An Algorithm for the Inverse Solution of Geodesic Sailing without Auxiliary Sphere

Wei-Kuo Tseng

(Department of Merchant Marine, National Taiwan Ocean University)

(E-mail: wilco0917@yahoo.com.tw)

An innovative algorithm to determine the inverse solution of a geodesic with the vertex or Clairaut constant located between two points on a spheroid is presented. This solution to the inverse problem will be useful for solving problems in navigation as well as geodesy. The algorithm to be described derives from a series expansion that replaces integrals for distance and longitude, while avoiding reliance on trigonometric functions. In addition, these series expansions are economical in terms of computational cost. For end points located at each side of a vertex, certain numerical difficulties arise. A finite difference method together with an innovative method of iteration that approximates Newton’s method is presented which overcomes these shortcomings encountered for nearly antipodal regions. The method provided here, which does not involve an auxiliary sphere, was aided by the Computer Algebra System (CAS) that can yield arbitrarily truncated series suitable to the users accuracy objectives and which are limited only by machine precisions.

KEYWORDS


1. INTRODUCTION. The definition of the connecting edge or line between two vertices is a minor great elliptic arc in Microsoft SQL Server (2014), whereas the more sophisticated definitions are the geodesics in the main geography databases (Oracle® Spatial Developer’s Guide, 2013; ArcGIS Resource, 2014; Hipparchus® Tutorial and Programmer’s Guide, 2004; IBM DB2 Universal Database 9.1., 2012). Great ellipse sailing has been studied in our previous work (Tseng et al., 2012a; 2013) and considering that the producers of Geographic Information Systems (GIS) and Electronic Chart Display and Information Systems (ECDIS) may adopt those databases in their systems, the geodesic algorithms need to be reviewed and studied.

The geodesic is of interest because it is the shortest path between two points on the Earth. In most terrestrial applications, the Earth is treated as a spheroid by adopting the World Geodetic System (WGS) 84 datum. Geodesics can also be
used in the application of the United Nations Convention on maritime boundaries at sea; other uses involve distance measuring in GIS and ECDIS and governing rules of the Federal Aviation Administration bounding areas (Sjöberg, 2007; 2012).

Usual algorithms for the geodesic can be roughly divided into two groups: (a) numerical integration schemes and (b) series expansion of elliptic integrals. Group (a) can be further divided into integration schemes based on simple differential relationships of the spheroid (Kivioja, 1971; Jank and Kivioja, 1980; Thomas and Featherstone, 2005), or by numerical integration of elliptic integrals that are usually functions of elements in the spheroid and its corresponding auxiliary sphere (Saito, 1970; 1979; Sjöberg, 2007; 2012). Group (b) includes the original method of Bessel (1826) that uses functions of elements in the spheroid related to a corresponding auxiliary sphere and various modifications to his method (Rainsford, 1955; Vincenty, 1975a; Bowring, 1983; 1984; Karney, 2013).

The inverse geodetic problem on the spheroid is to determine the geodesic arc length between two endpoints and the azimuths of the arc. The more complete solution for the Clairaut constant (or the vertex latitude) which is compared with the solution provided by Sjöberg (2007; 2012) is presented in this paper. If the two given points are not nearly antipodal, each azimuth and location of the geodesic is unique, while for the fixed points in the “nearly antipodal regions” and when the sum of the latitudes of two are equal to zero, there are two geodesics mirrored in the Equator and with complementary azimuths at each point. In the special case when the endpoints are located at the poles of the spheroid, all meridians are geodesics (Sjöberg, 2012). The special role played by the change of variable, the series integration in terms of CAS, finite difference method to obtain the vertices’ latitude of a geodesic and an innovative iterative method proposed to replace Newton’s method implemented by Karney (2013) makes this method different from others currently available.

The article is organised as follows: In Section 2, the basic equations needed and series solution for longitude in terms of latitude on a geodesic with specific phases to determine the actual geodesic are presented; Section 3 derives the series for the geodesic’s arc length. Section 4 treats the inverse problem in the general cases and discusses some special issues relevant to the special cases, and at the end of Section 4, the flow chart of the algorithm is provided for the reader’s convenience. Section 5 gives numerical examples with various conditions including the general case, the special case of geodesics passing two endpoints on nearly antipodal regions. The conclusions are summarised in Section 6.

2. FORMULA FOR LONGITUDE DIFFERENCE IN TERMS OF REDUCED LATITUDE ON A GEODESIC. Geodesic calculations all involve the Clairaut constant or latitude of the vertex along a geodesic, the relationship between the azimuth \( \alpha \) and the reduced latitude \( \beta \) can be expressed by the relation (Clairaut, 1735).

\[
c = \cos \beta_V = \cos \beta \sin a
\]  

where \( e \) is the first eccentricity, \( a, b \) is the semi-major and semi-minor axes, \( c \) is the Clairaut constant of a geodesic, \( \beta_V \) is the latitude of vertex along the geodesic, and \( \beta \) is
given by the relation shown below:

$$\tan(\beta) = \sqrt{1 - e^2} \tan(\varphi),$$

(2)

where $\varphi$ is geodetic latitude.

As mentioned, it is desirable to have a succinct theoretical statement of the solutions to the geodesic, covering all cases in hand before developing these problems' numerical solutions. Referring to Figure 1, it can be seen that the inverse problem is equivalent to solving the geodesic triangle $\Delta NP_1P_2$ given two sides and their longitude difference ($\Delta \lambda$). This can be done in more than one way. By using the reduced latitude as the variable of integration such as those presented in Jank and Kivioja (1980) give the longitude and distance integrals as shown in Equations (3) and (4). If the Clairaut constant or latitude of the vertex and one point are given, then the latitude and longitude of any point $P$ of the geodesic are related by Equation (3).

$$\lambda(\beta) = \omega + c \sum_{k=1}^{\infty} C_k^{1/2} (-e^2)^k \int \frac{\cos^{2k-1} \beta}{\cos^2 \beta - c^2} \cos \beta d\beta$$

(3)

The latitude and distance are related by Equation (4)

$$S(\beta) = a \int \frac{1 - e^2 \cos^2 \beta}{\cos^2 \beta - c^2} \cos \beta d\beta$$

(4)

where the semi-major axis $a$ is 6378137 and eccentricity $e$ is 0.0818191908426215 as adopted in the current WGS-84 ellipsoid model for the Earth.

