



Theory, strict formula derivation and algorithm development for the computation of a geodesic polygon area

Edward Nowak¹ · Joanna Nowak Da Costa²

Received: 23 July 2020 / Accepted: 21 January 2022 / Published online: 24 March 2022
© The Author(s) 2022

Abstract

The paper presents a coherent, original theory and an effective algorithm to calculate the surface area of a geodesic polygon, i.e. the area on an ellipsoid of revolution limited by geodesics, connecting pairs of subsequent vertices of known geographical coordinates. Using the cylindrical coordinate system and introducing an equatorial geodesic triangle, the novel strict recursive formulas were derived for the longitude on the ellipsoid and the distance along the geodesic and the area of a geodesic polygon. The area is determined by constant coefficients, not the values of finite series. The orthodromic properties of geodesics, which led to the detection of bifurcation lines composed of the double solutions of the inverse geodetic problem were analysed. The novel algorithm presented, integrating the problem with geodesic polygon area calculation, is based on the newly proposed parameterisation with two alternative working unknowns and a new formula to obtain the initial value of the geodesic C parameter. It allows solutions for all cases, including near-antipodal or near-vertices, and offers two solutions if bifurcation occurs. The algorithm presented was tested and verified using numerical examples from known publications. The numerical tests proved the accuracy of 1m^2 of the computed geodesic area using standard IEEE 754 Double Floating Point PC arithmetic, for geodesic lengths up to ten thousand km.

Keywords Area under geodesic · Geometrical geodesy · Recursive formulas · Bifurcation · Inverse geodesic problem · GNSS

List of symbols

a	Equator semi-axis	$r; \rho = \frac{r}{a}$	equator semi-axis (auxiliary dimensionless variable)
b	Polar semi-axis		Radial distance in cylindrical coordinate system, radius of circle of latitude (parallel); Normalised radius
f	Flattening of ellipsoid of revolution		Arc length of the geodesic on ellipsoid of revolution; Normalised geodesic arc length
e^2	Eccentricity of ellipsoid of revolution	$s; S = \frac{s}{a}$	Arc length of meridian; Normalised meridian arc length
φ	Geographic latitude		Azimuth (arc's azimuth)
β	Parametric latitude (also referred to as reduced latitude)	$m; \mu = \frac{m}{a}$	Clairaut's constant; Normalised Clairaut's constant
λ	Geographic longitude or angle of rotation (angular coordinate) in cylindrical coordinate system	α	Constant coefficients used in the λ, S series
$z; \zeta = \frac{z}{a}$	Axial position in cylindrical coordinate system; height normalised with	$c; C = \frac{c}{a}$	Base integrals used in the λ, S series
		$g_{1,i}$	Initial base integral used in the λ, S series; Geographic longitude on the sphere
		$G_{1,i}$	
		$C \cdot G_{1,0}; \omega$	

✉ Joanna Nowak Da Costa
joanna.nowakdc@wat.edu.pl

¹ Faculty of Geodesy and Cartography, Warsaw University of Technology, Warsaw, Poland

² Faculty of Civil Engineering and Geodesy, Military University of Technology, Warsaw, Poland

$G_{1,1}; \sigma$	Second initial base integral used in the λ, S series; Arc length of the geodesic on the sphere
$Ls(C^2); Le(C^2, \rho_P^2)$	Decomposition coefficients (multipliers) of geographic longitude, used in the λ series
$Ss(C^2); Se(C^2, \rho_P^2)$	Decomposition coefficients of arc length of the geodesic, used in the S series
A	Surface area on the ellipsoid of revolution
T	Function of the area distribution along the meridian
J_i	Base integrals of the area distribution along the meridian
$g_{2,k}$	Constant coefficients used in the series defining area under the geodesic
$G_{2,k}$	Base integrals used in the series defining area under the geodesic
$Ar(C^2, \rho_P^2)$	Auxiliary function simplifying the formula for the calculation of the area under geodesic
BIF	Boolean variable determining whether bifurcation occurs
\tilde{C}	Starting estimate of the Clairaut's constant

1 Introduction

The computation of the surface area of polygons, whose sides are arcs of geodesics, is related to the knowledge of the exact parameters of geodesics equations. A geodesic is a unique, locally minimising length, curve joining two points on any surface, e.g. a straight line is a geodesic on a plane. Geodesics are used in classical geodetic problems of computing the geodetic coordinates, azimuths and distances on the surface of an ellipsoid of revolution. The formula of the accuracy of e^8 for calculating the area of a figure between the geodesic and the equator bounded by two meridians (referred to as the area under the geodesic) was formulated by Nielsen in 1989, but the practical application of this formula is difficult in terms of calculation. As a consequence, although storing and calculating data by their original geographical coordinates – currently being achieved easily and more and more accurately, e.g. from Global Navigation Satellite Systems – seems the ideal solution (as noticed also by many others researchers i.e. Canters and De Genst 1997, Leung et al. 2004, Chrisman 2017), users of data that define spatial location are mostly using coordinates expressed in traditional planar coordinate systems, based on cartographic map projections that allow for quick calculations of the area

of polygons bounded by straight line segments on projected ellipsoids. The pragmatic choice is an equal-area projection, such as the Albers equal-area conic projection – used in the USA (Snyder 1987), the Lambert Azimuthal Equal-Area projection – recommended in the EU for statistical analysis and display (Annoni et al. 2001) or even Mollweide projection – still in use for world or sky maps (Snyder 1987). Even if an equal-area projection is used, however, the polygon border segments, i.e. the projected geodesics, do not coincide with straight lines, which is an important source of error of the surface area calculated in this manner. Moreover, many of these projections cover only a part of the globe, which generates additional calculation problems when calculating areas outside these mapping zones or outside the boundaries of the projected ellipsoid surface.

As an introduction to the topic of calculating the surface area on an ellipsoid, let us analyse the task of calculation of the surface area of a triangle on a sphere of radius $R = 7 \cdot 10^6 = 7,000,000$ m with the accuracy of 1m^2 , using the formula $S = R^2 \varepsilon$, i.e. $S = 4.91 \cdot 10^{13} \cdot \varepsilon$, where ε is the spherical excess. The errors are propagated according to the formula $\sigma_s = 4.9 \cdot 10^{13} \cdot \sigma_\varepsilon [\text{m}^2]$, $\sigma_s = 1 [\text{m}^2]$, hence $\sigma_\varepsilon = 2 \cdot 10^{-14}$. Consequently, the angles must be determined with at least fourteen digits of precision. In contrast, the determination of the position $x = R\varphi = 7 \cdot 10^6 \varphi$ with the accuracy of $\sigma_x = 1 \cdot 10^{-3} = 1$ mm, means that the geographical coordinates need to be calculated with a precision of ten digits. In this context, the issue of the determination of the surface area of a polygon delimited by geodesics on an ellipsoid casts new light on the classical geodetic inverse problem and requires a decidedly higher computational precision.

The research presented here describes a new theory and approach to obtaining the explicit equation in cylindrical coordinates and the arc length of a geodesic on an ellipsoid of revolution and the surface area of a geodesic polygon, i.e. the area bounded by arcs of geodesics connecting vertices with known geographical coordinates such as Global Navigation Satellite System (GNSS) derived coordinates. The goal was to derive strict formulas and develop an algorithm that can quickly find the solution with high accuracy. Furthermore, the solution should be applicable to all cases, including near-antipodal and near-vertices cases.

The focus of this research includes the determination of the border path and surface area of a contour delineated by any feature on the surface of the Earth or any surface that is modelled by an ellipsoid of revolution. The feature's boundaries and surface area are determined unambiguously, regardless of the mapping projection and the accuracy of the computed result is only limited by the floating-point representation of the computing system used. For example, high accuracy of the boundary and surface area is needed not only for the maintenance of a cadastre at the national level, but also at the international level or to support the calculation

of charges associated with the use of a given contour, such as territorial seas and the leases for the exploitation of the shelf's natural resources.

The three most important novel contributions of this theory and approach to accurately calculate the surface area of a geodesic polygon are as follows.

1. A new geometry on an ellipsoid is introduced, namely an equatorial geodesic triangle (see Sect. 5.1). This surface shape, bounded by three geodesics, simplifies determination of the area under the geodesic and, consequently, the area of a geodesic polygon on an ellipsoid that equals the total areas under the geodesics stretched on each pair of successive polygon vertices.
2. The difference between the reverse and forward azimuths (α_{12}) of the arc of a geodesic is directly determined, based on Napier's rules, thus avoiding subtraction which reduces the accuracy of the surface area calculation (see Sect. 7.2).
3. An innovative parameterisation of the inverse problem with two alternative working unknowns (see Sect. 7.1) to solve all cases is presented.

We extend the work by Karney (2013), Danielsen (1989) and Vincenty (1975, 1976) in several ways:

- 1 A coherent original theory of ellipsoid geometry is presented; the novel, strict recursive formulas for the equation in cylindrical coordinates and the arc length of a geodesic, the area under the geodesic and the surface area of the geodesic polygon are derived; the area is determined by constant coefficients, not the values of finite series.
- 2 A cylindrical coordinate system is used and the integration is performed with respect to the radius of the circle of latitude, not to the frequently used reduced geodetic latitude.
- 3 The solution is applicable to any ellipsoid of revolution adopted as the reference surface, whether oblate or prolate.
- 4 The phenomenon of bifurcation, i.e. the existence of two different geodesic arcs of the same length connecting a given pair of points, leading to two solutions of the inverse geodetic problem, is detected.
- 5 A novel algorithm for solving the inverse problem, extended by the area under the geodesic calculation, is developed; it uses Newton's iterative process, and is based on the newly proposed parameterisation with two alternative working unknowns, providing a solution for all cases, including near-antipodal or near-vertices ones, and offers two solutions if bifurcation occurs.
- 6 A new formula is given to estimate the initial value of the geodesic C parameter, covering all cases, including near-antipodal or near-vertices ones, which replaces the

fourth degree equation for Jacobi's (1891) astroid frequently used.

- 7 The accuracy of the resulting values, i.e. the geodesic arc length, longitude and the area under the geodesic, are only limited by the computer hardware.

2 Related works

The two fundamental geodetic operations are the problems of the geodesic on an ellipsoid: the direct problem is to find the position along the geodesic given a starting point, an initial heading (azimuth) and a distance; while the inverse problem is to find the distance and the forward and reverse azimuths given the latitude and longitude of the two points. Although these problems are fundamentally important, only few researchers have attempted to solve them from a theoretical standpoint, i.e. to derive the formulas and fully examine the existence of a solution. The aim of the published formulas or methods for solving these problems was usually to offer specific practical applications (e.g. triangulation), which typically referred to geodesics of a few tens of kilometres. Long geodesic arcs, antipodal and near-antipodal cases were rarely analysed. This section presents a non-exhaustive selection of significant contributions in the research of the length of geodesics and area on an ellipsoid.

The issue of geodesics first appeared in the works of Clairaut (1733) and it was applied in geodesy in the form of the direct and inverse problem by Bessel (1826). The latter was the first to deal with long geodesics. Bessel developed a series expansion of elliptic integrals (distance and longitude) using an auxiliary sphere, while Helmert (1880) perfected the technology of the practical computation of the classical geodetic problems. Modifying the original Bessel method, Vincenty (1975, 1976) provided the formulas and algorithm for solving the problems that arose when technological progress brought the beginning of the computer era and improved accuracy of measurements. Although Vincenty did not present a derivation of the formulas, they have been published on the official website of the American National Geodetic Survey agency (www.ngs.noaa.gov) along with the original Fortran code, and they are widely known. Vincenty used the expansion of elliptical integrals into a series to modify Helmert's iterative method and achieved high computational efficiency, which was extremely important for military applications, to which the formulas were then dedicated. Using Vincenty's formulas, however, a failure to converge for near-antipodal points was noticed, which is proven by a very high number of iterations. This fact was mentioned by numerous authors, e.g. Rapp (1993), Karney (2013) or Tseng (2014).

The studies by Kivioja (1971) mark the beginning of a new trend, namely numerical integration of elliptical integrals. Its supporters, e.g. Thomas and Featherstone (2005), developed the numerical integration of the differential equation of the geodesic. The parallel numerical approach was initiated by Saito (1970) and then developed by Sjöberg (2006, 2012). This involves the numerical integration of elliptical integrals, which are usually functions of the elements of the ellipsoid and the auxiliary sphere.

As far as the solutions for the longitude and distance along the geodesic are concerned, it is also worth noting the approaches using a recursion relation: by Levallois and Dupuy (1952) and by Pittman (1986). Pittman expressed the distance integral by a recursion formula consisting of base integrals. The delta lambda formula (longitude) involves double adding up (i.e. a series within a series). These are the only solutions that resemble some of the presented in this study, except that the recursion was used only for deriving formulas, not for calculations. Moreover, Pittman's formula, in particular, is very complex in computational terms, and its assumption that the initial approximation of the parameter is of spherical origin, makes it impossible to solve antipodal tasks. Pittman's significant contribution was to accelerate the convergence of the inverse geodetic problem's solution by using Newton's iterative process. Unfortunately, this method does not provide a solution if the end point of the geodesic arc is localised near its turning point, a problem of which Pittman is also aware. Moreover, Pittman's use of an approximate derivative increases the number of iterations; and this derivative equals zero if the end points of the geodesic arc lie on the same latitude.

The only formula to determine the surface area under a geodesic in geodetic coordinates on an ellipsoid was presented by Danielsen (1989), but this was not widely accepted at first. Danielsen derived a series to develop the formula to calculate the surface area under the geodesic and provided its six elements, each of which is a short finite series. The application of the formula requires the subtraction of two large, close numbers, calculated by a series, with a consequent sharp drop in the numerical accuracy of the result. Despite its imperfections, Danielsen's work represents a milestone in this field alongside that of Karney (2013), who popularised and extended it. Karney also used Newton's iterative process to accelerate the inverse geodetic problem convergence.

