taken from [5].

The compactness of the recommended solutions is due to the use of nested equations to compute elliptic terms and of only three trigonometric functions: sine, cosine, and arc tangent. Nesting reduces the number of operations involving storage and retrieval of intermediate results (particularly when programming in assembly language), reduces the length of the program and the time of execution, and minimizes the possibility of underflow.

2. NOTATION

a, b, major and minor semiaxes of the ellipsoid.

f, flattening $= (a-b)/a$.

ϕ, geodetic latitude, positive north of the equator.

L, difference in longitude, positive east.

s, length of the geodesic.

α_1, α_2, azimuths of the geodesic, clockwise from north; α_2 in the direction $P_1 P_2$ produced.

α, azimuth of the geodesic at the equator.

$u^2 = \cos^2 \alpha (a^2-b^2)/b^2$.

U, reduced latitude, defined by $\tan U = (1-f) \tan \phi$.

λ, difference in longitude on an auxiliary sphere.

σ, angular distance $P_1 P_2$ on the sphere.

σ_1, angular distance on the sphere from the equator to P_1.

σ_m, angular distance on the sphere from the equator to the midpoint of the line.

3. Direct Formula

$$\tan \sigma_1 = \tan U_1/\cos \alpha_1. \tag{1}$$

$$\sin \alpha = \cos U_1 \sin \alpha_1. \tag{2}$$

$$A = 1+\frac{u^2}{16384}\{4096+u^2[-768+u^2(320-175u^2)]\}. \tag{3}$$

$$B = \frac{u^2}{1024}\{256+u^2[-128+u^2(74-47u^2)]\}. \tag{4}$$

$$2\sigma_m = 2\sigma_1+\sigma. \tag{5}$$

$$\Delta\sigma = B \sin \sigma\{\cos 2\sigma_m+\tfrac{1}{4}B[\cos \sigma(-1+2\cos^2 2\sigma_m) -\tfrac{1}{6}B \cos 2\sigma_m(-3+4\sin^2\sigma)(-3+4\cos^2 2\sigma_m)]\}. \tag{6}$$

$$\sigma = \frac{s}{bA}+\Delta\sigma. \tag{7}$$

Eq. (5), (6), and (7) are iterated until there is a negligible change in σ. The first approximation of σ is the first term of (7).

$$\tan \phi_2 = \frac{\sin U_1 \cos \sigma+\cos U_1 \sin \sigma \cos \alpha_1}{(1-f)[\sin^2 \alpha+(\sin U_1 \sin \sigma-\cos U_1 \cos \sigma \cos \alpha_1)^2]^{\frac{1}{2}}}. \tag{8}$$

$$\tan \lambda = \frac{\sin \sigma \sin \alpha_1}{\cos U_1 \cos \sigma-\sin U_1 \sin \sigma \cos \alpha_1}. \tag{9}$$

$$C = \frac{f}{16}\cos^2 \alpha[4+f(4-3\cos^2 \alpha)]. \tag{10}$$

$$L = \lambda-(1-C)f \sin \alpha\{\sigma+C \sin \sigma[\cos 2\sigma_m+C \cos \sigma(-1+2\cos^2 2\sigma_m)]\}. \tag{11}$$

$$\tan \alpha_2 = \frac{\sin \alpha}{-\sin U_1 \sin \sigma+\cos U_1 \cos \sigma \cos \alpha_1}. \tag{12}$$

If the terms in u^8 and B^3 are omitted, $\Delta\sigma$ will give a maximum error of less than 0·00005″. Therefore the following simplified equations may be used for lesser accuracy:

$$A = 1+\frac{u^2}{256}[64+u^2(-12+5u^2)]. \tag{3a}$$

4

$$B = \frac{u^2}{512}[128+u^2(-64+37u^2)]. \tag{4a}$$

$$\Delta\sigma = B\sin\sigma[\cos 2\sigma_m + \tfrac{1}{4}B\cos\sigma(-1+2\cos^2 2\sigma_m)]. \tag{6a}$$