Applying binomial theorem, Equation (1) and integral techniques expands Equation (3) to the expansion:

$$\lambda(\beta) = \omega + c \sum_{k=1}^{\infty} C_k^{1/2} (-e^2)^k \int \frac{\cos^{2k-1} \beta}{\cos^2 \beta - c^2} \cos \beta d\beta$$

(5)

where

$$\omega = \arcsin \left( \frac{\tan \beta}{\tan \beta_V} \right) = \int \frac{c}{\sqrt{\cos^2 \beta - c^2 \cos \beta}} d\beta, \quad \text{and} \quad \tan \beta_V = \sqrt{1 - e^2 / c}.$$
The vertex latitude $\beta_V$ is always positive here. Equation (5) gives the longitude formula for the geodesic, and when the eccentricity equals zero, this longitude formula Equation (5) is the same as Napier’s rules. Further using the change of variable gives the simplified integral as:

$$\lambda(\beta_1, \beta_2) = \omega_{12} + c \sum_{k=1}^{\infty} C_k^{1/2} (-e^2)^k \int_{\sin \beta_1}^{\sin \beta_2} \frac{(1 - x^2)^{k-1}}{\sqrt{1 - e^2 - x^2}} dx$$

(6)

where

$$x = \sin \beta, \quad \omega_{12} = \arcsin \left( \frac{xc}{\sqrt{1 - x^2} \sqrt{1 - e^2}} \right), \quad \text{and} \quad -\beta_V \leq \beta \leq \beta_V.$$

By applying the binomial theorem again this expands the numerator in the right-hand side integrand of Equation (6) and this result can be exactly integrated as the following:

$$\lambda(\beta_1, \beta_2) = \omega_{12} + c \sum_{k=1}^{\infty} (M_{k,1} \cdot \sigma + M_{k,2} \cdot \mu) x^{2k}$$

(7)

where $u = \sqrt{1 - e^2 - x^2}$ and $\sigma = \arctan(u)$.

The expansion of Equation (6) can be conveniently integrated to arbitrarily order by CAS such as Maple® etc. which yields the two coefficient matrices by truncating the expansion at order $e^{10}$ up to $M_4$, as shown below:

$$M_{*,1} = \begin{bmatrix} -\frac{1}{2} & 0 \\ -\frac{1}{16} - \frac{1}{16} e^2 \\ -\frac{3}{128} - \frac{1}{64} e^2 - \frac{3}{128} e^4 \\ -\frac{25}{2048} - \frac{15}{2048} e^2 - \frac{15}{2048} e^4 - \frac{25}{2048} e^6 \end{bmatrix}$$

(8)

and

$$M_{*,2} = \begin{bmatrix} 0 \\ -\frac{1}{16} x \\ -\frac{5}{128} x - \frac{3}{128} x e^2 + \frac{1}{64} x^3 \\ -\frac{55}{2048} x - \frac{5}{256} x e^2 - \frac{25}{2048} x e^4 + \frac{65}{3072} x^3 + \frac{25}{3072} x^3 e^2 - \frac{5}{768} x^5 \end{bmatrix}$$

(9)

where $M_{*,i} = [M_{1,i} \ldots M_{4,i}]^T$ and $i = 1, 2$.

The latitudes between interval $[-\beta_V, \beta_V]$ cannot sketch all the geodesic segments of arbitrary length, so to overcome this disadvantage, one must introduce another variable $t$ with interval $[-\infty, \infty]$ and also imply the following properties of
Equation (10) which is related to latitudes.

\[ \beta(t) = (-1)^p \Delta t + (-1)^{p+1} \beta_V \]  

where \( p = \text{FIX}(t + \beta_V, 2\beta_V) \) is the number of phase \((2\beta_V)\) \([\text{FIX}(..)\) rounds the element of (..) to the nearest integer\] and \( \Delta t = \text{REM}(t + \beta_V, 2\beta_V) \) is the fraction of variable \( t \) in regard to the phase \((\text{REM} \text{ computes the remainder after division})\).

As the relationship between variables \( t \) and latitude is many-to-one, Equation (10) is a periodical function that oscillates between two opposite vertices as shown in Figure 2.

The integral must be transformed into the following conditional function modified to accommodate variable \( t \). But because although the range of variable \( t \) is \([-\infty, \infty]\), when integrating Equation (7) it will only give one corresponding value in the range of \([-\beta_V, \beta_V]\), thus to solve this problem, one must first determine how many phases \((\beta_V)\) does the variable \( t \) have in regard to Equation (7). If even phases, then use the conditional function Equation (11) (top left rectangle in Figure 2), whereas if it equals odd phases, then use the other conditional (bottom left rectangle in Figure 2).

\[
\lambda(0,t) = 2p\Delta\lambda_V + \lambda(0,(-1)^p\beta(t))
\]

i.e.

\[
\lambda(0,t) = \begin{cases} 
2p\Delta\lambda_V + \lambda(0,\beta(t)), & p \in 2n \\
2p\Delta\lambda_V + \lambda(\beta(t),0), & p \in 2n+1 
\end{cases}
\]  

where the longitude difference \( \Delta\lambda_V = \lambda(0, \beta_V) \) of one phase equals one quarter cycle wrapping the Equator and the latitude \( \beta(t) \) defined in Equation (10).