3 Equations for geodesics on a surface of revolution

When r , z , λ , cylindrical coordinates are used on the surface of revolution, a curve is determined by the arc length differential and azimuth (Fig. 1) defined by formulas (1) and (2),

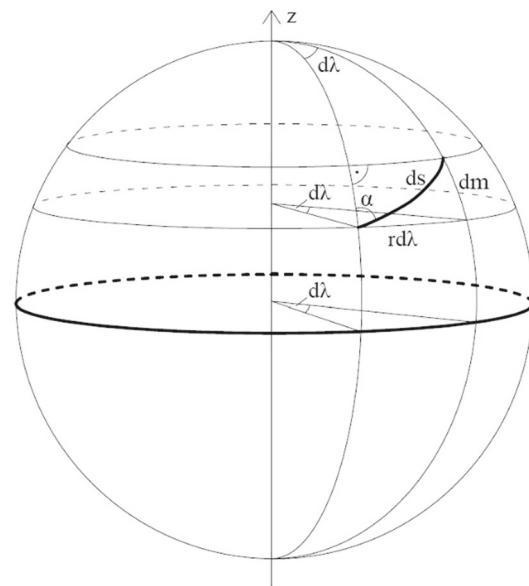


Fig. 1 Differential elements of a curve in cylindrical coordinates

respectively:

$$ds^2 = dm^2 + r^2 d\lambda^2 \quad (1)$$

$$\sin \alpha = r \frac{d\lambda}{ds} \quad (2)$$

where: r is the radius of a parallel (the distance of a point of the surface of revolution from the axis of rotation), λ is the longitude (also referred to as geodetic longitude if datum is specified), m is the meridian arc length, α is the azimuth of a curve in a given point.

If a curve on a surface of revolution is a geodesic, it is characterised by Clairaut's relation (Clairaut 1733; Pressley 2010), also known as Clairaut's constant:

$$r \sin \alpha = c = \text{const} \quad (3)$$

It should be noted that this relation discloses the sinusoidal character of geodesics on a surface of revolution and the $r \geq |c|$ constraint for the radius of a parallel. It can be demonstrated that the radius takes the minimum value, i.e. $r_P = |c|$, if and only if $P = P_0$ is a turning point of a geodesic.

Substituting (2) into (3) we get a new equation which determines the relationship between the longitude and the distance along a geodesic:

$$r^2 d\lambda = c ds \quad (4)$$

From (4) it is important to note that the sign of the constant c determines the sign of $d\lambda$.

Substituting ds from Eq. (4) into Eq. (1) we get consecutively

$$dm^2 + r^2 d\lambda^2 = \frac{r^4}{c^2} d\lambda^2, \quad dm^2 = \frac{r^2(r^2 - c^2)}{c^2} d\lambda^2$$

$$d\lambda = \frac{c}{r\sqrt{r^2 - c^2}} dm \quad (5)$$

Substituting $d\lambda$ from Eq. (4) into Eq. (5) we get successively

$$ds = \frac{r}{\sqrt{r^2 - c^2}} dm \quad (6)$$

After integrating Eq. (5) and (6), we obtain the explicit equation of a geodesic in cylindrical coordinates (7), i.e. longitude as a function of radius (often simply referred to as longitude), and the geodesic arc length (8) (often referred to as distance):

$$\lambda(c, r) = c \int \frac{r^{-1} dm}{\sqrt{r^2 - c^2}} \quad (7)$$

$$s(c, r) = \int \frac{r^{+1} dm}{\sqrt{r^2 - c^2}} \quad (8)$$

The similarity of both integrals was exposed in the notation, which consequently leads to a very efficient algorithm of their simultaneous determination.

4 Equations of geodesics on the ellipsoid of revolution

An ellipsoid of revolution in cylindrical coordinates is described with equation

$$\frac{r^2}{a^2} + \frac{z^2}{b^2} = 1, \quad -\pi < \lambda \leq \pi \quad (9)$$

Hence

$$\frac{dz}{dr} = -\frac{\sqrt{1 - e^2} r}{\sqrt{a^2 - r^2}}$$

$$dm = \sqrt{1 + \left(\frac{dz}{dr}\right)^2} dr = -\sqrt{\frac{a^2 - e^2 r^2}{a^2 - r^2}} dr \quad (10)$$

where $e^2 = 1 - \frac{b^2}{a^2} = f(2 - f)$, $f = \frac{a-b}{a}$, a – equatorial semi-axis, b – polar semi-axis.

It should be noted that the parameter which is commonly marked as e^2 can have a negative value, i.e. any ellipsoid of revolution can be adopted as the reference surface, whether oblate or prolate.

Substituting Eq. (10) into Eq. (7) and Eq. (8) and introducing dimensionless values $\rho = \frac{r}{a}$, $S = \frac{s}{a}$, $\mu = \frac{m}{a}$, $C = \frac{c}{a}$, we will get a description of the geodesic arc from the equator to the turning point (also referred to as the furthermost point of a geodesic), i.e. the longitude (11) and the distance (12) of the point of the radius $\rho = \rho_P$, respectively:

$$\lambda(C; \rho_P) = -C \int_1^{\rho_P} \frac{\sqrt{1 - e^2 \rho^2}}{\sqrt{1 - \rho^2}} \frac{\rho^{-1}}{\sqrt{\rho^2 - C^2}} d\rho \quad (11)$$

$$S(C; \rho_P) = - \int_1^{\rho_P} \frac{\sqrt{1 - e^2 \rho^2}}{\sqrt{1 - \rho^2}} \frac{\rho^{+1}}{\sqrt{\rho^2 - C^2}} d\rho \quad (12)$$

Note that the following property occurs: $|C| \leq \rho_P \leq 1$.

4.1 Integrals describing the arc of a geodesic from the equator to the turning point

Introducing an auxiliary binomial series

$$(1 - e^2)^{1/2} = \sum_{i=0}^{\infty} g_{1,i};$$

$$g_{1,0} = 1; \quad g_{1,i>0} = \frac{2i-3}{2i} e^2 g_{1,i-1} \quad (13)$$

of which the constant coefficients are expressible by a Newton's generalised binomial coefficient as $g_{1,i} = \binom{1/2}{i} (-e^2)^i$, e.g. in Tseng (2014), however this requires more number of calculations than the proposed recursive formula (13).

We will convert (11) and (12) integrals into a linear combination of basic integrals

$$\begin{aligned} \lambda(C; \rho_P) &= -C \int_1^{\rho_P} \frac{\rho^{-1} \sqrt{1 - e^2 \rho^2}}{E(C^2; \rho^2)} d\rho \\ &= -C \int_1^{\rho_P} \frac{\rho^{-1}}{E(C^2; \rho^2)} \sum_{i=0}^{\infty} g_{1,i} \rho^{2i} d\rho = \\ &= -C \sum_{i=0}^{\infty} \int_1^{\rho_P} \frac{g_{1,i} \rho^{2i-1} d\rho}{E(C^2; \rho^2)} = C \sum_{i=0}^{\infty} g_{1,i} G_{1,i}(C; \rho_P) \end{aligned} \quad (14)$$

$$\begin{aligned} S(C; \rho_P) &= - \int_1^{\rho_P} \frac{\rho^{+1} \sqrt{1 - e^2 \rho^2}}{E(C^2; \rho^2)} d\rho \\ &= - \int_1^{\rho_P} \frac{\rho^1}{E(C^2; \rho^2)} \sum_{i=0}^{\infty} g_{1,i} \rho^{2i+1} d\rho \\ &= - \sum_{i=0}^{\infty} \int_1^{\rho_P} \frac{g_{1,i} \rho^{2i+1} d\rho}{E(C^2; \rho^2)} = \sum_{i=0}^{\infty} g_{1,i} G_{1,i+1}(C; \rho_P) \end{aligned} \quad (15)$$

where basic integrals are as follows:

$$G_{1,i}(C; \rho_P) = - \int_1^{\rho_P} \frac{\rho^{2i-1} d\rho}{E(C^2; \rho^2)} \quad (16)$$

$$\begin{aligned} E(C^2; \rho^2) &= \sqrt{(1-\rho^2)(\rho^2-C^2)} \\ &= \sqrt{-\rho^4 + (1+C^2)\rho^2 - C^2} \end{aligned} \quad (17)$$

Separating the derivative of the radicand, we will get

$$\begin{aligned} G_{1,i}(C; \rho_P) &= \int_1^{\rho_P} \frac{-4\rho^3 + 2(1+C^2)\rho}{2E(C^2; \rho^2)} \frac{\rho^{2i-4}}{2} d\rho \\ &\quad - \frac{1+C^2}{2} \int_1^{\rho_P} \frac{\rho^{2i-3} d\rho}{E(C^2; \rho^2)} \end{aligned}$$

Integrating the first integrand by parts, we will get the following expression

$$\begin{aligned} G_{1,i}(C; \rho_P) &= \frac{\rho_P^{2i-4}}{2} E(C^2; \rho_P^2) \\ &\quad - \frac{2i-4}{2} \int_1^{\rho_P} \rho^{2i-5} \frac{-\rho^4 + (1+C^2)\rho^2 - C^2}{E(C^2; \rho^2)} d\rho \\ &= \frac{\rho_P^{2i-4}}{2} E(C^2; \rho_P^2) \\ &\quad - (i-2) \int_1^{\rho_P} \frac{-\rho^{2i-1} + (1+C^2)\rho^{2i-3} - C^2\rho^{2i-5}}{E(C^2; \rho^2)} d\rho \\ &= \frac{\rho_P^{2i-4}}{2} E(C^2; \rho_P^2) \\ &\quad - (i-2) \left(G_{1,i} - (1+C^2) G_{1,i-1} + C^2 G_{1,i-2} \right) \end{aligned}$$

Now, the following functional equation can be written as

$$\begin{aligned} G_{1,i} &= \frac{\rho_P^{2i-4}}{2} E(C^2; \rho_P^2) \\ &\quad - (i-2) \left(G_{1,i} - (1+C^2) G_{1,i-1} + C^2 G_{1,i-2} \right) \\ &\quad + \frac{1+C^2}{2} G_{1,i-1} \end{aligned}$$

which, once rearranged, leads to a recursive relation enabling effective determination of basic integrals.

$$G_{1,0}(C; \rho_P) = \frac{1}{C} \sin^{-1} \left(\frac{C \sqrt{1-\rho_P^2}}{\rho_P \sqrt{1-C^2}} \right) \quad (18)$$

$$G_{1,1}(C; \rho_P) = \sin^{-1} \left(\frac{\sqrt{1-\rho_P^2}}{\sqrt{1-C^2}} \right) \quad (19)$$

$$\begin{aligned} G_{1,i>1} &= \frac{E(C^2; \rho_P^2) \rho_P^{2i-4} + (2i-3)(1+C^2) G_{1,i-1} - (2i-4)C^2 G_{1,i-2}}{2i-2} \end{aligned} \quad (20)$$

The two initial basic integrals can be obtained by substituting $e^2 = 0$ into formulas (11) and (12). They show the longitude (18) $C \cdot G_{1,0}(C; \rho_P) \stackrel{\text{def}}{=} \omega$ and arc length (19) $G_{1,1}(C; \rho_P) \stackrel{\text{def}}{=} \sigma$ of the geodesic on the sphere, which can be referred to as a special (boundary) case of an ellipsoid.

Since the basic integrals are positive and $0 \leq \rho \leq 1$, the integrand (16) cannot increase with an increase in the i index, and hence $G_{1,i+1} \leq G_{1,i}$. This, in turn, allows us to show the convergence of series (14) and (15) based on Eq. (13).

$$\begin{aligned} \lambda(C; \rho_P) &\leq C \cdot G_{1,0} \sum_{i=0}^{\infty} g_{1,i} = C \cdot G_{1,0} (1-e^2)^{1/2} \\ S(C; \rho_P) &\leq G_{1,1} \sum_{i=0}^{\infty} g_{1,i} = G_{1,1} (1-e^2)^{1/2} \end{aligned}$$

4.2 Decomposition of formulas which determine the longitude and distance of a geodesic

The recursive formula (18, 19, 20) suggests that the basic integrals (20) depend on the two initial integrals, i.e. (18, 19). Let us calculate the next integral:

$$G_{1,2} = \frac{1}{2} E(C^2; \rho_P^2) + \frac{1+C^2}{2} G_{1,1}$$

It turns out that the third term is eliminated and this and successive integrals are functions with the following form

$$G_{1,i>1} = \gamma_i(C^2) G_{1,1} + \varepsilon_i(C^2, \rho_P^2) E(C^2; \rho_P^2) \quad (21)$$

The new coefficients are determined from the converted recursive formula (20)

$$\begin{aligned} \gamma_0 &= 0, \quad \gamma_1 = 1, \\ \gamma_{i>1} &= \frac{2i-3}{2i-2} (1+C^2) \gamma_{i-1} - \frac{2i-4}{2i-2} C^2 \gamma_{i-2} \end{aligned} \quad (22)$$

$$\begin{aligned} \varepsilon_1 &= 0, \quad \varepsilon_2 = \frac{1}{2}, \\ \varepsilon_{i>2} &= \frac{P^{2i-4}}{2i-2} + \frac{2i-3}{2i-2} (1+C^2) \varepsilon_{i-1} - \frac{2i-4}{2i-2} C^2 \varepsilon_{i-2} \end{aligned} \quad (23)$$

Table 1 Variation range of the introduced L_s , S_s , L_e , S_e multipliers

	Max	Min	Max–min
L_s	$-3.349998 \cdot 10^{-3}$	$-3.352811 \cdot 10^{-3}$	$2.812693 \cdot 10^{-6}$
L_e	$-2.807980 \cdot 10^{-6}$	$-2.819790 \cdot 10^{-6}$	$1.181280 \cdot 10^{-8}$
S_s	$9.983243 \cdot 10^{-1}$	$9.966472 \cdot 10^{-1}$	$1.677109 \cdot 10^{-3}$
S_e	$-1.067570 \cdot 10^{-3}$	$-1.679220 \cdot 10^{-3}$	$3.523550 \cdot 10^{-6}$

Substituting (21) into formulas (14) and (15) we get the formulas which decompose integrals into spherical distance and position impacts (Eq. 24–27).

$$\lambda(C; \rho_P) = \omega + C (\sigma L_s(C^2) + E(C^2; \rho_P^2) L_e(C^2, \rho_P^2)) \quad (24)$$

$$S(C; \rho_P) = \sigma S_s(C^2) + E(C^2; \rho_P^2) S_e(C^2, \rho_P^2) \quad (25)$$

$$L_s(C^2) = \sum_{i=1}^{\infty} g_{1,i} \gamma_i; \quad S_s(C^2) = \sum_{i=1}^{\infty} g_{1,i-1} \gamma_i \quad (26)$$

$$L_e(C^2, \rho_P^2) = \sum_{i=2}^{\infty} g_{1,i} \varepsilon_i; \quad S_e(C^2, \rho_P^2) = \sum_{i=2}^{\infty} g_{1,i-1} \varepsilon_i \quad (27)$$

In this way, the elliptical integrals are expressed by combinations of elementary functions and polynomial series. Tseng (2014) obtained a similar character of the formulas for λ and S .