4. Inverse Formula

$$\lambda = L \text{ (first approximation)}. \tag{13}$$

$$\sin^2\sigma = (\cos U_2 \sin\lambda)^2 + (\cos U_1 \sin U_2 - \sin U_1 \cos U_2 \cos\lambda)^2. \tag{14}$$

$$\cos\sigma = \sin U_1 \sin U_2 + \cos U_1 \cos U_2 \cos\lambda. \tag{15}$$

$$\tan\sigma = \sin\sigma/\cos\sigma. \tag{16}$$

$$\sin\alpha = \cos U_1 \cos U_2 \sin\lambda/\sin\sigma. \tag{17}$$

$$\cos 2\sigma_m = \cos\sigma - 2\sin U_1 \sin U_2/\cos^2\alpha. \tag{18}$$

λ is obtained by eqn. (10) and (11). This procedure is iterated starting with eqn. (14) until the change in λ is negligible.

$$s = bA(\sigma - \Delta\sigma), \tag{19}$$

where $\Delta\sigma$ comes from eqn. (3), (4), and (6).

$$\tan\alpha_1 = \frac{\cos U_2 \sin\lambda}{\cos U_1 \sin U_2 - \sin U_1 \cos U_2 \cos\lambda}. \tag{20}$$

$$\tan\alpha_2 = \frac{\cos U_1 \sin\lambda}{-\sin U_1 \cos U_2 + \cos U_1 \sin U_2 \cos\lambda}. \tag{21}$$

As in the direct solution, the simplified equations (3a), (4a), and (6a) may be used when a maximum error of less than 1·5 mm is acceptable.

The inverse formula may give no solution over a line between two nearly antipodal points. This will occur when λ, as computed by eqn. (11), is greater than π in absolute value.

5. Accuracy Checks

Elliptic terms have their maximum effect on angular and geodesic distances over north-south lines. Independent checks on distances were obtained by using the direct formula to compute the latitudes of forepoints in azimuths of 0°, and eqn. (1) and (2) of [7] (quoted from [1]) to compute meridional arcs corresponding to the same lines. The latitudes of standpoints were from 0° to 80° in increments of 10° and the distances were in multiples of 2000 km up to 18 000 km, which gave 81 test lines. The maximum disagreement was 0·01 mm.

Rainsford's eqn. (11) gives coefficients A_0, A_2, A_4, and A_6, to include terms in f^3, for computation of $(\lambda - L)$ by an equation corresponding to eqn. (11) of this paper. If the latter equation is rewritten in conventional form, we obtain the following approximate coefficients:

$$\left.\begin{aligned} A_0' &= 1-C. \\ A_2' &= C(1-C). \\ A_4' &= C^2(1-C)/2. \\ A_6' &= 0. \end{aligned}\right\} \quad (22)$$

The maximum errors in $(\lambda - L)$ due to omission of terms in f^3 in the A' coefficients are then given by

$$\left.\begin{aligned} \delta_0 &= (A_0' - A_0) f \sin \alpha\pi. \\ \delta_2 &= (A_2' - A_2) f \sin \alpha. \\ \delta_4 &= (A_4' - A_4) f \sin \alpha. \\ \delta_6 &= (A_6' - A_6) f \sin \alpha. \end{aligned}\right\} \quad (23)$$

These errors were computed for lines in equatorial azimuths of 2·5° to 87·5°, in increments of 2·5°. δ_0 attained a maximum of +3 in the 6th decimal of a second. The maximum value of δ_2 was −1 in the 6th decimal. The remaining errors were at most in the 8th decimal.

The values δ_0 and δ_2 are given in Table I; δ_0 was computed for $\sigma = 3{\cdot}1$ radians; δ_2 applies to lines over which $\sin \sigma \cos 2\sigma_m = 1$.

The recommended direct and reverse solutions duplicate each other perfectly if the values obtained from the previous computation are used without rounding. This was to be expected, since one formula was obtained by reversing the other. They were tested independently on five examples (a) to (e) given by Rainsford [3], using direct and inverse subroutines prepared in FORTRAN IV by this writer and the IBM 7094 (Model 1) computer of DMAAC Geodetic Survey Squadron. The programs iterate until the change in σ in the direct or λ in the inverse computation diminishes in absolute value to 10^{-12} radians or less. The first example is on the Bessel Ellipsoid and the remaining ones are on the International. The parameters are

Bessel: $a = 6377397{\cdot}155$ m, $1/f = r = 299{\cdot}1528128$

International: $a = 6378388{\cdot}000$ m, $1/f = r = 297$.

