J Geod (2005) 79: 413–420
DOI 10.1007/s00190-005-0487-5

ORIGINAL ARTICLE

C-D. Zhang · H.T. Hsu · X.P. Wu · S.S. Li
Q.B. Wang · H.Z. Chai · L. Du

An alternative algebraic algorithm to transform Cartesian to geodetic coordinates

Received: 20 January 2003 / Accepted: 28 July 2005 / Published online: 29 October 2005

Abstract The algorithm to transform from 3D Cartesian to geodetic coordinates is obtained by solving the equation of the Lagrange parameter. Numerical experiments show that geodetic height can be recovered to 0.5 mm precision over the range from -6×10^6 to 10^{10} m.

Keywords Coordinate transformation · Lagrange · Cartesian coordinates · Geodetic coordinates

1 Introduction

The transformation between 3D Cartesian coordinates (x_h, y_h, z_h) and geodetic coordinates (B, L, h) is a basic task that is frequently encountered in geodesy, notably with horizontal datum transformations and GPS positioning. The direct calculation from geodetic to Cartesian coordinates is straightforward. However, the inverse problem, which this paper addresses, is not so.

There are several methods for computing geodetic latitude B and height h from (x_h, y_h, z_h) coordinates. They range from simple iteration of a pair of formulae (Heiskanen and Moritz 1967; Torge 1980) to various closed analytical formulae (Bowring 1976; Vaníček and Krakiwski 1982), and from numerical iterative methods involving trigonometric functions (Borkowski 1989; Lin and Wang 1995; Fukushima 1999) to vector methods (Pollard 2002).

Electronic Supplementary material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00190-005-0487-5

C-D. Zhang (✉) · H.T. Hsu
Key Laboratory of Dynamical Geodesy,
Institute of Geodesy and Geophysics,
Chinese Academy of Science,
Wuhan 430077, PR China
Tel.: +86-371-7975331
E-mail: 13607665382@163.net

C-D. Zhang · X.P. Wu · S.S. Li · Q.B. Wang · H.Z. Chai · L. Du
Department of Geodesy,
Zhengzhou Institute of Surveying and Mapping,
Information Engineering University,
Zhengzhou 450052, PR China

These methods are stable and efficient in different ranges of B and h, but so far only the closed-form algebraic method of Vermeille (2002) is available. Nevertheless, an alternative algebric solution to the inverse problem exists and is proven theoretically in this paper, supported by numerical experiments. A MATLAB computer program is provided as electronic supplementary material (ESM) with this paper.

2 Theory

As is well-known, the geodetic height h at a point $P_h(B, L, h)$ is the distance from the reference ellipsoid to the point in a direction normal to the ellipsoid, in which B is geodetic latitude and L is geodetic longitude. Geometrically, the geodetic height h is the minimum of the distances between the point $P_h(B, L, h)$ and the surface of the reference ellipsoid (Fig. 1).

The 3D Cartesian equation of the reference ellipsoid is

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{b^2} = 1 \tag{1}$$

where a is the semi-major axis and b the semi-minor axis. Therefore, the geodetic height h can be transformed to a minimum problem in the form

$$h^2 = \min[(x_h - x)^2 + (y_h - y)^2 + (z_h - z)^2] \tag{2}$$

where (x_h, y_h, z_h) are the Cartesian coordinates of the point $P_h(B, L, h)$. From Eqs. (1) and (2), we can obtain the corresponding Lagrange equation as

$$h^2 = \min\left\{(x_h - x)^2 + (y_h - y)^2 + (z_h - z)^2 + ab\lambda\left[\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{b^2} - 1\right]\right\} \tag{3}$$

where λ is the Lagrange parameter. Through partial differentiation of Eq. (3) with respect to x, y, z respectively, we obtain the following equations:

$$\begin{aligned} &ax_h - (a + b\lambda)x = 0 \\ &ay_h - (a + b\lambda)y = 0 \\ &bz_h - (b + a\lambda)z = 0 \end{aligned} \tag{4}$$

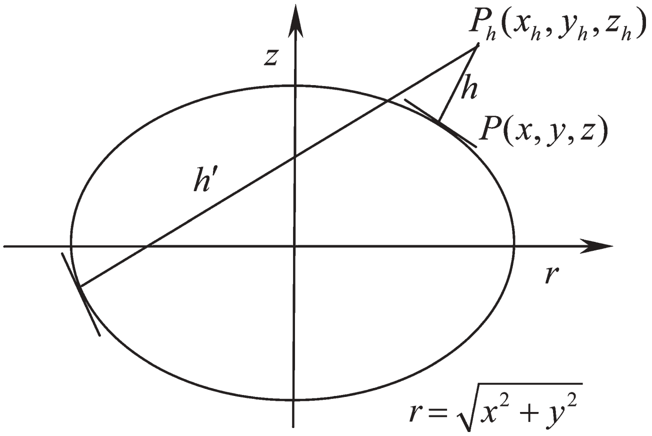


Fig. 1 Cross-section of the ellipsoid of revolution and geometrical definition of the geodetic height