The practical application of formulas (24) and (25) is facilitated by the properties of the multipliers. L_s , L_e , S_e change in very narrow ranges (Table 1), while the $\gamma_i = 1$ (for $i > 1$) relation occurs on the equator (i.e. $C = 1$). Therefore, based on Eq. (13):

$$\begin{aligned} S_s(C^2 = 1) &= \sum_{i=1}^{\infty} g_{1,i-1} \\ &= \sum_{i=0}^{\infty} g_{1,i} = \sqrt{1 - e^2} = 1 - f; \end{aligned}$$

$$L_s(C^2 = 1) = \sum_{i=1}^{\infty} g_{1,i} = -f$$

Formulas (24), (25) express the effect of the change of the polar axis (in respect to the equatorial axis) on the deformation of the sphere. The main influence of the flattening f propagates proportionally to the length σ . The remaining nonlinear part is proportional to the function $E(C^2; \rho_P^2)$. It follows from Eq. (17) that the function E equals zero at the equator $E(C^2; \rho_P^2 = 1) = 0$ and at the turning points of a

geodesic $E(C^2; \rho_P^2 = C^2) = 0$, while its maximum is

$$\max E(C^2; \rho_P^2) = E\left(C^2; \rho_P^2 = \frac{1+C^2}{2}\right) = \frac{1-C^2}{2}.$$

4.3 Solution in spherical coordinates

Section 4.1 presents integrals providing the solution for the arc of a geodesic from the equator to the turning point of a geodesic. Let us introduce new variables leading to the extension of the solution to the entire geodesic.

A sphere with geographical coordinates, marked by Karney as $\omega = \lambda^{\text{sfer}}$ and $\beta = \varphi^{\text{sfer}}$, is a boundary case of $e^2 \rightarrow 0$ ellipsoid. In order to achieve full correspondence between the geodesics on the ellipsoid and the great circle on the sphere, we assume identical radii and azimuths, following Bessel (1826); hence

$$\rho = \cos(\beta) \quad (28)$$

where β is the parametric (also referred to as reduced) latitude related to the geodetic latitude φ by equation

$$\tan(\beta) = (1-f)\tan(\varphi) \quad (29)$$

The essence of extending the solution beyond the part of the arc of the geodesic line from the equator to its turning point, is the following substitution:

$$\sqrt{1 - \rho^2} = \zeta = \sin(\beta) \quad (30)$$

which allows to obtain the solution also on the southern hemisphere.

Using Clairaut's relationship, we get the extension

$$\sqrt{\rho^2 - C^2} = \rho \cos(\alpha) \quad (31)$$

$$E(\alpha, \beta) = \zeta \rho \cos(\alpha) = \sin(\beta) \cos(\beta) \cos(\alpha) \quad (32)$$

that (a) has a geometric interpretation and (b) determines the signs of the decomposition components. The extension defined by formulas (31–32) is fundamentally important since it makes it possible to apply formulas (24) and (25) to any arcs of geodesic. Moreover, it replaces the artificial determination of combinations of signs of individual components that occurs in Pittman's (1986) and Tseng's (2014) methods.

Focusing more closely on the function E , which constitutes the nonlinear component of the longitude and the distance of the geodesic line defined by formulas (24) and (25), it is noted that it changes its sign when a geodesic crosses the equator ($\beta = 0$) or passes through its turning point

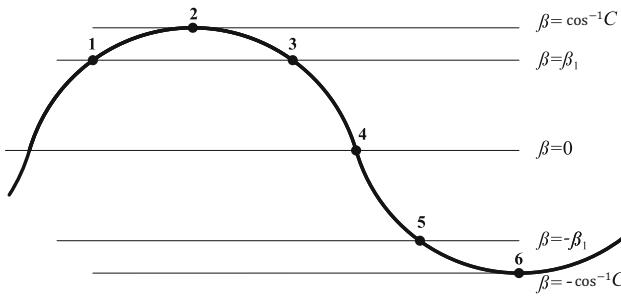


Fig. 2 Points of a geodesic representing some characteristic properties of the function E

(Fig. 2). This is a direct consequence of the properties of the turning point of a geodesic, which can be made apparent by differentiating relation (3)

$$dr \sin(\alpha) + r \cos(\alpha) d\alpha = 0$$

hence

$$\frac{dr}{d\alpha} = -\frac{r^2 \cos(\alpha)}{c}$$

The sign of the cosine changes when passing through the point(s) $P = P_0$ where $\alpha = \pm\pi/2$, which is precisely the turning point of a geodesic, thus $r_P = |c| \Leftrightarrow \rho_P = |C|$. Considering Eq. (28), β takes extreme values there, marking the northernmost or southernmost points of a geodesic.

Let examine the extreme values of the nonlinear terms in formulas (39) and (40). If $\rho_1^2 = \rho_2^2$, then, according to formula (42), we have $Le_1 = Le_2$; $Se_1 = Se_2$, and from formula (17), it follows that $|E_1| = |E_2|$. As a result, the absolute values of the nonlinear terms of both formulas become equal, i.e. $|E_1 Le_1| = |E_2 Le_2|$, $|E_1 Se_1| = |E_2 Se_2|$. Let us also note that the signs of the function E are determined by the rule derived from Eq. (32). As a result, the maximum of nonlinear influences occurs when points 1, 2 are situated on the same parallel, while for points located on antipodes the influences disappear (Fig. 2).

The remaining relevant properties of the function E for the points 1 to 6 lying on the geodesic line illustrated on Fig. 2 are as follows: $E_2 = E_4 = E_6 = 0$, $E_3 = -E_1$, $E_5 = E_1$.

4.4 Extension of the solutions to the entire geodesic

Using Eqs. (28–32) we reformulate Eqs. (18) and (19), i.e. the longitude and arc length of the geodesic on the sphere, in the following way

$$\sin(\omega) = \frac{C \sin(\beta)}{\sqrt{1 - C^2} \cos(\beta)} \quad (33)$$

$$\sin(\sigma) = \frac{\sin(\beta)}{\sqrt{1 - C^2}} \quad (34)$$

All solutions of the equations are determined by formulas for any integer k ($k \in \mathbb{Z}$):

$$\sigma(C, \beta, k) = k\pi + (-1)^k \sin^{-1}\left(\frac{\sin(\beta)}{\sqrt{1 - C^2}}\right) \quad (35)$$

$$\omega(C, \beta, k) = k\pi + (-1)^k \sin^{-1}\left(\frac{C \sin(\beta)}{\sqrt{1 - C^2} \cos(\beta)}\right) \quad (36)$$

Substituting them into (24) and (25), we will get

$$S(C; \rho_P, k) = \sigma(C, \beta, k) Ss(C^2) + E(C^2; \rho_P^2) Se(C^2, \rho_P^2) \quad (37)$$

$$\lambda(C; \rho_P, k) = \omega(C, \beta, k) + C \left(\sigma(C, \beta, k) Ls(C^2) + E(C^2; \rho_P^2) Le(C^2, \rho_P^2) \right) \quad (38)$$

Hence, it can be concluded that a geodesic is infinite and, analogous to a sinusoid, it oscillates around the equator with a $2\pi(1 + |C|L_s)$ period, intersecting the equator in points $\beta = 0$; $\lambda_k = k\pi(1 + |C|L_s)$. In practical applications, the geodesics do not exceed half of the period and finding the solution for $k = 0$ or $k = 1$ is not usually problematic.

4.5 Geodesic arc of length not exceeding half of the period

An arc of the geodesic with C parameter connects two points with reduced latitudes β_1, β_2 if $|C| \leq \rho_i = \cos(\beta_i)$ for $i = 1, 2$. Then

$$\begin{aligned} \lambda_{12} &= \lambda(C; \beta_2) - \lambda(C; \beta_1) \\ &= \omega_{12} + C \left(\sigma_{12} Ls(C^2) + E_2 Le_2 - E_1 Le_1 \right) \end{aligned} \quad (39)$$

$$\begin{aligned} S_{12} &= S(C; \beta_2) - S(C; \beta_1) \\ &= \sigma_{12} Ss(C^2) + E_2 Se_2 - E_1 Se_1 \end{aligned} \quad (40)$$

$$\omega_{12} = \omega_2 - \omega_1; \quad \sigma_{12} = \sigma_2 - \sigma_1; \quad (41)$$

$$E_i = E(\alpha_i, \beta_i); \quad Le_i = Le(C^2, \rho_i^2); \quad Se_i = Se(C^2, \rho_i^2) \quad (42)$$

There are two possible cases: if $\sigma_{12} < \pi$ then the spherical triangle meets the requirements for the Euler triangle, which enables the use of spherical trigonometry, or if $\sigma_{12} = \pi$ then the spherical triangle degenerates to a biangle.

5 Area on an ellipsoid of revolution

Using the area differential on the ellipsoid $d\text{Area} = r d\lambda dm = a^2 \rho d\mu d\lambda$ (Fig. 1), we determine the area distribution T along the meridian, which is the area (dA) measured from the equator to the parallel with radius ρ_P , between two differentially close ($d\lambda$) meridians

$$dA = a^2 T(\rho) d\lambda \quad (43)$$

$$T(\rho_P) = \int_1^{\rho_P} \rho d\mu = - \int_1^{\rho_P} \frac{\sqrt{1 - e^2 \rho^2}}{\sqrt{1 - \rho^2}} \rho d\rho \quad (44)$$

Using the binomial series (13) we transform the integral (44) into a following form:

$$\begin{aligned} T(\rho_P) &= - \int_1^{\rho_P} \frac{\rho \sqrt{1 - e^2 \rho^2}}{\sqrt{1 - \rho^2}} d\rho \\ &= - \int_1^{\rho_P} \frac{\rho}{\sqrt{1 - \rho^2}} \sum_{i=0}^{\infty} g_{1,i} \rho^{2i} d\rho \\ &= - \sum_{i=0}^{\infty} g_{1,i} \int_1^{\rho_P} \rho^{2i} \frac{\rho}{\sqrt{1 - \rho^2}} d\rho \end{aligned}$$

In consequence, a linear combination of basic integrals is obtained as follows:

$$T(\rho_P) = \sum_{i=0}^{\infty} g_{1,i} J_i(\rho_P) \quad (45)$$

$$J_i(\rho_P) = \int_1^{\rho_P} \rho^{2i} \frac{(-\rho)}{\sqrt{1 - \rho^2}} d\rho \quad (46)$$

Integrating by parts, we have

$$\begin{aligned} J_i(\rho_P) &= \rho_P^{2i} \sqrt{1 - \rho_P^2} - 2i \int_1^{\rho_P} \rho^{2i-1} \frac{(1 - \rho^2)}{\sqrt{1 - \rho^2}} d\rho \\ &= \rho_P^{2i} \sqrt{1 - \rho_P^2} + 2i J_{i-1}(\rho_P) - 2i J_i(\rho_P) \end{aligned}$$

hence the recursive formula

$$J_{i>0}(\rho_P) = \frac{1}{2i+1} \left(\rho_P^{2i} \sqrt{1 - \rho_P^2} + 2i J_{i-1}(\rho_P) \right) \quad (47)$$

$$J_0(\rho_P) = \sqrt{1 - \rho_P^2} \quad (48)$$

Since all these integrals depend on the initial one, any of these integrals can be written as

$$J_i(\rho_P) = J_0(\rho_P) \sum_{k=0}^i \tau_{i,k} \rho_P^{2k} \quad (49)$$

where

$$\tau_{i,i} = \frac{1}{2i+1}; \quad \tau_{i,k<i} = \frac{2i}{2i+1} \tau_{i-1,k} \quad (50)$$

Substituting (49) into (45), we get

$$T(\rho_P) = \sqrt{1 - \rho_P^2} \sum_{k=0}^{\infty} t_k \rho_P^{2k} \quad (51)$$

$$t_k = \sum_{i=k}^{\infty} g_{1,i} \tau_{i,k} \quad (52)$$

Another method of calculating t_k constants is worth analysing. Coefficients $\tau_{i,k}$ compose a triangular table where $0 \leq k \leq i$; $i \geq 0$. The k^{th} column of the table is necessary in order to compute t_k . Let us compare the coefficients of the k^{th} and $k-1^{st}$ column in the same line. Using formula (50), we calculate successively

$$\tau_{k,k} = \frac{1}{2k+1}; \quad \tau_{k,k-1} = \frac{2k}{(2k-1)(2k+1)} \quad (53)$$

$$\begin{aligned} \tau_{i,k} &= \frac{2i}{2i+1} \tau_{i-1,k}; \\ \tau_{i,k-1} &= \frac{2i}{2i+1} \tau_{i-1,k-1} \quad \text{for } i = k+1, k+2, \dots \end{aligned}$$