Again when variable \( t \) is applied to Equation (1), it will only give one corresponding value in the range of \([-\beta_V, \beta_V]\). We solve this problem by first determining how many phases \((2\beta_V)\) does variable \( t \) have in regard to \(2\beta_V\). If the phases of variable \( t \) are even phases, then use the first condition of Equation (12) (top right rectangle in Figure 2), whereas then otherwise use the second condition of Equation (12) (bottom right
rectangle in Figure 2), using Equation (1) gives the course function below on waypoints along the geodesic.

\[ \alpha(t) = \pm \left[ \MOD(p, 2) \cdot \pi + (-1)^p \sin^{-1}(\cos \beta V / \cos \beta(t)) \right] \]

i.e.

\[ \alpha(t) = \pm \begin{cases} 
\sin^{-1}(\cos \beta V / \cos \beta(t)), & p = 2n \\
\pi - \sin^{-1}(\cos \beta V / \cos \beta(t)), & p = 2n + 1
\end{cases} \]  

(12)

If the geodesic is westbound, the positive sign (+) is adopted, otherwise the negative sign (−) is adopted.

The two conditions of Equation (12) mean that the courses are northeast bound or southeast bound as shown in Figure 2, because the default travel direction in the canonical configurations is eastbound.

Apply Equation (11) to calculate the longitudes of the path departing from the first node to the fifth node (passing 2 cycles) where \( t \) has the interval \([0°, 720°]\).

Figure 3 shows the oscillations of a geodesic on the spheroid and a great circle on the sphere. Point F on the geodesic crosses the Equator at longitude 0, heading in a northeast direction reaching a maximum northern latitude, then descends in a southeast direction crossing the Equator at a certain longitude, reaching a maximum southern latitude, then ascends in a northeastern direction crossing the Equator again at point F1 (the red line). This is one cycle of the geodesic, but the longitude difference of the end of one cycle does not equal 360° due to the eccentricity of the spheroid and is not a closed curve, hence the geodesic does not repeat after a complete cycle, unlike its great circle counterpart which repeats after a complete cycle at point F2 (the brown line).

The geodesic curve will lag a certain longitude compared to a great circle after a complete cycle. Therefore after passing many circuits, it will wrap a band on the spheroid between two opposite vertices shown in Figure 4.
Applying Equation (11) gives the longitude difference between the south and north vertices:

\[ 2\Delta \lambda_{V} = 2\lambda(0, \beta_{V}) = \pi \left( 1 + c \sum_{k=1}^{\infty} M_{2k,1} e^{2k} \right) \]

\[ \approx \pi \left( 1 - \frac{1}{2} e^{2} c - \left( \frac{1}{16} c + \frac{1}{16} c^{3} \right) e^{4} - \left( \frac{3}{128} c + \frac{1}{64} c^{3} + \frac{3}{128} c^{5} \right) e^{6} \right. \]

\[ \left. - \left( \frac{25}{2048} c + \frac{15}{2048} c^{3} + \frac{15}{2048} c^{5} + \frac{25}{2048} c^{7} \right) e^{8} + O(e^{10}) \right) \]  

(13)

where \( x = \sin(\beta_{V}) = \sqrt{1 - c^{2}} \).

The nearly antipodal regions are located within the interval \([\Delta \lambda_{V}, \pi]\); there are two geodesics mirrored in the Equator and with complementary azimuths at each point (Figure 5). As stated before, in the special case when the given points are located at the poles of the spheroid, all meridians are geodesics (Sjöberg, 2012).
The track of the two mirrored geodesics for various starting latitudes can be described by the definition of variable $t$ shown below:

$$ t_1 = \begin{cases} \beta_1, & a_1 \leq \frac{\pi}{2}, \\ \text{sign}(\beta_1)2\beta_V - \beta_1, & a_1 > \frac{\pi}{2}, \end{cases} \quad \beta_V \geq \beta_1, \quad (14) $$

and the longitudes and latitudes can be obtained by Equation (15):

$$ \Delta \lambda = \lambda(t) - \lambda(t_1), \quad t = [t_1, t_1 + 2\beta_V] \quad (15) $$

Using Equation (13) gives the nearly antipodal regions for various vertices shown in Figure 6.

Figure 6 shows that the longitude difference between the two opposite vertices depends on the vertex. It equals $\pi$ only when the constant $c = 0$ and is slightly shorter than $\pi$. The geodesic is not closed around the spheroid for $c \neq 0$, and its period is somewhat shorter than $2\pi$. Applying integration by parts gives a different integral for Equation (3):

$$ \lambda(\beta) = \omega \sqrt{1 - e^2 \cos^2 \beta} - \int \omega \frac{e^2 \sin \beta \cos \beta}{\sqrt{1 - e^2 \cos^2 \beta}} d\beta \quad (16) $$

The longitude difference between the two opposite vertices has a minimum value when $c$ equals one. As the latitude of the vertex approaches to zero, the LHS of integral Equation (16) vanishes. The exact formula of the longitude difference between two opposite vertices can be given or setting $c=1$ in Equation (13) also gives the approximate series Equation (17).

$$ 2\Delta \lambda_V = \lambda(\beta_1 = 0, \beta_2 \to 0, 1) = \pi \sqrt{1 - e^2} \approx \pi \left(1 - \frac{1}{2} e^2 - \frac{1}{8} e^4 + \frac{1}{16} e^6 - \frac{5}{128} e^8 - \ldots\right) \quad (17) $$

<table>
<thead>
<tr>
<th>A.R.</th>
<th>Vertex($\beta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>36.21</td>
<td>0</td>
</tr>
<tr>
<td>35.66</td>
<td>10</td>
</tr>
<tr>
<td>34.02</td>
<td>20</td>
</tr>
<tr>
<td>31.35</td>
<td>30</td>
</tr>
<tr>
<td>27.73</td>
<td>40</td>
</tr>
<tr>
<td>23.26</td>
<td>50</td>
</tr>
<tr>
<td>18.09</td>
<td>60</td>
</tr>
<tr>
<td>12.38</td>
<td>70</td>
</tr>
<tr>
<td>6.28</td>
<td>80</td>
</tr>
<tr>
<td>0.00</td>
<td>90</td>
</tr>
</tbody>
</table>
The binomial series of the second last term of Equation (17) with respect to $e^2$ is approximately the last term of the above. This is the so-called “lift-off longitude” (Rapp, 1993). The antipodal region (Schmidt, 2006) is located roughly within 36.2′ from the antipode at the Equator adopting the WGS-84 Earth datum.