By solving Eqs. (4), we get

$$\begin{aligned} x &= \frac{a}{a + b\lambda} x_h \\ y &= \frac{a}{a + b\lambda} y_h \\ z &= \frac{b}{b + a\lambda} z_h \end{aligned} \tag{5}$$

These are the relations between $P_h(x_h, y_h, z_h)$ and $P(x, y, z)$. Once the Lagrange parameter λ is known, geodetic latitude B and geodetic height h can be calculated directly by

$$\begin{aligned} \tan B &= \frac{1}{(1 - e^2)} \frac{z}{\sqrt{x^2 + y^2}} \\ h &= \mathrm{sign}(\lambda)\sqrt{(x_h - x)^2 + (y_h - y)^2 + (z_h - z)^2} \end{aligned} \tag{6}$$

where $e = \sqrt{a^2 - b^2}/a$ is the first numerical eccentricity of the reference ellipsoid. As can be seen from Fig. 1, if h is negative, the Lagrange parameter λ must be negative, and vice versa.

By substituting Eq. (5) into the equation for the biaxial ellipsoid (Eq. 1)

$$(a + b\lambda)^2 z_h^2 + (b + a\lambda)^2 (x_h^2 + y_h^2) = (a + b\lambda)^2 (b + a\lambda)^2 \tag{7}$$

we obtain a quartic equation in terms of the Lagrange parameter λ.

$$\begin{aligned} &\lambda^4 + 2\gamma\lambda^3 + \left(2 + \gamma^2 - \beta\right)\lambda^2 \\ &+2\left(\gamma - \frac{\alpha + \beta}{\gamma}\right)\lambda + (1 - \alpha) = 0 \end{aligned} \tag{8}$$

where

$$\begin{aligned} \alpha &= \frac{x_h^2 + y_h^2}{a^2} + \frac{z_h^2}{b^2} \\ \beta &= \frac{x_h^2 + y_h^2}{b^2} + \frac{z_h^2}{a^2} \\ \gamma &= \frac{b}{a} + \frac{a}{b} \end{aligned} \tag{9}$$

If $e = 0$, then $a = b$ and Eqs. (9) become

$$\begin{aligned} \alpha &= \frac{x_h^2 + y_h^2 + z_h^2}{a^2} \\ \beta &= \alpha \\ \gamma &= 2 \end{aligned} \tag{10}$$

and Eq. (8) becomes

$$\lambda^4 + 4\lambda^3 + (6 - \alpha)\lambda^2 + 2(2 - \alpha)\lambda + (1 - \alpha) = 0 \tag{11}$$

The solutions of Eq. (11) are $-1,\ -1,\ -1 - \sqrt{\alpha},\ -1 + \sqrt{\alpha}$ and $\sqrt{\alpha} - 1$, which is the one that we want. If $e > 0$, the correct solution of the Lagrange parameter λ must satisfy the condition $\lambda \approx \sqrt{\alpha} - 1$. In the next section, we will obtain the algebraic expression of this solution.

3 The algebraic algorithm

To solve Eq. (8), let us start with the substitution $t = \lambda + \gamma/2$ or $\lambda = t - \gamma/2$, which gives

$$t^4 + pt^2 + qt + r = 0 \tag{12}$$

where

$$\begin{aligned} p &= 2 - \beta - \frac{\gamma^2}{2} \\ q &= \beta\gamma - 2\frac{\alpha + \beta}{\gamma} \\ r &= 1 + \beta - \frac{\gamma^2}{2} - \frac{\beta\gamma^2}{4} + \frac{\gamma^4}{16} \end{aligned} \tag{13}$$

The classical cubic resolvent of a quartic equation (Eq. 12) is defined as

$$s^3 - ps^2 - 4rs + (4pr - q^2) = 0 \tag{14}$$

If s_0 is a positive real root of the cubic resolvent, then the solutions of the original quartic in Eq. (12) can be given by the solutions of (Note: s_0 corresponds to the cubic resolvent $s^3 - ps^2 - 4rs + (4pr - q^2) = 0$ with s being the unknown)

$$t^2 \pm \sqrt{s_0 - p}\left(t - \frac{q}{2(s_0 - p)}\right) + \frac{s_0}{2} = 0 \tag{15}$$

If $q \to 0$, the root s_0 also approaches p, which causes the solutions to be singular. In our application, the p, q and r values are all real numbers, and we only need a non-singular real solution.