Hence, after division

$$\tau_{i,k-1} = \tau_{i,k} \frac{\tau_{i-1,k-1}}{\tau_{i-1,k}} \quad \text{for } i = k+1, k+2, \dots$$

By using the previous recursive relation we get

$$\tau_{k,k-1} = \tau_{i,k} \frac{\tau_{k,k-1}}{\tau_{k,k}} \quad \text{for } i = k, k+1, \dots$$

After substituting (53), this becomes

$$\tau_{i,k-1} = \tau_{i,k} \frac{2k}{2k-1} \quad \text{for } i = k, k+1, \dots \quad (54)$$

Substituting into (52) we get

$$\begin{aligned} t_{k-1} &= \sum_{i=k-1}^{\infty} g_{1,i} \tau_{i,k-1} \\ &= g_{1,k-1} \tau_{k-1,k-1} + \sum_{i=k}^{\infty} g_{1,i} \tau_{i,k-1} \end{aligned}$$

$$\begin{aligned}
&= \frac{g_{1,k-1}}{2k-1} + \frac{2k}{2k-1} \sum_{i=k}^{\infty} g_{1,i} \tau_{i,k} \\
&= \frac{g_{1,k-1}}{2k-1} + \frac{2k}{2k-1} t_k
\end{aligned}$$

Hence, the recursive formula which replaces the computation of series (52) is

$$t_k = \frac{(2k-1)t_{k-1} - g_{1,k-1}}{2k} \quad (55)$$

We calculate the initial constant as

$$t_0 = \sum_{i=0}^{\infty} g_{1,i} \tau_{i,0} = \sum_{i=0}^{\infty} \psi_i$$

After taking into consideration formulas (13) and (50), we get

$$\psi_0 = g_{1,0} \tau_{0,0} = 1$$

$$\begin{aligned}
\psi_{i>0} &= g_{1,i} \tau_{i,0} = \frac{2i-3}{2i} e^2 g_{1,i-1} \frac{2i}{2i+1} \tau_{i-1,0} \\
&= \frac{2i-3}{2i+1} e^2 \psi_{i-1}
\end{aligned}$$

Extending the sequence for $i > 3$ results in the reduction in odd numbers in the numerator and denominator.

$$\begin{aligned}
\psi_1 &= \frac{-1}{3} \psi_0 = -\frac{1}{3}, \\
\psi_2 &= \frac{1}{5} \psi_1 = -\frac{1}{3 \cdot 5}, \\
\psi_3 &= \frac{3}{7} \psi_2 = -\frac{3}{3 \cdot 5 \cdot 7} = -\frac{1}{5 \cdot 7}
\end{aligned}$$

Hence, the final value of the constant

$$\begin{aligned}
t_0 &= 1 - \frac{e^2}{1 \cdot 3} - \frac{e^4}{3 \cdot 5} - \frac{e^6}{5 \cdot 7} - \dots \\
&\quad - \frac{e^{2n}}{(2n-1)(2n+1)} - \dots
\end{aligned} \quad (56)$$

For the oblate ellipsoid ($e^2 > 0$) the constant is determined with the well-known Mollweide's formula (Mollweide 1805; Biernacki 1949, p.254;p.266)

$$t_0 = \frac{1}{2} + \frac{(1-e^2)}{4e} \ln \left(\frac{1+e}{1-e} \right)$$

and for the prolate ellipsoid ($e^2 < 0$) the constant is determined with Wrinch's (1932) formula

$$t_0 = \frac{1}{2} + \frac{\sin^{-1} \epsilon}{2\epsilon \sqrt{1-\epsilon^2}}; \text{ where } \epsilon^2 = \frac{b^2-a^2}{b^2} = \frac{-e^2}{1-e^2}$$

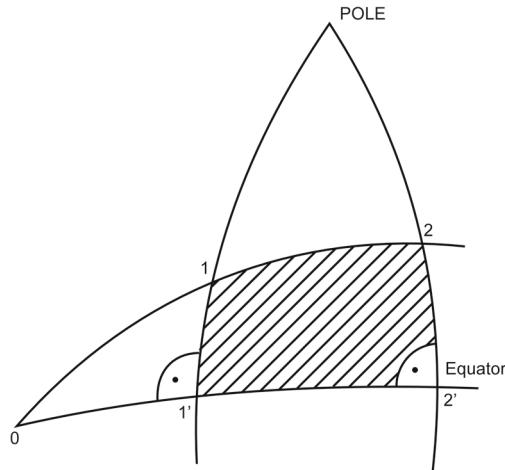


Fig. 3 The area under the geodesic (hatched area) and the equatorial geodesic triangle (1,0,1' or 2,0,2' figure)

Obtaining the exact values of the constant coefficients t_k is the key aspect of the theory of the area on the ellipsoid of revolution. The function $T(\rho)$ (defined at the beginning of Chapter 5) makes direct determination of the area of a geographic rectangle (A_{rec}) possible as:

$$\begin{aligned}
A_{\text{rec}}(\rho_1 < \rho < \rho_2, \lambda_1 < \lambda < \lambda_2) \\
= a^2(\lambda_2 - \lambda_1)(T(\rho_2) - T(\rho_1))
\end{aligned} \quad (57)$$

and, consequently, the area of an ellipsoid (denoted as A_{ell}) composed of two geographic rectangles

$$\begin{aligned}
A_{\text{ell}} &= 2 A_{\text{rec}}(0 < \rho < 1, -\pi < \lambda < \pi) \\
&= 4\pi a^2 (T(0) - T(1)) = 4\pi a^2 t_0
\end{aligned} \quad (58)$$

5.1 Area of the equatorial geodesic triangle

Let us reformulate Eq. (14) as

$$d\lambda = \frac{-C}{\sqrt{1-\rho^2} \sqrt{\rho^2-C^2}} \sum_{i=0}^{\infty} g_{1,i} \rho^{2i-1} d\rho \quad (59)$$

Substituting Eqs. (51) and (59) into (43), we will get the area of the equatorial geodesic triangle (denoted as A_{egt}), a geometry on an ellipsoid delimited by the geodesic, meridian and equator (e.g. shape 0,1,1', Fig. 3)

$$\begin{aligned}
A_{\text{egt}}(C, \rho_P) \\
&= -a^2 C \int_1^{\rho_P} \left(\sum_{j=0}^{\infty} t_j \rho^{2j} \right) \left(\sum_{i=0}^{\infty} g_{1,i} \rho^{2i-1} \right) \frac{d\rho}{\sqrt{\rho^2-C^2}} \\
&= a^2 A(C, \rho_P)
\end{aligned} \quad (60)$$

where $A(C, \rho_P)$ is the normalised area of the equatorial geodesic triangle that we use subsequently.

After multiplying the two series in (60) and ordering using the variable $k = i + j$, we get

$$A(C, \rho) = \sum_{k=0}^{\infty} \left(\sum_{j=0}^k t_j g_{1,k-j} \right) \left(-C \int_1^{\rho_P} \rho^{2k-1} \frac{d\rho}{\sqrt{\rho^2 - C^2}} \right) \quad (61)$$

Then, denoting

$$g_{2,k} = \sum_{j=0}^k t_j g_{1,k-j} \quad (62)$$

$$G_{2,k}(C, \rho_P) = -C \int_1^{\rho_P} \rho^{2k-1} \frac{d\rho}{\sqrt{\rho^2 - C^2}} = I_k(\rho_P) - I_k(1) \quad (63)$$

we get the area of interest as a linear combination of basic integrals

$$A(C, \rho_P) = \sum_{k=0}^{\infty} g_{2,k} G_{2,k}(C, \rho_P) \quad (64)$$

Let us calculate the integral

$$I_k(\rho) = -C \int \rho^{2k-1} \frac{d\rho}{\sqrt{\rho^2 - C^2}}$$

We convert the integrand to a form convenient for integration by parts (aiming at obtaining the derivative of the radicand in the numerator). We start with

$$I_k(\rho) = I_k = -C \int \rho^{2k-2} \frac{\rho d\rho}{\sqrt{\rho^2 - C^2}}$$

Then, integrating by parts, we get

$$\begin{aligned} I_k &= -C \rho^{2k-2} \sqrt{\rho^2 - C^2} + C(2k-2) \int \rho^{2k-3} \frac{\rho^2 - C^2}{\sqrt{\rho^2 - C^2}} d\rho \\ &= -C \rho^{2k-2} \sqrt{\rho^2 - C^2} + C(2k-2) \int \frac{\rho^{2k-1} - C^2 \rho^{2k-3}}{\sqrt{\rho^2 - C^2}} d\rho \\ &= -C \rho^{2k-2} \sqrt{\rho^2 - C^2} + (2k-2)(-I_k + C^2 I_{k-1}) \end{aligned}$$

Hence, the recursive formula for the integral I_k is

$$I_k = \frac{-C \rho^{2k-2} \sqrt{\rho^2 - C^2} + (2k-2)C^2 I_{k-1}}{2k-1} \quad (65)$$

Since the integral I_1 does not depend on I_0 (i.e. $I_1 = -C \sqrt{\rho^2 - C^2}$), the subsequent integrals have the form

$$I_{k>0} = C \sqrt{\rho^2 - C^2} \theta_k(C^2, \rho^2) \quad (66)$$

where the function θ_k is described by recursive formula

$$\theta_k = \frac{-\rho^{2k-2} + (2k-2)C^2 \theta_{k-1}}{2k-1} \quad (67)$$

Substituting (66) into (63), we get

$$\begin{aligned} G_{2,0}(C, \rho_P) &= -C \int_1^{\rho_P} \frac{d\rho}{\rho \sqrt{\rho^2 - C^2}} \\ &= \sin^{-1}\left(\frac{C}{\rho_P}\right) - \sin^{-1}(C) = \alpha - \alpha_0 \end{aligned}$$

$$\begin{aligned} G_{2,k>0}(C, \rho_P) &= I_k(\rho_P) - I_k(1) \\ &= C \left(\sqrt{\rho_P^2 - C^2} \theta_k(C, \rho_P) - \sqrt{1 - C^2} \theta_k(C, 1) \right) \\ &= \sin(\alpha) \cos(\alpha) \rho_P^2 \theta_k(C, \rho_P) \\ &\quad - \sin(\alpha_0) \cos(\alpha_0) \theta_k(C, 1) \end{aligned} \quad (68)$$

where $\alpha_0 = \sin^{-1}(C)$ is the azimuth under which a geodesic crosses the equator.

Substituting them subsequently into (64), we get

$$\begin{aligned} A(C, \rho_P) &= \sum_{k=0}^{\infty} g_{2,k} G_{2,k}(C, \rho_P) \\ &= g_{2,0} G_{2,0}(C, \rho_P) + \sum_{k=1}^{\infty} g_{2,k} G_{2,k}(C, \rho_P) \\ &= g_{2,0} (\alpha - \alpha_0) \\ &\quad + \sin(2\alpha) \frac{\rho_P^2}{2} \sum_{k=1}^{\infty} g_{2,k} \theta_k(C^2, \rho_P^2) \\ &\quad - \sin(2\alpha_0) \frac{1}{2} \sum_{k=1}^{\infty} g_{2,k} \theta_k(C^2, 1) \end{aligned}$$

And introducing the auxiliary function (Ar), which is the primitive function of the determined integral (60),

$$Ar(C^2, \rho_P^2) = \frac{\rho_P^2}{2} \sum_{k=1}^{\infty} g_{2,k} \theta_k(C^2, \rho_P^2) \quad (69)$$

Table 2 Estimated values of constant coefficients defined in Eqs. (13) and (62)

k	$g_{1,k}$	$g_{2,k}$
0	$1.00 \cdot 10^0$	$9.98 \cdot 10^{-1}$
1	$-3.35 \cdot 10^{-3}$	$-4.46 \cdot 10^{-3}$
2	$-5.60 \cdot 10^{-6}$	$-2.97 \cdot 10^{-6}$
3	$-1.88 \cdot 10^{-8}$	$-1.14 \cdot 10^{-8}$
4	$-7.85 \cdot 10^{-11}$	$-5.08 \cdot 10^{-11}$
5	$-3.68 \cdot 10^{-13}$	$-2.47 \cdot 10^{-13}$
6	$-1.85 \cdot 10^{-15}$	$-1.27 \cdot 10^{-15}$
7	$-9.71 \cdot 10^{-18}$	$-6.82 \cdot 10^{-18}$

we will get the normalised area of the equatorial geodesic triangle

$$\begin{aligned} A(C, \rho_P) = & g_{2,0} (\alpha - \alpha_0) + \sin(2\alpha) \operatorname{Ar}(C^2, \rho_P^2) \\ & - \sin(2\alpha_0) \operatorname{Ar}(C^2, 1) \end{aligned} \quad (70)$$

The use of the Ar function defined by the series (69), with fixed coefficients (62) and terms defined recursively (67), distinguishes formula (70) from the approximate formulas of Danielsen (1989) and Karney (2013) that use the coefficients defined by finite series.

The estimated values of constant coefficients (Table 2) indicate that the constants $g_{1,k} \approx e^{-2k}$. These eight coefficients are sufficient to determine the surface area of the Earth's ellipsoid ($a^2 \approx 4,068 \cdot 10^{13} [\text{m}^2]$) with an accuracy of 1 cm², as long as the resolution of the floating point implementation allows it.

Since new formulas have been obtained, let us test them on the example of the surface area of the ellipsoid.

A biangle is a boundary case of an equatorial triangle. It is formed when the geodesic reaches the equator having crossed the turning point, then $\alpha = \pi - \alpha_0$. For the meridian $C = 0$; $\alpha_0 = 0$; the biangle holds $\frac{1}{4}$ of the ellipsoid. This observation allows to check the calculations by computing the ellipsoid area using the following alternative formula

$$A_{\text{ell}} = 4\pi g_{2,0} a^2 \quad (71)$$

Based on formula (62), we obtain $g_{2,0} = t_0$, $g_{1,0} = t_0$, which in turn confirms that the result is consistent with formula (58).