3. ARC LENGTH IN TERMS OF REDUCED LATITUDE ON A GEODESIC. Geodesic distances are very important as they are the shortest distances. Applying the binomial theorem, Equation (1), and integral techniques expands Equation (4) to the series Equation (18):

$$S(\beta) = a \cdot \sigma + a \sum_{k=1}^{\infty} C_k^{1/2} (-e^2)^k \int \frac{\cos^{2k+1} \beta}{\sqrt{\cos^2 \beta - c^2}} d\beta$$

where

$$\sigma = \arctan \left( \frac{\sin \beta}{\sqrt{\cos^2 \beta - c^2}} \right) = \int \frac{\cos \beta}{\sqrt{\cos^2 \beta - c^2}} d\beta.$$

Setting $x = \sin \beta$, thus using the change of variable transforms Equation (18) to Equation (19).

$$S(\beta_1, \beta_2) = a \cdot \sigma + a \sum_{k=1}^{\infty} C_k^{1/2} (-e^2)^k \int_{\sin \beta_1}^{\sin \beta_2} \frac{(1-x^2)^k}{\sqrt{1-c^2-x^2}} dx$$

Again, by using the binomial theorem, the numerator in the right-hand side integrand of Equation (19) may be expanded, and this result can be exactly integrated as Equation (20):

$$S(\beta_1, \beta_2) = a \cdot \sigma + a \sum_{k=1}^{\infty} (N_{k,1} \cdot \sigma + N_{k,2} \cdot \mu)e^{2k} \left| \frac{\sin \beta_2}{\sin \beta_1} \right|$$

where $u = \sqrt{1-c^2-x^2}$ and $\sigma = \arctan(u)$, and $\beta_V \leq \beta \leq -\beta_V$.

Integrating Equation (19) in terms of CAS or handwork gives the two matrices of coefficients by truncating the expansion at order $e^{10}$ and $N_5$ up to $e^8$ and $N_4$.

$$N_{*,1} = \begin{bmatrix}
-\frac{1}{4} & -\frac{1}{4} & c^2 \\
-\frac{3}{64} & -\frac{1}{32} & -\frac{3}{64} & c^4 \\
-\frac{5}{256} & -\frac{3}{256} & -\frac{3}{256} & -\frac{5}{256} & c^6 \\
-\frac{175}{16384} & -\frac{25}{4096} & -\frac{45}{8192} & -\frac{25}{4096} & -\frac{175}{16384} & c^8
\end{bmatrix}$$

(21)
and

\[
N_{i,2} = \begin{bmatrix}
-\frac{1}{4}x \\
-\frac{5}{64}x - \frac{3}{64}x^2 + \frac{1}{32}x^3 \\
-\frac{11}{256}x - \frac{1}{32}x^2 - \frac{5}{256}x^3 + \frac{13}{384}x^4 + \frac{5}{384}x^3c^2 - \frac{1}{96}x^5 \\
-\frac{465}{16384}x - \frac{365}{16384}x^2 - \frac{275}{16384}x^3 + \frac{175}{16384}x^4 - \frac{815}{24576}x^5 + \ldots \\
-\frac{75}{4096}x^2 + \frac{175}{24576}x^3 - \frac{125}{6144}x^5 - \frac{35}{6144}x^5c^2 + \frac{5}{1024}x^7 \\
\end{bmatrix}
\]

(22)

where \(N_{i,j} = [N_{i,1} \ldots N_{i,4}]^T\) and \(i = 1, 2\).

Again, since the latitudes between the interval \([-\beta_V, \beta_V]\) cannot sketch all the geodesic segments of arbitrarily length, one must introduce the variable \(t\) as presented in Equation (10). The integral Equation (20) must be transformed into the following conditional function modified to accommodate variable \(t\).

\[
S(0, t) = 2p\Delta S_V + S(0, \beta V, \beta V) \quad (23)
\]

where \(\Delta S_V = S(0, \beta V)\) is the distance from node to vertex as Equation (24) and the \(p\) as defined by Equation (11).

\[
\Delta S_V = S(0, \beta V) = \frac{a\pi}{2} \left(1 + \sum_{k=1}^{\infty} N_{2i,1} \cdot e^{2k}\right) \approx a\pi\left(1 - \left(\frac{1}{4} + \frac{1}{4}c^2\right)\right)e^2 - \left(\frac{3}{64} + \frac{1}{32}c^2 + \frac{3}{64}c^4\right)e^4 \ldots
\]

(24)

The arc length of a geodesic is the meridian when the vertex is at the pole(s). This shows that the arc length between the two opposite vertices or two nodes depends on the Clairaut constant, and it shall equal the meridional distance when the constant \(c = 0\). Applying integration by parts gives a different integral Equation (25) for Equation (3).

\[
S(\beta) = a\sigma\sqrt{1 - e^2\cos\beta} - a\int \sigma \frac{e^2\sin\beta\cos\beta}{\sqrt{1 - e^2\cos^2\beta}} d\beta
\]

(25)

The geodesic arc length between the two opposite vertices has a minimum value when \(c = 1\). As before, as the latitude of vertex approaches zero, the LSH integral of Equation (25) vanishes, thus the exact formula of the curve length can be given or the setting \(c = 1\) of Equation (24) gives the series approximately.

\[
2S(0, \beta V, 0, \beta V) = a\pi(1 - f) \\
\approx \pi a\left(1 - \frac{1}{2}e^2 - \frac{1}{8}e^4 - \frac{1}{16}e^6 - \frac{5}{128}e^8 - \ldots\right)
\]

(26)

In the case of when \(c = 1\), the geodesic runs along the equator, Equation (26) is consistent with Equation (17). Applying Equations (17) or (26) gives the solution of...
antipodal region at the Equator.

\[ e = \pi \left( 1 - \sqrt{1 - e^2} \right) = \pi f. \]  

(27)

We apply the parameters according to WGS 84 to obtain the maximum distance \( ae \times 6718197165 \) km along the geodesic on the Equator which is less than one half circle of the Equator. The great ellipse or the normal section passing antipodes located on the Equator cannot be determined, so the track along the Equator may be chosen for convenience. This means that the maximum distance difference between a geodesic and a great ellipse (or normal section) is the distance of one half circle of the Equator minus the length of one half elliptic arc length of the meridian as shown in Equation (28).

\[ a_\pi - 2S(0, \pi/2, e = 0) = 20037508.343 - 20003931.435 \]

= \(-33576.908\) m = \(18.1300799\) nautical miles.  