The results are listed in Table II. Columns A and B show Rainsford's published (1955) data. Column C gives the amounts (in the 5th decimal of a second) by which the results of the direct subroutine differed from those in column B when the inputs were those of column A. The last column D lists the differences in azimuths and distances (the latter in millimetres) from the published (1955) results when the inverse routine was used with ϕ_1, ϕ_2, and L as data from the published results.

The disagreements shown in columns C and D can be explained. Rainsford states that in his example (c) there is a residual error in longitude of 3 in the fifth decimal and that in example (e) the disagreements in latitude and azimuth are 2 and 1 in the fifth decimal respectively. Distances obtained from the inverse solution and rounded off to the millimetre may be in error by up to 0·5 mm, which represents

TABLE I—MAXIMUM ERRORS IN (λ–L)*

α	10°	20°	30°	40°	50°	60°	70°	80°
δ_0	0·4	0·8	1·4	2·2	2·9	2·9	1·9	0·6
δ_2	−0·1	−0·2	−0·4	−0·6	−0·9	−0·9	−0·6	−0·2

* In the sixth decimal of a second.

TABLE II—RESULTS OF SOLUTIONS

Line	A	B	C	D
	ϕ_1	ϕ_2	$\delta\phi_2$	$\delta\alpha_1$
	α_1	L	δL	$\delta\alpha_2$
	s	α_2	$\delta\alpha_2$	δs
(a)	55° 45′ 00·00000″	−33° 26′ 00·00000″	−1·2	−0·4
	96 36 08·79960	108 13 00·00000	+0·7	−0·5
	14110526·170m	137 52 22·01454	−1·2	−0·4
(b)	37 19 54·95367	26 07 42·83946	−0·7	−0·2
	95 27 59·63089	41 28 35·50729	+1·2	−0·2
	4085966·703	118 05 58·96161	+0·5	−0·4
(c)	35 16 11·24862	67 22 14·77638	−2·0	−0·2
	15 44 23·74850	137 47 28·31435	+2·9	+0·3
	8084823·839	144 55 39·92147	+3·0	−0·7
(d)	1 00 00·00000	−0 59 53·83076	−0·2	−102·9
	89 00 00·00000	179 17 48·02997	+0·6	+102·6
	19960000·000	91 00 06·11733	−0·3	−0·2
(e)	1 00 00·00000	1 01 15·18952	+2·5	+0·4
	4 59 59·99995	179 46 17·84244	−0·2	−0·8
	19780006·558	174 59 59·88481	−0·3	+0·8

0·000015″ in the direction of the line. Another source of discrepancy in azimuths is the use in the inverse solution of rounded co-ordinates which were computed in the direct solution with more precision. This will now be investigated.

After Rapp [4] and other sources we have on a sphere

$$d\alpha_1 = (\sin\alpha_1/\tan\sigma)\,dU_1 - (\sin\alpha_2/\sin\sigma)\,dU_2 + (\cos U_2\cos\alpha_2/\sin\sigma)\,d\lambda. \tag{24}$$

We may write similarly

$$d\alpha_2 = (\sin\alpha_1/\sin\sigma)\,dU_1 - (\sin\alpha_2/\tan\sigma)\,dU_2 + (\cos U_1\cos\alpha_1/\sin\sigma)\,d\lambda. \tag{25}$$

We note that the displacement of the forepoint (in arc measure) in the direction at right angle to the direction of the line is given by

$$-\sin\alpha_2\,dU_2 + \cos U_2\cos\alpha_2\,d\lambda \tag{26}$$

which, when divided by $\sin\sigma$, gives the corresponding change in α_1. This change is precisely the same for a short line of angular length σ as for a line of supplementary length $\sigma' = \pi - \sigma$. At about 10 000 km a small change in U_1 does not affect the azimuth at the standpoint, as $\tan\sigma$ becomes infinite.

Errors in computed azimuths due to rounding of co-ordinates are large for short lines and decrease progressively up to about 10 000 km, after which they start increasing. Very large errors can be expected over lines between nearly antipodal points. Conversely, the error in position of the forepoint due to an error in azimuth attains its maximum value of 0·3 mm per unit of the 5th decimal of a second at a distance of about 10 000 km. The assumptions $dU = d\phi$ and $d\lambda = dL$ are not quite correct but they give us an idea about the magnitudes of changes in azimuths due to changes in co-ordinates.