5.2 Area under the geodesic

Danielsen (1989) called the area between a geodesic and the equator and the two meridians connecting to the end points of the arc of geodesic (Fig. 3) the area under the geodesic.

It should be noted that the area under the geodesic connecting points 1 and 2 on an ellipsoid (A_{12}) can be defined as the difference between the areas of two equatorial triangles, i.e. 2,0,2' and 1,0,1' (Fig. 3). Hence

$$\begin{aligned} A_{12} = & A(C, \beta_2) - A(C, \beta_1) \\ = & g_{2,0} \alpha_{12} + \sin(2\alpha_2) \operatorname{Ar}(C^2, \rho_2^2) \\ & - \sin(2\alpha_1) \operatorname{Ar}(C^2, \rho_1^2) \end{aligned} \quad (72)$$

The sign of the area depends on the arc orientation, i.e. $A_{21} = -A_{12}$. The area is positive for arcs oriented according to the direction of the azimuth increase. The accuracy of the area calculation mainly depends on the accuracy of the difference between the reverse and forward azimuths $\alpha_{12} = \alpha_2 - \alpha_1$ and therefore it should be calculated directly in the polar triangle.

Similarly to formulas (24) and (25), formula (72) shows the effect of the change of the polar axis on the deformation of the sphere, except that the main effect of the flattening f is transferred proportionally by the coefficient $g_{2,0}$. The azimuth difference α_{12} changes stepwise on the meridian, i.e. it takes the value π if points 1,2 are on either side of the pole, otherwise it equals 0. It follows that the pole causes problems in calculating the area.

5.3 Area of a geodesic polygon

The area of the geodesic polygon (A_{gp}) equals the total areas under the geodesics, which mark its border. Hence

$$\begin{aligned} A_{\text{gp}}(1, 2, 3, \dots, n, 1) = & (A_{12} + A_{23} + \dots + A_{n1}) a^2 \\ = & a^2 \sum_{i=1}^n A_{i,i+1} \end{aligned} \quad (73)$$

where n stands for the number of polygon vertices, and in the last component of the sum, the polygon vertex $n+1$ is equal to the first vertex ($n+1 \underset{\text{def}}{=} 1$). The surface area is positive if the polygon's border is travelled clockwise, i.e. according to the direction of the azimuth increase.

Let us analyse in detail the area of the geodesic triangle shown in Fig. 4. It can be seen from the figure that

$$\begin{aligned} A_{\text{gp}}(1, 2, 3, 1) = & A_{\text{gp}}(1, 2, 2', 1', 1) + A_{\text{gp}}(2, 3, 3', 2', 2) \\ & - A_{\text{gp}}(1, 3, 3', 1', 1) \end{aligned}$$

While, by changing the last area description to anticlockwise, we get

$$\begin{aligned} A_{\text{gp}}(1, 2, 3, 1) = & A_{\text{gp}}(1, 2, 2', 1', 1) \\ & + A_{\text{gp}}(2, 3, 3', 2', 2) + A_{\text{gp}}(3, 1, 1', 3', 3) \\ = & (A_{12} + A_{23} + A_{31}) a^2 \end{aligned}$$

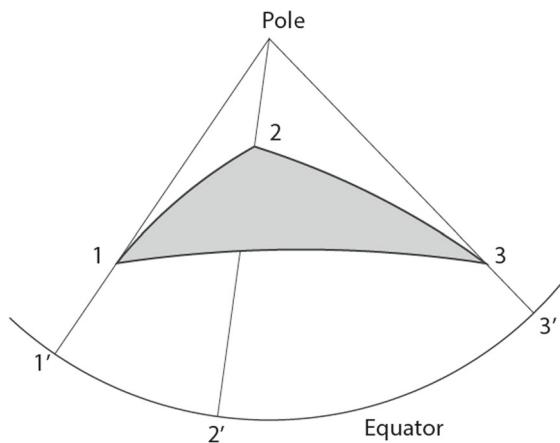


Fig. 4 The area of a geodesic triangle (grey area)

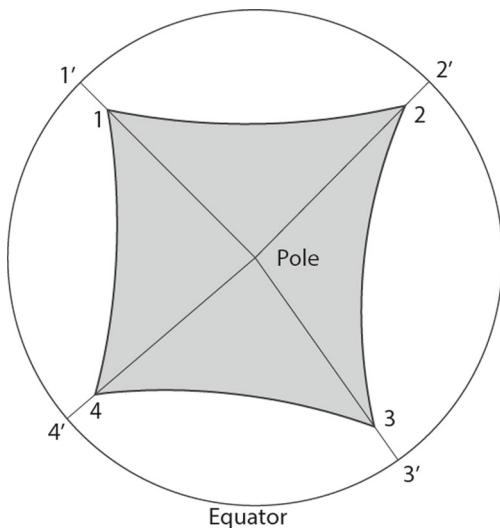


Fig. 5 A circumpolar polygon (grey area)

which is in accordance with formula (73).

Worth analysing also is the case of the geodesic polygon delimiting the area, which includes the pole, namely a circumpolar polygon (Fig. 5). Applying formula (73) to the circumpolar polygon, we obtain the following negative result:

$$\begin{aligned} A_{gp}(1, 2, 3, 4, 1) &= -A_{gp}(1, 2, 2', 1', 1) - A_{gp}(2, 3, 3', 2', 2) \\ &\quad - A_{gp}(3, 4, 4', 3', 3) - A_{gp}(4, 1, 1', 4', 4) \end{aligned}$$

It is the only case when we get a negative result travelling along the polygon circumference according to the direction of the azimuth increase. As a matter of fact, this value applies to an area, which completes half of the ellipsoid area (the area marked in white in Fig. 5). While the area of the circumpolar polygon (A_{cmp}) is defined as:

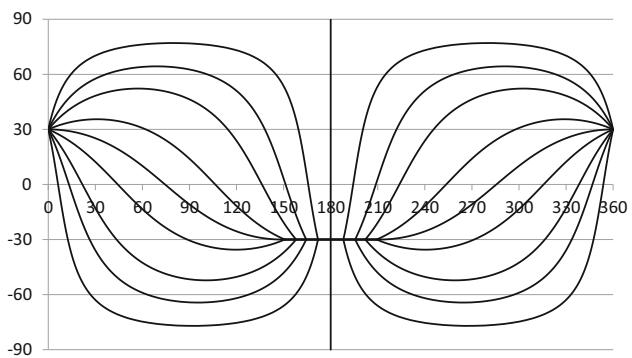


Fig. 6 A bundle of orthodromic geodesics starting in the origin point creates a line of bifurcation nodes on the circle of latitude passing through the antipodal point. The Plate Carrée projection (normal cylindrical, equidistant along the meridians) was applied; the cylinder is developed by unrolling from a cut along the image of the meridian passing through the starting point of the bundle

$$A_{cmp}(1, 2, 3, \dots, n, 1) = \frac{1}{2} A_{ell} + a^2 \sum_{i=1}^n A_{i,i+1} \quad (74)$$

6 Orthodromic path of a geodesic on an ellipsoid: The bifurcation phenomenon

Analysing the path of a geodesic connecting two equator points lying on the antipodes, one can see that two geodesics connect these two points, i.e. the equator and a meridian. On an oblate ellipsoid of revolution, the curve of the meridian is shorter and constitutes an orthodrome. The problem of the unequivocalness of a geodesic line, therefore, comes down to the question of when a geodesic constitutes an orthodrome.

6.1 Bifurcation of Clairaut's equation. Bifurcate antipodal nodes

Considering that a geodesic is defined by Clairaut's equation ($\rho \sin \alpha = C$), it can be noted that there are two corresponding geodesics with the same parameter C , passing through a given point with β_0, λ_0 coordinates. Their azimuths are $\alpha_1 = \sin^{-1}(C/\rho_0)$, $\alpha_1' = \pi - \alpha_1$, respectively. According to the theory of differential equations (the azimuth is a function of the first derivative, see formula (2)), this phenomenon is termed bifurcation (Kusza and Bailey 1982).

Let us analyse the path of geodesic lines in Fig. 6:

- The geodesics, with the initial azimuth $-\pi/2 < \alpha_1 < \pi/2$, rise towards the north to reach the northern turning point for a minimal radius of $\rho = |C|$, after which they descend.

- ii. The other geodesics descend towards the south to reach the southern turning point for a minimal radius of $\rho = |C|$, after which they rise.
- iii. The corresponding geodesic lines intersect in the nodes lying on the opposite parallel $(-\beta_0)$ on the other (antipodal) side of the equator plane; the curve of each of the lines corresponds to half a period.

Starting from the origin point, the geodetic arcs:

- i. smaller than half of the period, do not intersect and are orthodromes, because they uniquely determine the shortest path from the origin point to a given point of the ellipsoid;
- ii. equal to half a period, intersect at the nodes with their bifurcation counterparts, i.e. they constitute two arcs of the same length from the origin point to the node, and both are orthodromes;
- iii. greater than half a period, are not orthodromes because they always intersect with the above-described orthodromes covering the entire ellipsoid.

6.2 Bifurcation line of the nodes for an oblate ellipsoid

The nodes of all geodesics coming from a specific point form a bifurcation line assigned to the point, shown in Fig. 6 as the bold part of the parallel. The bifurcation line is the geodesic uniqueness border: the starting point and a node link two orthodromes with the same length and azimuths, which complete π ; for instance $\alpha_1 = 30^\circ$, $\alpha_2 = 150^\circ$ and $\alpha_{1'} = 150^\circ$, $\alpha_{2'} = 30^\circ$.

The beginning of the bifurcation line lies at the turning point of the geodesic with azimuths $\alpha_1 = 90^\circ$, $\alpha_2 = 90^\circ$; and its parameter $C = \rho_0$. Substituting it into formula (39), we will get

$$\lambda_{\min} = \pi \left(1 + |C| Ls(C^2) \right)$$

Hence, the Boolean variable, which determines whether bifurcation occurs, is determined as:

$$BIF = (\beta_2 = -\beta_1) \text{ and } \left(|\lambda_{12}| \geq \pi \left(1 + |C| Ls(C^2) \right) \right) \quad (75)$$

Let us consider the case of the bifurcation $\sigma_{12} = \omega_{12} = \pi$ at the equator. Then $C = \pm 1$, $\alpha_1 = \alpha_2 = \pm \frac{\pi}{2}$, $E_1 = E_2 = 0$, $Ls = -f$, $Ss = 1 - f$. And from Eq. (40), we obtain $s_{12} = aS_{12} = a(1 - f)\pi = b\pi$. It follows that the equator is an orthodrome, provided its curve segment does not exceed the length of a semicircle with a radius equal

to the semi-minor axis of an ellipsoid. This can be illustrated by imagining an ellipsoid in which the semi-minor axis decreases to zero. In such a case, the ellipsoid becomes a flat disc, and the shortest line between the points on the equator becomes a chord.

6.3 Bifurcation line of the nodes for prolate ellipsoid

In case the equator semi-axis is smaller than the polar semi-axis, the half-period is greater than π , since the parameter f for such an ellipsoid is negative, as can be seen from Eq. (10). Before reaching the parallel $-\beta_0$, the geodesic crosses the symmetrical geodesic with the azimuth $\alpha = -\alpha_1$. The intersection point lies on the antipodal meridian $\lambda_0 + \pi$, which is a bifurcation line for the prolate ellipsoid ($e^2 < 0$).

If the semi-axes are equal, the bifurcation line is reduced to the antipodal point, where all geodesics of a bundle meet.

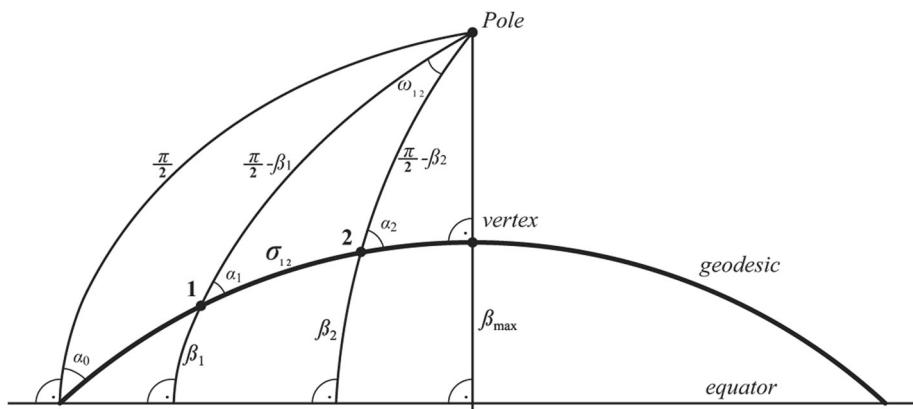
7 The inverse problem extended with the area under the geodesic

The analysis of the problematic cases of the inverse problem, i.e. those involving the antipodes and the turning points of a geodesic, revealed issues related to the choice of parameters constituting the main equation, which manifest themselves as: (i) the lack of monotonicity of the function forming the equation; (ii) the consequent limitation of the applicability of Newton's method, which requires the existence of a nonzero derivative; and (iii) the difficulty of determining the initial estimate for the iterative process.

The algorithms for solving the inverse problem are numerical solutions of an equation with one unknown. The classical methods of Bessel (1826) and their modifications, including the popular Vincenty's (1975) formulas, use ω_{12} as a working variable, while Karney (2013) assumes, after Helmert (1880), that the azimuth α_1 constitutes a variable of the main equation. None of the methods used to date solves all cases of the inverse problem, and the reason for this is the selection of the controlled variable used in the iterations. The methods based on ω_{12} variable encounter problems in near-antipodal areas, while the methods based on α_1 , do not provide a solution if the case involves the area near the turning point of a geodesic. Therefore, we will revise the parameterisation of the indirect problem with the aim to obtain solutions in all cases. We will use two functionally connected variables for parameterisation. The inverse problem is symmetrical with respect to the end points, consequently, the algorithm should not prefer any of them.