(28)

In our previous work (Tseng et al., 2013), the maximum error for a longitude difference of \(90^\circ\) from the Equator to a certain latitude is about 7 metres at approximately \(45^\circ\) latitude. The results mean that nearly antipodal points at lower latitudes will generate bigger errors between the great ellipse and geodesic.

4. INVERSE PROBLEM OF GEODESIC SAILING. The inverse problem is intrinsically more complicated than the direct problem because the given longitude \(\lambda_{12}\) is related to the corresponding unknown equatorial azimuth \(\alpha_0\) or latitude \(\beta_V\) of the geodesic vertex. Thus, the inverse problem inevitably becomes a root-finding exercise. This problem can be tackled as follows.

Assume that \(\beta_V\) is known. Solve the hybrid geodesic problem: given \(\beta_1, \beta_2,\) and \(\beta_V,\) find the calculated \(\lambda_{12}\) corresponding to the first intersection of the geodesic with the parallel of latitude \(\beta_2\) which is the resulting longitude difference of given initial \(\lambda_{12}\) in general cases; so adjust \(\beta_V\) using Newton’s method, secant method, or other root-finding methods until the correct \(\lambda_{12}\) is obtained. The two endpoints can be put in a canonical configuration,

\[ \beta_1 \leq 0, \quad \beta_1 \leq \beta_2 < -\beta_1, \quad 0 \leq \beta_{12} \leq \pi. \]  

(29)

This may be accomplished by 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 \(\beta_V = [|\beta_1|, \pi/2]\) and \(\alpha_1 = [0, \pi]\) intersect the parallel of latitude \(\beta_2\) with \(\lambda_{12} = [0, \pi].\) Meridional (\(\lambda_{12} = 0\) or \(\pi\)) and equatorial (\(\beta_1 = \beta_2 = 0,\) and \(\lambda_{12} = (1 - f)\pi\)) geodesics are treated as special cases. The general case is solved by the analog of Newton’s method as outlined below. The longitude difference of a geodesic depends on its initial azimuth as northeast or southeast shown as the conditional Equation (30) that is constrained by Condition (29). If the course is southeast bound, then the reference initial point is near the destination and the geodesic passes the vertex and the variable \(t_1\) is set to \(-2\beta_V - \beta_1\). Otherwise, the geodesic does not pass the vertex; the variable \(t_1\) is set to \(\beta_1.\)

\[ \lambda(t_1, t_2) = t_2 = \beta_2, \quad t_1 = -2NS \cdot \beta_V + (-1)^{NS} \beta_1 \]  

(30)

where the direction logical variable \(NS\) is given in Equation (32).
The geodesic with vertex latitude $\beta_1$ departing from latitude $\beta_1$ intersects the parallel of the latitude $\beta_2$ at the longitude difference $\lambda_{NS}$ as shown in Equation (31):

$$\lambda_{NS} = \lambda(\beta_1, \beta_2, c = \cos(\beta_1))$$  \hspace{1cm} (31)

The following conditions give the azimuth of a geodesic. If longitude difference $\lambda_{12}$ is greater than the longitude difference $\lambda_{NS}$ of the intersection, the initial course of the geodesic is southeast bound, otherwise it is northeast bound as shown in Figure 5. The logical variable $NS$ is decided by Equation (32):

$$NS = \begin{cases} 1, & (\lambda_{12} - \Delta\lambda_{NS}) > 0 \\ 0, & (\lambda_{12} - \Delta\lambda_{NS}) < 0 \end{cases}$$  \hspace{1cm} (32)

Vincenty (1975a) used the iterative method of Helmert (1880) to solve the inverse problem and was aware of its failure to converge for nearly antipodal regions. In an unpublished report (Vincenty, 1975b) gives a modification of his method which deals with this case. Unfortunately, this method sometimes requires many thousands of iterations to converge, whereas the approach shown below is used to approximate Newton’s method and the innovative iterative method proposed here for nearly antipodal regions only requires a few iterations.

Karney (2013) uses Newton’s method, requires a good starting guess and uses more complex procedures which even involves firstly calculating the reduced length of the geodesic and solving a fourth-order polynomial merely to find a good starting point, and may involve a high computational cost. Using Newton’s Method solves the inverse problem by the derivative of longitude difference with respect to $c$ or $\beta_V$ as shown:

$$c = c - \frac{\lambda(t_1, t_2) - \lambda_{12}}{\partial\lambda(t_1, t_2) / \partial c}, \quad \beta_V = \beta_V - \frac{\lambda(t_1, t_2, c = \cos(\beta_V)) - \lambda_{12}}{\partial\lambda(t_1, t_2) / \partial \beta_V}$$  \hspace{1cm} (33)

Differentiating the longitude difference function with respect to $c$ gives

$$\frac{\partial\lambda_{12}}{\partial c} = \begin{cases} \frac{\partial\lambda(\beta_1, \beta_2)}{\partial c}, & NS = 0 \\ \frac{\partial\lambda(-\beta_V, \beta_V)}{\partial c} + \frac{\partial\lambda(-\beta_1, \beta_2)}{\partial c}, & NS = 1 \end{cases}$$  \hspace{1cm} (34)

where the derivative of the function with respect to $c$ in-between the two opposite vertices is

$$\frac{\partial\lambda(\beta_1', \beta_2')}{\partial c} = \int_{\beta_1}^{\beta_2'} \left( \frac{\sqrt{1 - e^2 \cos^2 \beta}}{\sqrt{\cos^2 \beta - c^2 \cos \beta}} + \frac{c^2 \sqrt{1 - e^2 \cos^2 \beta} \left( \cos^2 \beta - c^2 \right)^{3/2}}{(\cos^2 \beta - c^2)^{3/2} \cos \beta} \right) d\beta$$  \hspace{1cm} (35)