The change in distance is given by the well-known equation

$$ds = -M_1\cos\alpha_1\,d\phi_1 + M_2\cos\alpha_2\,d\phi_2 + N_2\cos\phi_2\sin\alpha_2\,dL, \tag{27}$$

in which M and N denote radii of curvature in the meridian and in the prime vertical respectively. This equation applies to lines of any length. A mean radius of the Earth may be used instead of M and N for approximate results. See [6], p. 55.

The disagreement of 0·001″ in azimuth over Rainsford's line (d) is justified

TABLE III—CHANGES IN INVERSE RESULTS DUE TO CHANGES IN CO-ORDINATES OF THE FOREPOINT

Line	$d\alpha_1/d\phi_2$	$d\alpha_1/dL$	$d\alpha_2/d\phi_2$	$d\alpha_2/dL$	$ds/d\phi_2$*	ds/dL*
(a)	$-0{\cdot}84$	$-0{\cdot}78$	$+0{\cdot}51$	$-0{\cdot}08$	$-0{\cdot}23$	$+0{\cdot}17$
(b)	$-1{\cdot}47$	$-0{\cdot}71$	$-1{\cdot}18$	$-0{\cdot}13$	$-0{\cdot}14$	$+0{\cdot}25$
(c)	$-0{\cdot}60$	$-0{\cdot}33$	$-0{\cdot}18$	$+0{\cdot}82$	$-0{\cdot}25$	$+0{\cdot}07$
(d)	$-583{\cdot}$	$-10{\cdot}3$	$+583{\cdot}$	$+10{\cdot}2$	$-0{\cdot}01$	$+0{\cdot}31$
(e)	$-1{\cdot}90$	$-21{\cdot}8$	$+1{\cdot}89$	$+21{\cdot}8$	$-0{\cdot}31$	$+0{\cdot}03$

* In mm per 0·00001″.

by the fact that this line is very sensitive to a change in latitude. The errors in distances and azimuths due to errors in ϕ_2 and L are given in Table III. They were obtained from inverse computations after changing the co-ordinates of the forepoints by small amounts.

The direct and inverse subroutines, as now written, require two to four iterations in most cases. Lines connecting nearly antipodal points may need considerably more repetitions in the inverse case. Rainsford's line (e) converged in the direct solution after four iterations but it needed 18 iterations in the inverse solution.

6. PROGRAMMING SUGGESTIONS

A useful FORTRAN IV function to evaluate arc tangent in double precision is DATAN 2(Y, X) which accepts the numerator and the denominator as arguments and gives the result between $-\pi$ and $+\pi$. This function may be used with seven equations of the above formulae.

Eqn. (1) and (18) become indeterminate over equatorial lines but this will not cause trouble, provided that division by 0 is excluded. In this case it is unimportant what values are computed by these equations, since $B = C = 0$, so that $\Delta\sigma$, L, and λ will be computed correctly.

Nested equations are designed for programming from right to left. Identical expressions, such as those found in (8) and (12) or (14) and (20), should be computed once, stored, and used as needed. Trigonometric identities may be used to obtain $\cos U$ and $\sin U$ from $\tan U$ and $\cos^2 \alpha$ from $\sin \alpha$.

If reverse azimuth is desired (i.e. the azimuth at P_2 to P_1) instead of α_2, the signs of the numerators and denominators should be reversed in eqn. (12) and (21).

ACKNOWLEDGMENTS

The author is grateful to Mr. Hume F. Rainsford for scrupulous checking of equations and numerical data given here; to Professor Richard H. Rapp for providing several test examples; and to both for helpful suggestions.

References

1. Adams, O. S. *Latitude developments connected with geodesy and cartography*, US Coast and Geodetic Survey Special Publication No. 67, 1949.
2. Bowring, B. R. The further extension of the Gauss inverse problem, *Survey Review*, No. 151, 1969.
3. Rainsford, H. F. Long geodesics on the ellipsoid, *Bull. Géod.*, No. 37, 1955.
4. Rapp, R. H. *Geometric geodesy notes*, The Ohio State Univ., 1969.
5. Sodano, E. M. General non-iterative solution of the inverse and direct geodetic problems, *Bull. Géod.*, No. 75, 1965.
6. Tobey, W. M. *Geodesy*, Geodetic Survey of Canada Publication No. 11, 1928.
7. Vincenty, T. The meridional distance problem for desk computers, *Survey Review*, No. 161, 1971.