In our algorithm, the initial data of our extended inverse geodetic problem are the same as for the classical inverse geodetic problem (i.e. latitude and longitude of two points: $\varphi_1, \lambda_1, \varphi_2, \lambda_2$), while the outputs (i.e. $\alpha_1, \alpha_2, s_{12}, Ar_{12}$,

Fig. 7 Polar spherical triangle 1,2,P and path of the geodesic. The coordinate system of the sphere is defined by longitude ω and latitude β



Type, *BIF*) additionally include the numerical value of the surface area under geodesic (Ar_{12}), the *Type* variable specifying the result of dynamic parameter choice, taking the value C or S (discussed further in ch.7.1) and a Boolean flag indicating the occurrence of bifurcation (*BIF*).

7.1 Analysis of the inverse problem. Novel parameterisation

Considering the above reasoning, we decided to base the parameterisation of the inverse geodetic problem on two variables, namely ω_{12} and C . The former is Bessel's (1826) well-known variable, and it represents the longitude on the auxiliary sphere (Fig. 7). It has been used by most authors of geodetic algorithms, including Vincenty (1975).

In contrast, the latter variable (i.e. C) was introduced by Helmert (1880) as: $C = \rho_{\min}$. The geometric interpretation of this variable is shown in Fig. 7 and is described by the following formula.

$$C = \rho_{\min} = \sin \alpha_0 = \cos \beta_{\max}$$

Karney (2013) and Pittman (1986) used variants of this variable. Karney replaced it by α_0 that $\alpha_0 = \sin^{-1} C$. The variable $\cos \beta_{\max}$ was used by Pittman (1986), denoted as $\cos \psi_0$.

The fundamental equation is composed of the given values difference (i.e. $\lambda_{12} = \lambda_2 - \lambda_1$) and its value is expressed by formula (39)

$$\delta = \lambda_{12} - \omega_{12} - C \left(\sigma_{12} Ls(C^2) + E_2 Le_2 - E_1 Le_1 \right) = 0 \quad (76)$$

Equation (76) is interpreted differently depending on the type of relations between the C and ω_{12} variables. If bifurcation occurs, then the spherical triangle (Fig. 7) degenerates into a digon; $\sigma_{12} = |\omega_{12}| = \pi$ and the main equation becomes a function of the C parameter. The inverse problem has two solutions that meet Clairaut's relation. Such a

case is assigned the discriminant *BIF* = 1 in the output of the authors' implementations. Otherwise, the polar spherical triangle links C and ω_{12} . Then, the inverse problem has one solution, and it is assigned the discriminant *BIF* = 0.

The fundamental Eq. (76) includes three variables C , ω_{12} , σ_{12} , which describe a polar spherical triangle (Fig. 7). In order to solve all cases, we need formulas which describe a biangle as a boundary case for triangle $\sigma_{12} = |\omega_{12}| = \pi$. We use a sine formula for polar spherical triangle converted into the form of Clairaut's equation's

$$\rho_1 \sin(\alpha_1) = \rho_2 \sin(\alpha_2) = \frac{\rho_1 \rho_2 \sin(\omega_{12})}{\sin(\sigma_{12})} = C \quad (77)$$

hence

$$C \sin(\sigma_{12}) - \rho_1 \rho_2 \sin(\omega_{12}) = 0 \quad (78)$$

and cosine formula for the same triangle

$$\cos(\sigma_{12}) = \zeta_1 \zeta_2 + \rho_1 \rho_2 \cos(\omega_{12}) \quad (79)$$

An inverse problem will be solved iteratively with Newton's method. The creation of a system of equations for two variables C , ω_{12} requires differentiation of Eqs. (76), (78) and (79) eliminating σ_{12} . Having differentiated (79), we therefore get

$$\sin(\sigma_{12}) d\sigma_{12} = \rho_1 \rho_2 \sin(\omega_{12}) d\omega_{12}$$

After substituting (77), we will get

$$d\sigma_{12} = Cd\omega_{12} \quad (80)$$

This equation allows for elimination of $d\sigma_{12}$ from the other equations.

We develop the fundamental Eq. (76) into a Taylor's series, neglecting the terms of the second and higher orders:

$$\delta(C + dC, \omega_{12} + d\omega_{12}) \cong \delta(C, \omega_{12}) \\ + \frac{\partial \delta}{\partial C} dC + \frac{\partial \delta}{\partial \omega_{12}} d\omega_{12} = 0$$

Hence

$$d_1 dC + d_2 d\omega_{12} = \delta \quad (81)$$

And the details of the differentiation are presented below.
Let start with variable C .

$$d_1 = -\frac{\partial \delta}{\partial C} = (Ls\sigma_{12} + Le_2 E_2 - Le_1 E_1) + C\sigma_{12} \frac{\partial Ls}{\partial C}$$

From formula (26) we get

$$Ls(C^2) \cong g_{1,1} + g_{1,2} \frac{1+C^2}{2} + g_{1,3} \frac{3+2C^2+3C^4}{8}$$

$$\frac{\partial Ls}{\partial C} = Cg_{1,2} + \frac{(C+3C^3)}{2}g_{1,3}$$

Finally

$$d_1 = (Ls\sigma_{12} + Le_2 E_2 - Le_1 E_1) \\ + C^2\sigma_{12} \left(g_{1,2} + \frac{(1+3C^2)}{2}g_{1,3} \right) \quad (82)$$

In turn, differentiating with respect to ω_{12} , yields

$$d_2 = -\frac{\partial \delta}{\partial \omega_{12}} = 1 + CLs \frac{d\sigma_{12}}{d\omega_{12}}$$

And after substituting (80), we get

$$d_2 = -\frac{\partial \delta}{\partial \omega_{12}} = 1 + C^2 Ls \quad (83)$$

Derivative d_2 is close to zero and greater than d_1 by two orders of magnitude. It gives clear preferences towards variable ω_{12} and, consequently, allowed the use of direct iterations in Vincenty's method. Unfortunately, Vincenty's approach turned out to have some shortcomings in respect to near-antipodal arcs.

Differentiating Eq. (78) and substituting (80), we will get

$$h_1 dC + h_2 d\omega_{12} = 0 \quad (84)$$

$$h_1 = \sin(\sigma_{12}) \quad (85)$$

$$h_2 = C \cos(\sigma_{12}) \frac{d\sigma_{12}}{d\omega_{12}} - \rho_1 \rho_2 \cos(\omega_{12})$$

$$= C^2 \cos(\sigma_{12}) - \rho_1 \rho_2 \cos(\omega_{12}) \quad (86)$$

Equations (81) and (84) form a Newton's method system of equations for unknowns dC , $d\omega_{12}$.

In order to investigate Eq. (86), we will use formula (3)

$$h_2 = C^2 \cos(\sigma_{12}) - \rho_1 \rho_2 \cos(\omega_{12}) \\ = \rho_1 \rho_2 (\sin(\alpha_1) \sin(\alpha_2) \cos(\sigma_{12}) - \cos(\omega_{12}))$$

and the cosine formula for the angles of polar spherical triangle (Fig. 7) ($\cos(\omega_{12}) = \cos(\alpha_1) \cos(\alpha_2) + \sin(\alpha_1) \sin(\alpha_2) \cos(\sigma_{12})$), then

$$h_2 = -\rho_1 \rho_2 \cos(\alpha_1) \cos(\alpha_2) \quad (87)$$

Equation (84) has the following specific characteristics:

- i. Coefficient h_1 equals zero in the case of bifurcation, resulting in $d\omega_{12} = 0$;
- ii. Coefficient h_2 equals zero in the case when an end point of geodesic arc coincides with its turning point or when the geodesic arc coincides with the equator, resulting in $dC = 0$;
- iii. Both coefficients h_1 , h_2 are equal to zero, i.e. the equation is identically satisfied for all values of $d\omega_{12}$ and dC , when both end points of geodesic arc coincide with its turning points, but this is an extreme case of bifurcation, i.e. $d\omega_{12} = 0$, and Eq. (81) is sufficient.

The solution of the system of Eqs. (81), (84)

$$D = d_1 h_2 - d_2 h_1; \\ \text{if } D = 0 \text{ then } dC = \frac{-\delta}{d_1} \text{ else } dC = h_2 \frac{\delta}{D}; \\ d\omega_{12} = -h_1 \frac{\delta}{D} \quad (88)$$

reveals a direct impact of the coefficients of homogeneous Eq. (83) on the values of the unknowns. In practice, only the unknown with the greater effective impact on the free term of Eq. (80) is determined.

$$|h_1 d_2| < |h_2 d_1| \quad (89)$$

If inequality (89) is true, dC is determined (and $Type = C$ assigned to the solution) or, otherwise, $d\omega_{12}$ is determined ($Type = S$). The choice of the dynamic (adaptive) parameter determines the way the polar triangle is defined, and results in fewer iterations.

After inserting the approximate values of d_1 , d_2 , the inequality takes the following form

$$\sin(\sigma_{12}) < |h_2|f\sigma_{12}$$

It is true for $\sigma_{12} \approx \pi$, then the estimation turns into

$$\pi - \sigma_{12} < |h_2|f\sigma_{12}$$

hence

$$\sigma_{12} > \frac{\pi}{1 + |h_2|f} \quad (90)$$

This means that the *Type C* solution is more effective in near-antipodal cases of an inverse problem.

The adaptative parameter choice was verified numerically. The numerical derivative of $\frac{d\lambda_{12}}{dC}$ was found to be equal to $\frac{\delta}{dC}$, and the numerical derivative of $\frac{d\lambda_{12}}{d\omega_{12}}$ was found to be equal to $\frac{\delta}{d\omega_{12}}$.

The Newton's iterative process is stopped when $|\delta| \leq \varepsilon$, where ε is an arbitrary stop-criterion. Then, in the last stage of our extended inverse problem, the distance and area are calculated according to formulas (40) and (73).

The (81) and (84) system of equations is an indirect solution of the complex theoretical problem of the reduced length of the geodesic that was analysed by Gauss (1828) and Christoffel (1868). For example, Helmert's (1880) solution, later used by Karney (2013), requires laborious calculation of a specific series twice.

7.2 Two types of polar spherical triangles

The one type of polar spherical triangle (*Type = S*) is defined by the determined value of ω_{12} and the constant values β_1 , β_2 . This is a classical problem of spherical trigonometry, and its solution will be based on Napier's rules that enable direct determination of α_{12} . If in any spherical triangle, two sides a , c are given and the angle B between them, then the cotangential formulas define the angles as follows:

$$\cot(A) = \frac{\cos(a)\sin(c) - \sin(a)\cos(c)\cos(B)}{\sin(a)\sin(B)} = \frac{q_1}{q_2} \quad (91)$$

$$\cot(C) = \frac{\cos(c)\sin(a) - \sin(c)\cos(a)\cos(B)}{\sin(c)\sin(B)} = \frac{q_3}{q_4} \quad (92)$$

It results from the sine formula that $q_2 = \sin(b)\sin(A)$. While from the 'five-part rule' of spherical trigonometry, it results that $q_1 = \sin(b)\cos(A)$ hence, $\sin(b) = \sqrt{q_1^2 + q_2^2} = \sqrt{q_3^2 + q_4^2} = q_5$. This allows the calculation of the missing

triangle side:

$$\tan(b) = \frac{q_5}{\cos(a)\cos(c) + \sin(a)\sin(c)\cos(B)} \quad (93)$$

Using these formulas for a polar triangle, we get

$$\tan(\sigma_{12}) = \frac{q_5}{\zeta_1\zeta_2 + \rho_1\rho_2\cos(\omega_{12})} \quad (94)$$

$$q_1 = \rho_2\sin(\omega_{12}),$$

$$q_2 = \rho_1\zeta_2 - \rho_2\zeta_1\cos(\omega_{12}),$$

$$q_5 = \sqrt{q_1^2 + q_2^2} \quad (95)$$

We get the parameter of the geodesic from modified formula (77)

$$C = \frac{\rho_1\rho_2\sin(\omega_{12})}{\sin(\sigma_{12})} = \frac{\rho_1\rho_2\sin(\omega_{12})}{q_5} \quad (96)$$

While, the azimuths and their difference will be derived directly from Napier's rules:

$$\tan\left(\frac{\alpha_1 + \alpha_2}{2}\right) = \frac{\cos\left(\frac{\beta_1 + \beta_2}{2}\right)\sin\left(\frac{\omega_{12}}{2}\right)}{\sin\left(\frac{\beta_2 - \beta_1}{2}\right)\cos\left(\frac{\omega_{12}}{2}\right)} \quad (97)$$

$$\tan\left(\frac{\alpha_2 - \alpha_1}{2}\right) = \frac{\sin\left(\frac{\beta_1 + \beta_2}{2}\right)\sin\left(\frac{\omega_{12}}{2}\right)}{\cos\left(\frac{\beta_2 - \beta_1}{2}\right)\cos\left(\frac{\omega_{12}}{2}\right)} \quad (98)$$

The second type of the polar spherical triangle is defined by two sides and the C parameter of the third side. It is used for near-antipodal regions and for bifurcation, where the triangle shape is better defined by the C parameter. The case is assigned *Type = C* since formula (3) allows the azimuths to be calculated using $\sin(\alpha_i) = \frac{C}{\rho_i}$.