If the segment of the geodesic passes the vertex, the derivative of longitude difference function with respect to $c$ or $\beta_V$ involves the segment of geodesic from the south vertex to the north vertex, which is non-differentiable as shown below.

$$\frac{\partial\lambda(-\beta_V, \beta_V, c)}{\partial c} = \infty$$  \hspace{1cm} (36)
By using the finite difference method, the derivative can be completely replaced in all situations to avoid the non-differentiable condition as the following Equation (37).

\[
\frac{\partial \lambda}{\partial \beta_V} \approx \frac{\Delta h[\lambda](\beta_V)}{h} \approx \lim_{h\to 0} \frac{\lambda(\beta_V + h) - \lambda(\beta_V)}{h}
\] (37)

In the general case, the Clairaut constant and the reduced latitude of the vertex can be represented in terms of longitude difference and latitudes of the two given endpoints.

The starting point for Newton’s method can be approximately obtained by the spherical model. In rectangular coordinates, any point on the unitary sphere can be represented as coordinates of

\[
\vec{P} = (\cos \beta \cos \omega, \cos \beta \sin \omega, \sin \beta)
\] (38)

where \( \omega \) can be replaced with \( \lambda \) approximately for the first iteration.

The normal to the plane of a great circle is the cross product of the two vectors as shown:

\[
\vec{N} = [x_N \ y_N \ z_N] = \vec{P}_1 \times \vec{P}_2
\] (39)

The latitude of the vertex equals to the angle between the \( Z \)-axis and the normal to the plane spanned by the two vectors.

\[
\tan \beta_V = \sqrt{\lambda_N^2 + y_N^2 / z_N}
\] (40)

Using the algebra symbolic system and the trigonometric identities expands the last expression as Equation (41). Sjöberg (2007; 2012) has also proven Equation (41) by spherical trigonometric methods.

\[
\tan \beta_V = \sqrt{\tan^2 \beta_1 + \tan^2 \beta_2 - 2 \tan \beta_1 \tan \beta_2 \cos \lambda_{12} / \sin \lambda_{12}}
\] (41)

Using trigonometric identity rewrites the formula for the Clairaut constant \( c \) as Equation (42).

\[
c = \cos \beta_V = \sin \lambda_{12} / \sqrt{\tan^2 \beta_1 + \tan^2 \beta_2 - 2 \tan \beta_1 \tan \beta_2 \cos \lambda_{12} + \sin \lambda_{12}}
\] (42)

If the two endpoints are located on the equator (\( \beta_1 = \beta_2 = 0 \)), then by applying Equations (13) and (17), the various cases for the Clairaut constant \( c \) can be determined by the following conditional Equation (43):

\[
c = \begin{cases} 
1, & \lambda_{12} \leq \pi \sqrt{1 - e^2} \\
0, & \lambda_{12} = \pi \\
\cos(\beta_V), & \pi \sqrt{1 - e^2} < \lambda_{12} \leq \pi
\end{cases}
\] (43)

If the two endpoints are located in the nearly antipodal regions (\( \beta_1 + \beta_2 = 0 \) and \( 2\Delta \lambda_V \leq \lambda_{12} \)), then applying Equations (13) and (17), the various cases for the Clairaut constant \( c \) can be determined by the following conditional Equation (44):

\[
c = \begin{cases} 
0, & \lambda_{12} \leq \pi \\
\cos(\beta_V), & 2\Delta \lambda_V < \lambda_{12}
\end{cases}
\] (44)
where the vertex reduced latitude ($\beta_V$) is subject to Equation (13) as the following repetitively:

$$1 - \frac{\lambda_{12}}{\pi} = -c \sum_{k=1}^{\infty} M_{2k,1} \cdot e^{2k}$$  (45)

Truncating the series toward the first order gives the approximate Clairaut constant $c$ as shown in Equation (46):

$$c_k = \frac{2(\pi - \lambda_{12})}{e^2 \pi}$$  (46)

The accurate value can be re-approached several times by the following iterative computation Equation (47) or Newton’s method:

$$c_{k+1} = \frac{1 - \Delta \lambda}{\pi} \left( c = c_k \right)$$  (47)

Equation (47) shows that the solution to the Clairaut constant has unique values for the two ends in the nearly antipodal regions.

Applying the method of the asteroid plane provided by Karney (2013) can accelerate the speed of convergence. But the faster method implies a higher computational cost to solve the quartic equation before starting the iteration. Once the Clairaut constant and the coordinates of the end points of the geodesic are known, the azimuths of the geodesic at the end points may be determined from Equation (48)

$$\alpha_1 = NS \cdot \pi + (-1)^{NS} \sin^{-1}(\cos \beta_V / \cos \beta_1)$$  (48)

and

$$\alpha_2 = \cos \beta_V / \cos \beta_2$$  (49)

Procedures for solving the geodesic inverse problem are summarised in Figure 7, which gives all the required algorithms clearly for the readers’ convenience.

5. NUMERICAL RESULTS FOR DISTANCE AND LONGITUDE EQUATIONS. Three examples are represented in this section. The first two samples are cited from Tables 4 and 5 of Karney (2013). Those data obtained from his results are compared with the results computed by the methods proposed here, which are also compared with the results obtained from the great ellipse (GE) provided in our previous work (Tseng et al., 2012a). Analysis of the efficiency and accuracy between such comparisons are shown in Tables 2 and 3. It can be seen that the innovative iterative method proposed here for nearly antipodal regions or general cases only requires a few iterations.