In order to obtain an unambiguous solution, we check whether the polar triangle is obtuse. We analyse the cosine formula that determines the triangle side opposite to the azimuth

$$\begin{aligned} \cos\left(\frac{\pi}{2} - \beta_2\right) &= \zeta_2 = \zeta_1 \cos(\sigma_{12}) \\ &\quad + \sin\left(\frac{\pi}{2} - \beta_1\right) \sin(\sigma_{12}) \cos(\alpha_1) \end{aligned}$$

The second segment of the equation is negative, if $\alpha_1 > \frac{\pi}{2}$. Moving the first segment of the equation to the left side, we obtain the sign of the cosine of the missing angle:

$$m_1 = \sqrt{\rho_1^2 - C^2}; \text{ if } \zeta_2 - \zeta_1 \cos(\sigma_{12}) < 0 \text{ then } m_1 = -m_1 \quad (99)$$

$$m_2 = \sqrt{\rho_2^2 - C^2}; \text{ if } \zeta_1 - \zeta_2 \cos(\sigma_{12}) \geq 0 \text{ then } m_2 = -m_2 \quad (100)$$

Finally, the azimuths are determined from the formulas

$$\tan(\alpha_i) = \frac{C}{m_i} \quad (101)$$

and their difference from the direct formula

$$\tan(\alpha_2 - \alpha_1) = \frac{C(m_1 - m_2)}{m_1 m_2 + C^2} \quad (102)$$

Let us determine ω_1 from formula (33)

$$\begin{aligned} \sin(\omega_1) &= \frac{C \zeta_1}{\sqrt{1 - C^2} \rho_1} \\ \cos(\omega_1) &= \frac{\sqrt{(1 - C^2) \rho_1^2 - C^2 \zeta_1^2}}{\sqrt{1 - C^2} \rho_1} = \frac{\sqrt{\rho_1^2 - C^2}}{\sqrt{1 - C^2} \rho_1} \end{aligned}$$

After substituting formula (99)

$$\cos(\omega_1) = \frac{m_1}{\sqrt{1 - C^2} \rho_1}$$

Finally, we get

$$\tan(\omega_{12}) = \frac{C(m_1 \zeta_2 - m_2 \zeta_1)}{m_1 m_2 + C^2 \zeta_1 \zeta_2} \quad (103)$$

Hence, using formula (34)

$$\sin(\sigma_1) = \frac{\zeta_1}{\sqrt{1 - C^2}}$$

we get

$$\tan(\sigma_{12}) = \frac{m_1 \zeta_2 - m_2 \zeta_1}{m_1 m_2 + \zeta_1 \zeta_2} \quad (104)$$

7.3 Initial approximation

In order to obtain the starting point for Newton's method, we shall begin with $\lambda_{12} = \lambda_2 - \lambda_1$ normalisation so that $|\lambda_{12}| \leq \pi$. It determines the signs of the following values $\text{sign}(C) = \text{sign}(\omega_{12}) = \text{sign}(\lambda_{12})$. We calculate $C_{\max} = \min(\rho_1, \rho_2)$, $Ls_m = Ls(C_{\max})$, which enables the determination of the BIF variable value from formula (75). If

bifurcation occurs, then $\sigma_{12} = \pi$; $\omega_{12} = \pi \text{ sign}(\lambda_{12})$, otherwise σ_{12} is calculated from formula (93) assuming that $\omega_{12} = \lambda_{12}$.

The components of the fundamental Eq. (76) are estimated based on Table 1. Ls is the function of unknown C but it varies in very narrow ranges, and the extreme value $Ls = Ls_m$ is taken for the initiation. Le takes low values, and $Le = 0$ is assumed for the initiation. Thus substituting the fundamental equation simplified this way:

$$\omega_{12} = \lambda_{12} - Ls_m \sigma_{12} C \quad (105)$$

into (77), we get

$$C \sin(\sigma_{12}) = \rho_1 \rho_2 \sin(\lambda_{12} - Ls_m \sigma_{12} C)$$

Expanding the sine into a Taylor's series, neglecting the terms of the second and higher order one gets the relation

$$C \sin(\sigma_{12}) \cong \rho_1 \rho_2 \sin(\lambda_{12}) - Ls_m \sigma_{12} \rho_1 \rho_2 \cos(\lambda_{12}) C$$

Hence, the formula for the initial value of the C parameter is as follows:

$$\tilde{C} = \frac{\rho_1 \rho_2 \sin(\lambda_{12})}{\sin(\sigma_{12}) + Ls_m \sigma_{12} \rho_1 \rho_2 \cos(\lambda_{12})} \quad (106)$$

Although the approximation obtained is weaker than the 4th degree equation of Jacobi's (1891) astroid, the tests revealed that it does not affect the number of iterations (see Sect. 7.5). As a result, the proposed algorithm is characterised by relatively low computational complexity.

Using the value of the variable $C = \tilde{C}$ in formula (76), the approximation of the value of the second variable can be improved.

$$\tilde{\omega}_{12} = \lambda_{12} - \tilde{C} \left(\sigma_{12} Ls(\tilde{C}^2) + E_2 Le_2 - E_1 Le_1 \right) \quad (107)$$

7.4 Algorithm for calculating the surface area of a geodesic polygon

This pseudocode outlines how to calculate the area of a geodesic polygon, given a list of the geographic coordinates of its vertices. FUNCTION AreaPolygon uses FUNCTION AreaArc to calculate the area under the geodesic.

```

FUNCTION AreaArc( $\varphi_1, \varphi_2, \lambda_{12}$ )
IF  $\lambda_{12} < -180$  THEN  $\lambda_{12} = \lambda_{12} + 360$ 
IF  $\lambda_{12} > 180$  THEN  $\lambda_{12} = \lambda_{12} - 360$ 
Ros =  $Pi/180$ ;  $\lambda_{12}$  =  $\lambda_{12}$  Ros;  $\varphi_1$  =  $\varphi_1$  Ros;  $\varphi_2$  =  $\varphi_2$  Ros
 $\beta_1, \beta_2, \rho_1, \rho_2, \zeta_1, \zeta_2$  from (28-30)
 $C_{\max} = \min(\rho_1, \rho_2)$ ;  $L_{S_m} = Ls(C_{\max})^2$  from (26); BIF from (75)
IF BIF THEN
     $\sigma_{12} = \pi$ ;  $\omega_{12} = \pi \operatorname{sign}(\lambda_{12})$ 
    ELSE
         $\omega_{12} = \lambda_{12}; \sigma_{12}$  from (94)
    ENDIF
    TYPc=BIF; C from (106);
    IF abs(C) >  $C_{\max}$  THEN  $C = C_{\max} \operatorname{sign}(\lambda_{12})$ 
    IF not(BIF) THEN  $\omega_{12}, h1, h2, d1, d$ , TYPc from (105,83-86,89)
    it=0
    ITER:
    IF TYPc THEN
         $\alpha_{12}, \alpha_1, \alpha_2$  from (99-102)
        IF not(BIF) THEN  $\omega_{12}, \sigma_{12}$  from (103-104)
        ELSE
             $\sigma_{12}, C, \alpha_{12}, \alpha_1, \alpha_2$  from (94,96-98)
        ENDIF
         $\delta$  from (76;26,27,32);
        IF abs( $\delta$ ) > eps THEN
            h1, h2, d1, d2 from (83-87)
            IF not(BIF) THEN TYPc from (89)
             $dC, d\omega_{12}$  from (88)
            IF TYPc THEN  $C = C + dC$  ELSE  $\omega_{12} = \omega_{12} + d\omega_{12}$ 
            It=it+1; GOTO ITER
        ELSE
            AreaArc from (73;69)
        ENDIF
    ENDIF
FUNCTION AreaPolygon( $F_1, L_1, F_2, L_2, \dots, F_n, L_n$ )
 $F_{n+1} = F_1; L_{n+1} = L_1; S = 0$ 
FOR j=1 TO n
     $\varphi_1 = F_j; \varphi_2 = F_{j+1}; \lambda_{12} = L_{j+1} - L_j$ 
     $S = S + \operatorname{AreaArc}(\varphi_1, \varphi_2, \lambda_{12})$ 
ENDFOR
AreaPolygon=S

```

7.5 Comments on numerical application

The ellipsoid area amounts to $5.1 \cdot 10^{14}$. This means that in order to record only the numerical values of the area with $1m^2$ precision, we have to be able to use numbers with 15 decimal digits (49 binary digits). The common IEEE 754 Double Floating Point Format PC standard has a 52-bit mantissa, which is sufficient to record values with such high precision. Still, making such precise real calculations is difficult, which is illustrated by the following numerical relation in double precision arithmetic, which contradict mathematics.

Example. Inequality $1 + \varepsilon > 1$ is not met by $0 \leq \varepsilon \leq 2^{-52} \cong 2.22 \cdot 10^{-16}$.

7.6 Numerical tests

The method was verified by comparing the results obtained in IEEE 754 Double Floating arithmetic (52 bits of mantissa) with the results obtained in IEEE 754 Extended Floating Point arithmetic (63 bits of mantissa). The tests data and results are listed in Table 3. In the first column the geodetic latitudes of the given points and the longitude difference are displayed. The distance, the forward and reverse azimuths, i.e. the solution of the inverse problem, are displayed in the second column. The last column lists the value of the surface area under the geodesic obtained in double and extended arithmetic (denoted as ‘Area double’ and ‘Area extended’, respectively), the difference between these results (denoted as ‘diff’), the type specifying the result of dynamic parameter choice, taking the value C or S (as discussed in ch.7.1, and denoted as ‘Type’), bifurcation occurrence indicator (‘BIF’) and the number of iterations (‘Iter’).

The relative accuracy of the area ranges from $5 \cdot 10^{-14}$ for the longest arcs to $5 \cdot 10^{-12}$ for the shortest ones. The tests revealed that $1m^2$ accuracy is achieved for distances under 10,000 km. It needs to be emphasised that the presented innovative iterative algorithm is characterised by a low number of iterations (Iter, Table 3), compared to the number of iterations when Rapp’s (1993) method was applied. Consequently, a

Table 3 Numerical results of the extended inverse problem solved using the authors' algorithm. Ist column: the inverse problem input data; IIInd column: the results, i.e. distance and the forward and reverse azimuths; IIIrd column: the extension, where (Area) is the area under geodesics obtained using double and extended arithmetic, (diff) is their difference, (Type) is the parametrisation type, (Iter) is the number of iterations and (BIF) is the bifurcation occurrence indicator. Adopted ellipsoid: GRS80, $a = 6378137$ m, $f = 298.257222101$ (IAG and IAU recommended). Authors of test data: Karney (2013) – tests no. 3, 4 and 10; Nowak & Nowak Da Costa – remaining tests

	φ_1 [°]	s_{12} [m]	Area double [m ²]
	φ_2 [°]	α_1 [°]	Area extended [m ²]
	λ_{12} [°]	α_2 [°]	Type Iter BIF diff
1	-30	20,000,239,4375782	95,545,707,564,560,4375
	30	22,4966622933548	95,545,707,564,558,3438
	179.8	157,503337706645	C 2 BIF 2.0938
2	-60.0832522871723	19,995,538,1198384	0.0000
	60.0832522871723	90,0000000000000	0.0000
	179.698437014927	90,0000000000000	C 0 BIF 0.0000
3	-30	19,989,832,8274572	-101,790,744,815,440,5312
	29.9	161,890524809383	-101,790,744,815,440,5469
	179.8	18,0907371727648	C 3 0.0156
4	40,0000000000000	10,000,000,0000000	84,275,623,420,833,5938
	41,7933102054876	30,0000000000000	84,275,623,420,833,5469
	137,844900043235	149,090169317859	S 2 0.0469
5	0	9,997,769,05991920	42,426,932,221,844,8594
	60.0832522871723	30,0000000000000	42,426,932,221,844,8516
	89,8492185074635	90,0000000000000	S 2 0.0078
6	0	6,662,472,71812859	13,757,767,314,573,6738
	45	35,4100589057817	13,757,767,314,573,6738
	45	54,8907738286376	S 2 0.0000
7	49	771,856,420145830	-4,953,242,880,753,5127
	53	-51,3997986214807	-4,953,242,880,753,5127
	-9	-58,4040881483856	S 2 0.0000
8	49	1331,18994804168	5,336,632,538,4134
	49.01	33,3369708839879	5,336,632,538,4132
	0.01	33,3445185523180	S 1 0.0003
9	30	96,4862802512827	353,022,972,1751
	30	89,9997500000000	353,022,972,1740
	0.001	90,0002500000000	S 1 0.0011
10	-30.12345	4,94420828437249	-17,717,125,1039
	-30.12344	77,0435335454647	-17,717,125,1039
	0.00005	77,0435084522292	S 1 0.0000
11	25,7877777777778	1,670,050,03285709	5,340,270,554,894,0654
	32,3333333333333	60,6586569955724	5,340,270,554,894,0674
	15,4741666666667	68,2225774455895	S 2 -0.0020
12	32,3333333333333	1,570,002,92992898	-531,808,178,986,7905
	18,2500000000000	-173,161873198869	-531,808,178,986,7904
	-1,7500000000000	-173,915270017159	S 2 -0.0001
13	18,2500000000000	1,642,829,80344637	-3,654,170,119,224,9297
	25,7877777777778	-57,0496199543674	-3,654,170,119,224,9312
	-13,7241666666667	-62,2275665606725	S 2 0.0015

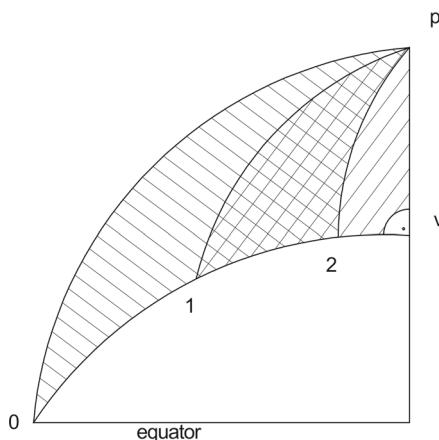


Fig. 8 Polar spherical triangles used in different approaches to solve the inverse geodetic problem

quick convergence of the proposed algorithm was achieved and proven.