The first example is specified by $\phi_1 = -30\cdot12345^\circ$, $\phi_2 = -30\cdot12344^\circ$, and $\lambda_{12} = 0.00005^\circ$ as shown in Figure 8 (left). Because the two points are not nearly antipodal, an initial guess for the reduced latitude $\beta_V$ of the vertex is found from Equation (41). However, in cases of when two points are very close to each other, the line is short
enough that the error in $\beta_V$ is negligible; the solution of this inverse problem is completed by using finite difference or Newton’s method as shown in Table 1.

As the two points are very close to each other, the line is very short and the differences between the geodesic and GE are very small; both sailing methods may be applied here, even using rhumbline or plane sailing (Tseng et al., 2012b) gives near identical results (shown in Table 1).

A second example is specified by $\phi_1 = -30^\circ$, $\phi_2 = 29.9^\circ$, and $\lambda_{12} = 179.8^\circ$ in Figure 8 (right). In this case, the points are nearly antipodal, an initial guess for the Clairaut constant $c$ is found by solving Equation (46), then the solution of this problem is
completed by using finite difference or Newton’s method, the result is shown in Table 2.

The function $[\text{dist,az}] = \text{distance(lat1,lon1,lat2,lon2,ellipsoid)}$ provided in the Mapping Toolbox® by MATLAB® computes the geodesic distance and azimuth assuming that the points lie on the reference ellipsoid defined by the input ellipsoid. This distance function of MATLAB adopts Vincenty’s algorithm which sometimes requires many thousands of iterations or converges to a wrong answer in nearly antipodal regions, whereas the approximated Newton’s method and innovative iterative method for nearly antipodal regions as described here only requires a few iterations. As shown in Table 2, Karney (2013) uses much more complex procedures which even involve calculating the reduced length of the geodesic and solving a fourth-order polynomial to merely find a good starting point for his method.

In the case of closely near antipodal regions, using the MATLAB® distance function for this problem gives the incorrect result; the initial course and distance have huge errors and cannot be accepted. The discrepancies between the results computed by the proposed methods described here and of those provided by Karney (2013) are negligible; the distance difference between the two is approximately 2.5 cm for this very long geodesic. These results show that the new methods proposed here can be accepted in practical sailing and geodesic calculations. The discrepancies have arisen from truncating the expansion of Equations (7) and (10) both at order $e^{10}$ leaving the series up to $e^8$, the truncated series generates errors in $ae^8$ order which are centimetre-level.

As shown in Table 3, the GE distance is 10812·32 m (5·9 nautical miles) further than the distance of geodesic with the same input values. Fortunately in most cases of navigation, the endpoints of practical sailing are seldom located closely in nearly antipodal regions.

The third numerical example presented has a very long navigational route with longitude differences between two endpoints of about 145°. The route is from Valparaiso-Chile (32° 59·998′ S, 71° 36·675′ W) to Yokohama-Japan (34° 26·178′ N, 139° 51·39 E).

The results of Example 3 are shown in Table 3. The geodesic is slightly more curved than the great ellipse (see Figure 9). The eccentricity is exaggerated and set to 0·5 to show the difference between the GE and geodesic. In this example the discrepancy between the two distances (GE: 9242·561583 nautical miles, geodesic: 9242·558019

---

Figure 8. The track (left) of geodesic between very closely points. The tracks (right) of geodesic and great ellipse from two nearly antipodal points.
nautical miles) is approximately 6.6 metres, for which the GE can be accepted for the practical purposes of navigation. As the track doesn’t pass through the vertex, the differences of the two tracks are very small which can be negligible in practice. Also, the geodesic distance obtained from the truncated series provided here generates tiny discrepancies compared to those obtained from the different truncated series terms of the in-built function “geodesicinv.m” in MATLAB™; both generate near consistent results.

Table 1. Comparing the results of calculations with two very close endpoints.

<table>
<thead>
<tr>
<th>φ₁</th>
<th>β₁</th>
<th>λ₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ₂</td>
<td>β₂</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>a₁</td>
<td>a₂</td>
</tr>
<tr>
<td>Iteration.3</td>
<td>77·04355157491</td>
<td>37·70352648164</td>
</tr>
<tr>
<td>Karny</td>
<td>77·04355345237</td>
<td>77·04350844913</td>
</tr>
<tr>
<td>Geo-K</td>
<td>-0·0000180325</td>
<td>-0·0000180325</td>
</tr>
<tr>
<td>GE</td>
<td>77·0435347354</td>
<td>77·04350856022</td>
</tr>
<tr>
<td>Geo-GE</td>
<td>0·00001810377</td>
<td>0·00001810377</td>
</tr>
<tr>
<td>Rhumb</td>
<td>77·04352097777</td>
<td>77·04352097777</td>
</tr>
<tr>
<td>Geo-Rh</td>
<td>0·00001259714</td>
<td>0·00001259714</td>
</tr>
</tbody>
</table>

Note: Geo-K = differences between the result calculated here and Karney’s result

Table 2. Comparing the results of calculations for Example 2.

<table>
<thead>
<tr>
<th>φ₁</th>
<th>β₁</th>
<th>λ₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ₂</td>
<td>β₂</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>a₁</td>
<td>a₂</td>
</tr>
<tr>
<td>Iteration.4</td>
<td>161·89051809996</td>
<td>161·89051809996</td>
</tr>
<tr>
<td>Iteration.5</td>
<td>161·89052473571</td>
<td>161·89052473571</td>
</tr>
<tr>
<td>Iteration.6</td>
<td>161·89051780675</td>
<td>161·89051780675</td>
</tr>
<tr>
<td>Iteration.7</td>
<td>161·89051748643</td>
<td>161·89051748643</td>
</tr>
<tr>
<td>Iteration.8</td>
<td>161·89052806001</td>
<td>161·89052806001</td>
</tr>
<tr>
<td>Iteration.9</td>
<td>161·89052470289</td>
<td>161·89052470289</td>
</tr>
<tr>
<td>Karny</td>
<td>161·89052473633</td>
<td>161·89052473633</td>
</tr>
<tr>
<td>Geo-K</td>
<td>0·00001259714</td>
<td>0·00001259714</td>
</tr>
<tr>
<td>GE</td>
<td>119·81278909040</td>
<td>60·08786527687</td>
</tr>
<tr>
<td>Geo-GE</td>
<td>42·07773561248</td>
<td>42·07773561248</td>
</tr>
<tr>
<td>MATLAB</td>
<td>249·965610768079</td>
<td>249·965610768079</td>
</tr>
<tr>
<td>Geo-M</td>
<td>88·0750860652</td>
<td>88·0750860652</td>
</tr>
</tbody>
</table>

Table 3. Determining route from Yokohama to Valparaiso.

<table>
<thead>
<tr>
<th>φ₁</th>
<th>β₁</th>
<th>λ₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ₂</td>
<td>β₂</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>a₁</td>
<td>a₂</td>
</tr>
<tr>
<td>Iteration.4</td>
<td>95·6463916608066</td>
<td>95·6463916608066</td>
</tr>
<tr>
<td>Iteration.5</td>
<td>95·6463916578752</td>
<td>95·6463916578752</td>
</tr>
<tr>
<td>GE</td>
<td>95·6463916578635</td>
<td>95·6463916578635</td>
</tr>
<tr>
<td>Geo-Ge</td>
<td>-0·1545300</td>
<td>-0·1545300</td>
</tr>
<tr>
<td>MATLAB</td>
<td>95·6463916578635</td>
<td>95·6463916578635</td>
</tr>
<tr>
<td>Geo-M</td>
<td>0·0000000000117</td>
<td>0·0000000000117</td>
</tr>
</tbody>
</table>
As one can see from Table 3, our method converges very quickly after the first few iterations. The $\Delta \lambda$ change of the calculated longitude difference and the given $\lambda_{12}$ also converge very quickly. In the computer program designed by us, the tolerance is set to $10^{-10}$ degree, the program designed for geodesic sailing here generates 50 linearly equally spaced points between the two latitudes $\beta_1$ and $\beta_2$. The behaviour of change in longitude decreases as the latitude approaches to the Equator as shown in Table 4.

<table>
<thead>
<tr>
<th>NO.</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Azimuth</th>
<th>NO.</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Azimuth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34:43630</td>
<td>139:85650</td>
<td>95:64639</td>
<td>26</td>
<td>0:03103</td>
<td>−140:35266</td>
<td>124:75132</td>
</tr>
<tr>
<td>24</td>
<td>2:78550</td>
<td>−144:30285</td>
<td>124:65418</td>
<td>49</td>
<td>−31:62536</td>
<td>−78:28040</td>
<td>105:41146</td>
</tr>
</tbody>
</table>

Figure 9. The track from Yokohama, Japan to Valparaiso, Chile.

As one can see from Table 3, our method converges very quickly after the first few iterations. The $\Delta \lambda$ change of the calculated longitude difference and the given $\lambda_{12}$ also converge very quickly. In the computer program designed by us, the tolerance is set to $10^{-10}$ degree, the program designed for geodesic sailing here generates 50 linearly equally spaced points between the two latitudes $\beta_1$ and $\beta_2$. The behaviour of change in longitude decreases as the latitude approaches to the Equator as shown in Table 4.

6. CONCLUSIONS. This study presents algorithms for determining the Clairaut constant or the vertex latitude from two given points on the spheroid. If the given points are not in the antipodal regions (AR) governed by Equation (13), the iterative formula of Equation (33) solves the vertex latitude. If the given points are in the AR, the alternative iterative procedure described in Equation (47) solves the Clairaut constant, in which in all cases the Clairaut constant and the vertex latitude are unique. Once the Clairaut constant or the vertex latitude is determined, the length, azimuths and waypoints along the geodesic are easily obtained. Outside the nearly AR the geodesic is unique, while for the endpoints within the AR there are generally dual geodesics mirrored in the equator, i.e. the azimuths are complementary at each point.
Our technique to solve the inverse geodetic problem is more accurate and complete than solutions found in other literatures mainly from the central role played by the initial course or Clairaut constant, where partial ideas of this were originally presented by Sjöberg (2007; 2012). Once the constant or vertex latitude is given, the solution is straightforward. In this way we avoid the difficulties of finding the iterative solutions and the starting guess points in the AR which are as reported e.g. in Vincenty (1975a) and Karney (2013). With aid of CAS and basic calculus knowledge, the mathematical derivations presented here are more straightforward and understandable. The variable $t$ is introduced to replace reduced latitude which has the interval between $[-\infty, \infty]$, using the conditional longitude and distance functions with the variable $t$ can sketch any point on the geodesic. The methods provided here do not involve any auxiliary sphere and corresponding arguments which then need many spherical trigonometry formulae; our methods can reach arbitrary truncated series as the user’s need is limited only by ability of the computer used.

The expressions derived here are suitable to both the syntax of computer algorithms and research purposes in the areas of sailing and cartographical computation in GIS and ECDIS environments. The differences of distances between the great ellipse and geodesic sailing passing two antipodal points will generate the greatest discrepancy whereas on the Equator, it can reach a maximum of about 18,130,799 nautical miles. This is not acceptable for practical purposes of navigation and ECDIS, and there may be alternatives to solve this problem, such as a compromised solution with two separated great ellipses sailing, which may be worth studying in the future. The proposed algorithm for geodesic sailing provides extremely high accuracies with efficiencies comparable to those algorithms provided by previous literatures. Numerical tests show that the algorithms of the geodesic provided here gives near identical results, yet it is faster and more direct than previous literatures, and can also be used in nearly antipodal regions where MATLAB’s distance function often does not work correctly. The aim of this paper is to facilitate navigators and designers of GIS or electronic charts to design navigational software more efficiently, accurately, easily and much more understandably.

ACKNOWLEDGEMENT

This work was supported in part by the National Science Council of Taiwan, Republic of China, under grant NSC 98-2410-H-019-019-, NSC 99-2410-H-019-023-, 100-2410-H-019-018-, 101-2410-H-019-025-, and 102-2410-H-019-016-.

REFERENCES