We covered the calculation of the area of a geodesic polygon using the Bermuda Triangle as an example. The extended inverse problems listed in the last three rows of Table 3 (lines 11–13) allow the calculation of the area of a geodesic triangle: Miami $25^{\circ}47'16''\text{N}$, $80^{\circ}13'27''\text{W}$, Bermuda $32^{\circ}20'\text{N}$, $64^{\circ}45'\text{W}$, Puerto Rico $18^{\circ}15'\text{N}$, $66^{\circ}30'\text{W}$. The sum of the areas under successive geodesics equals $1,154,292,256,682 \text{ m}^2$.

Additionally, for the sake of comparison with the published results of the calculation of the area under the geodesic, i.e. $84,275,623.42235 \text{ km}^2$ given $\varphi_1 = 40^\circ$, $\alpha_1 = 30^\circ$ and $s_{12} = 10000\text{km}$ (Karney 2013), the calculations were made for ellipsoid WGS-84: $a = 6,378,137$, $1/f = 298.257223563$ (adopted by Karney as the reference surface) obtaining $84,275,623,422,354.4844 \text{ m}^2$ in double arithmetic and $84,275,623,422,354.4531 \text{ m}^2$ in extended arithmetic.

7.7 Comment on existing methods

The main distinctive feature of the methods described for solving the inverse geodetic problem on arc 1–2 (Fig. 8) is the employment of different polar spherical triangles for the same purpose. Namely, Karney's or Tseng's method uses two triangles, i.e. $(0, P, 1)$ and $(0, 2, P)$; Pittman's method uses both the $(1, P, V)$ and $(2, P, V)$ triangles; while our method, like Vincenty's, use only one, i.e. the target triangle, i.e. $(1, P, 2)$.

In those methods that use two triangles, in each iteration, the series to determine the longitude and distance for the two arcs are computed. Subtraction of the resulting large, often similar, numbers is then performed, resulting in a decrease in accuracy (number of significant digits). Additionally, in Pittman's method, the same series are also calculated for the

geodesic arc from the equator to the turning point. In Karney's method, two dedicated series are calculated and then subtracted to determine the reduced length, which directly increases the number of calculations but also reduces the number of iterations by three times compared to Tseng's method.

The second distinguishing feature of our proposed approach is the use of recursively computed base functions common to both series, which greatly improves the computation of these series. In addition, our method uses strict derivatives that reduces the number of iterations.

Finally, it is worth emphasising that all methods of solving the inverse geodetic problem use Clairaut's formula (1733), most of them in its classical form, i.e.

$$\cos \beta \sin \alpha = \sin \alpha_0 = C$$

Pittman's and Tseng's methods, on the contrary, use the following form

$$\cos \beta \sin \alpha = \cos \beta_{\text{turningPoint}} = C \geq 0$$

which implicitly limits the azimuths to the interval $0 \leq \alpha \leq \pi$ (note that in case this inequality is not satisfied, the right side of the Clairaut formula can be negative, while its left side is always positive). Consequently, such an approach complicates the algorithms as it requires additional operations, such as sign analysis of individual components.

In contrast to our work, the expansion of integrals into series was implemented by computer algebra systems in the work of Karney (2013) and Tseng (2014). Our approach, though laborious and time-consuming, resulted in the possibility of recursion discovery and brought us to an efficient (in a mathematical sense) and effective (in a computational sense) solution.

8 Conclusions

In this paper, we aimed to develop a coherent theory of the ellipsoid's geometry and accomplished our goal elaborating original, rigorous and implement convenient formulas for a geodesic polygon area, the longitude and the arc length of a geodesic based directly on geographical coordinates, e.g. from GNSS. While there were a number of partial approaches focusing on practical solutions for a longitude and distance along a geodesic for which approximate solutions were sufficient, we are unaware of any previous work that simultaneously attains the above-mentioned threefold goal.

In deriving the elliptical integral formulas, a system of cylindrical coordinates was used, whereby integration was

carried out in reference to the normalised radius of parallel ρ , and not in reference to the reduced latitude, as most authors have done since Bessel. When developing the integrals for the geodesic longitude and arc length into a series, their unique similarity was used, i.e. it helped to use the same basic functions. A direct consequence of this method, in addition to being convenient to apply, is the decomposition of a series into two parts: depending on the distance and on the azimuth. This decomposition makes it possible to solve only one spherical triangle instead of two, as used by Danielsen (1989) and Karney (2013). This contributes indirectly to solving the issue of function periodicity, with which some scientists, such as Bessel (1826) and Pittman (1986), struggled, because the distance-dependent part of the formula is interconnected with periodicity.

The other significant objective of this study was to provide the formulas implementation algorithm to effectively obtain results with an accuracy that is limited only by the computer arithmetic applied. This objective was achieved though (a) formulas, defined with the use of constant factors and not with finite series values, as in previous formulas (cf. Danielsen 1989), as well as (b) the use of Napier's rules for the direct determination of the difference between the forward and reverse azimuths, and consequently (c) the elimination of their subtraction that reduces the accuracy of the surface area calculation.

The provided formulas are of recursive form that facilitates and simplifies their coding, which further contributes to a greater efficiency of our algorithm seizing Newton's rapidly converging iterative process and a new simple method to determine the geodesic C parameter's initial value for all cases, including those that previously required solving a 4th degree equation describing Jacobi's (1891) astroid.

Newton's iterative process was constructed for two alternative parameters, C and ω . It is an indirect solution to a difficult theoretical problem of a reduced length of the geodesic investigated by Gauss (1828) and Christoffel (1868), and it replaces Helmert's solution (1880), which requires labour-consuming calculation of a special series. The proposed parameterisation with two alternative working unknowns includes all cases and, in particular, the cases when derivatives do not exist: near the turning point of a geodesic, where Karney's (2013) and Tseng's (2014) algorithm lose their convergence, and in the near-antipodal case, where Vincenty's (1975) algorithm diverges.

The orthodromic characteristics of a geodesic were analysed, which helped to detect the bifurcation lines on which there are two orthodromically equivalent solutions of a geodetic inverse problem, offered by the algorithm presented. New solutions were obtained of a polar geodesic triangle for near-antipodal and bifurcation cases defined by the C parameter value.

The algorithm presented was tested and verified using numerical examples from known publications and our original ones. The numerical tests, performed in extended and double arithmetic, proved the accuracy of 1m² of the computed geodesic area using standard IEEE 754 Double Floating Point arithmetic, for geodesic lengths up to 10,000 km.

The method described converges rapidly, despite the fact that the accuracy of the solution of the inverse geodetic problem $1 \cdot 10^{-15}$, necessary to calculate the area, typically requires additional iteration. The number of iterations, like accuracy, depends on the length of the geodesic (cf. Tab.3): for short arcs, dominant in the applications, only one iteration is sufficient to obtain the final result, with the relative accuracy of the surface area being $5 \cdot 10^{-12}$; for long arcs, on the other hand, up to three iterations are sufficient and the relative accuracy is $5 \cdot 10^{-14}$.

Summing up, we have achieved our goal of developing strict formulas, which are convenient for use in applications and implementations, and an innovative algorithm, which is characterised by its relative simplicity (executable in only a few steps and computationally less complex), its rapid convergence, its high accuracy (depending only on the floating point implementation), and its universal character (applicable for any ellipsoid of revolution and all cases, including near-antipodal cases, cases near the turning point of a geodesic or bifurcation case).

Acknowledgements J. Nowak Da Costa acknowledges the support of this work by the Institute of Geospatial Engineering and Geodesy, Faculty of Civil Engineering and Geodesy of Military University of Technology in Warsaw, Poland. The authors would like to thank the Editor and the anonymous Reviewers whose constructive criticisms and insightful comments, helped to significantly improve this article.

Author contributions E.N. designed research; J.NDC. and E.N. performed research; E.N. and J.NDC analysed data; J.NDC wrote manuscript.

Funding This research was funded by the Military University of Technology in Warsaw, Poland under the statutory grant no. UGB/22-785/2022/WAT.

Data availability All data generated or analysed during this study are included in this published article.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence,

unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Annoni A, Luze C, Guble E, Ihde J, editors. Map Projections for Europe. European Commission, Joint Research Centre; 2001. p. 131. <http://mapref.org/LinkedDocuments/MapProjectionsForEurope-EUR-2010.pdf> (last accessed 17 July 2020)
- Bessel FW (1826) On the computation of geographical longitude and latitude from geodetic measurements. Astronomische Nachrichten (Astronomical Notes), Band 4 (Volume 4), Number 86, Spalten 241–254, Altona 1826
- Biernacki, F., 1949. Teoria odwzorowań powierzchni dla geodetów i kartografów, Główny Urząd Pomiarów Kraju, Prace Geodezyjnego Instytutu Naukowo-Badawczego no.4 [Polish translated into English as Theory of Representation of Surfaces for Surveyors and Cartographers, pp. 53–147, U.S. Department of Commerce, Washington, D.C., 1965].
- Canters F, De Genst W (1997) Development of a low-error equal-area map projection for the European Union, Proceedings of the 18th ICA/ACI International Cartographic Conference, Stockholm, Sweden, 23–27 June, pp. 775–782.
- Chrisman NR (2017) Calculating on a round planet. Int J Geogr Inf Sci 31(4):637–657. <https://doi.org/10.1080/13658816.2016.1215466>
- Christoffel EB (1868) Allgemeine Theorie der geodätischen Dreiecke. Math Abhand König Akad der Wiss zu Berlin 8:119–176. <http://books.google.com/books?id=EEtFAAAcAAJ&pg=PA119>
- Clairaut AC (1733) Détermination géometrique de la perpendiculaire à la méridienne tracée par M Cassini. Mém. de l'Acad Roy des Sciences de Paris 1733–35:406–416.
- Danielsen J (1989) The area under the geodesic. Surv Rev 30(232):61–66. <https://doi.org/10.1179/003962689791474267>
- Gauss CF (1828) Disquisitiones Generales circa Superficies Curvas. Dieterich, Göttingen. <http://books.google.com/books?id=bX0AAAAAMAAJ> [translated into English by Morehead JC and Hiltebeitel AM as General investigations of curved surfaces of 1827 and 1825. Princeton Univ Lib (1902) <http://books.google.com/books?id=a1wTJR3kHwUC>]
- Helmert FR (1880) Die Mathematischen und Physikalischen Theorien der Höheren Geodäsie, volume 1 (Teubner, Leipzig). Translated into English by Aeronautical Chart and Information Center (St. Louis, 1964) as Mathematical and Physical Theories of Higher Geodesy, Part 1 (Chaps. pp. 5–7.)
- Jacobi CGJ (1891) Über die Curve, welche alle von einem Punkte ausgehenden geodätischen Linien eines Rotationsellipsoides berührt. In: Weierstrass KTW (ed) Gesammelte Werke, vol 7. Reimer, Berlin, pp 72–87
- Karney CFF (2013) Algorithms for geodesics. J Geodesy 87(1):43–55. <https://doi.org/10.1007/s00190-012-0578-z>
- Kivioja LA (1971) Computation of geodetic direct and indirect problems by computers accumulating increments from geodetic line elements. Bulletin Geodesique, No 99:55–63
- Kuszta B, Bailey JE (1982) Nonlinear model identification by analysis of feedback-stimulated bifurcation. IEEE Transactions on Automatic Control, 27 (1). pp. 227–228. ISSN 0018–9286, <https://doi.org/10.1109/TAC.1982.1102832>
- Leung Y, Ma JH, Goodchild MF (2004) A general framework for error analysis in measurement-based GIS Part 1: The basic measurement-error model and related concepts. J Geogr Syst 6(4):325–354
- Levallois JJ, Dupuy M (1952) Note sur le calcul des grandes géodésiques. Technical report, Institut Géographique National (IGN), Paris.
- Mollweide KB (1805) Ueber die vom Prof. Schmidt in Giessen in der zweyten Abtheilung seines Handbuchs der Naturlehre S. 595 angegebene Projection der Halbkugelfläche, Zach's Monatliche Correspondenz zur Beforderung der Erd- und Himmels-Kunde, vol. 12, Aug., 152–163.
- Pittman ME (1986) Precision direct and inverse solution of the geodesic. Surveying and Mapping 46(1):47–54
- Pressley A (2010) Elementary Differential Geometry. Second Edition. Springer. ISBN 978–1–84882–891–9
- Rapp RH (1993) Geometric Geodesy – Volume II (revised from past versions). Department of Geodetic Science, The Ohio State University, Columbus, Ohio 43210, March, 1993
- Saito T (1970) The computation of long geodesics on the ellipsoid by non-series expanding procedure. Bulletin Geodesique, No 98:341–374
- Sjöberg LE (2006) Direct solutions to the direct and indirect geodetic problems on the ellipsoid. Zeitschrift f Vermessungswesen 131:35–39
- Sjöberg LE (2012) Solutions to the ellipsoidal Clairaut constant and the inverse geodetic problem by numerical integration. Journal of Geodetic Science 2(3):162–171
- Snyder JP (1987) Map Projections - A Working Manual, U.S. Geological Survey Professional Paper 1395, Washington
- Thomas CM, Featherstone WE (2005) Validation of Vincenty's formulas for the geodesic using a new fourth-order extension of Kivioja's formula. J Surv Eng 131(1):20–26. [https://doi.org/10.1061/\(asce\)0733-9453\(2005\)131:1\(20\)](https://doi.org/10.1061/(asce)0733-9453(2005)131:1(20))
- Tseng WK (2014) An algorithm for the inverse solution of geodesic sailing without auxiliary sphere. J Navig 67:825–844. <https://doi.org/10.1017/S0373463314000228>
- Vincenty T (1975) Direct and inverse solutions on the ellipsoid with application of nested equations. Surv Rev 23(176):88–93. <https://doi.org/10.1179/sre.1975.23.176.88>
- Vincenty T (1976) Correspondence: solutions of geodesics. Surv Rev 23(180):294
- Wrinch DM (1932) Some problems concerned with inverted prolate spheroids. Phil Mag 14(95):1061–1070