To avoid this singularity, we re-derive the solutions from a Fourier expression. Suppose $t_j\,(j = 0, 1, 2, 3)$ are the roots of the quartic Eq. (12), then

$$\begin{aligned} \begin{pmatrix} t_0 \\ t_1 \\ t_2 \\ t_3 \end{pmatrix} &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & \left(e^{2\pi i/4}\right)^1 & \left(e^{2\pi i/4}\right)^2 & \left(e^{2\pi i/4}\right)^3 \\ 1 & \left(e^{2\pi i/4}\right)^2 & \left(e^{2\pi i/4}\right)^4 & \left(e^{2\pi i/4}\right)^6 \\ 1 & \left(e^{2\pi i/4}\right)^3 & \left(e^{2\pi i/4}\right)^6 & \left(e^{2\pi i/4}\right)^9 \end{pmatrix} \begin{pmatrix} 0 \\ u \\ w \\ v \end{pmatrix} \\ &= \begin{pmatrix} +w + (u + v) \\ -w + i(u - v) \\ +w - (u + v) \\ -w - i(u - v) \end{pmatrix} \end{aligned} \tag{16}$$

in which w, u, v are parameter variables and $e^{2\pi i/4} = i = \sqrt{-1}$. Hence, we have

$$\begin{aligned}&(t-t_0)(t-t_2)(t-t_1)(t-t_3)\\&\quad=\left[(t-w)^2-(u+v)^2\right]\left[(t+w)^2+(u-v)^2\right]\\&\quad=t^4-2(2uv+w^2)t^2-4(u^2+v^2)wt\\&\qquad+\left((2uv-w^2)^2-(u^2+v^2)^2\right)\\&\quad=0\end{aligned} \tag{17}$$

Comparing Eq. (17) with Eq. (12), we obtain

$$\begin{aligned}p&=-2(2uv+w^2)\\q&=-4(u^2+v^2)w\\r&=(2uv-w^2)^2-(u^2+v^2)^2\\&=\left(\frac{p}{2}+2w^2\right)^2-(u^2+v^2)^2\end{aligned} \tag{18}$$

From Eq. (18), an alternative cubic resolvent of Eq. (12) is derived as

$$s'^3+\frac{p}{2}s'^2+\left(\frac{p^2}{16}-\frac{r}{4}\right)s'-\frac{q^2}{64}=0 \tag{19}$$

where $s'=w^2$. If s'_0 is a positive real root of the cubic resolvent in Eq. (19), we have

$$w=-\mathrm{sign}(q)\sqrt{s'_0} \tag{20}$$

(Note that here s'_0 corresponds to the resolvent in Eq. (19), with s' being the unknown), and the sign function is defined as

$$y=\mathrm{sign}(x)=\begin{cases}1, & \text{if } x>0\\0, & \text{if } x=0\\-1, & \text{if } x<0\end{cases} \tag{21}$$

then the solutions of the parameter variables u and v are given by the solutions of Eq. (18). They are

$$\begin{aligned}u+v&=\sqrt{\sqrt{\left(\frac{p}{2}+2s'_0\right)^2-r}-\frac{p}{2}-s'_0}\\u-v&=\sqrt{\sqrt{\left(\frac{p}{2}+2s'_0\right)^2-r}+\frac{p}{2}+s'_0}\end{aligned} \tag{22}$$

and they are non-singular. By making the substitution

$$s'=-m-\frac{p}{6} \tag{23}$$

Eq. (19) becomes

$$m^3+3Em-2F=0 \tag{24}$$

where

$$\begin{aligned}E&=-\frac{r}{12}-\frac{p^2}{144}=-G^2\\F&=\frac{pr}{48}-\frac{q^2}{128}-\frac{p^3}{1728}=G^3+\varepsilon\end{aligned} \tag{25}$$

and

$$\begin{aligned}G&=\frac{1}{12}(4+\beta-\gamma^2)\\&=\frac{1}{12}\left(\frac{x_h^2+y_h^2}{b^2}+\frac{z_h^2}{a^2}-\frac{e^4}{1-e^2}\right)\\\varepsilon&=-\frac{1}{32\gamma^2}\left((\alpha+\beta)^2-\alpha\beta\gamma^2\right)\\&=\frac{e^4}{32(1-e^2)}\frac{x_h^2+y_h^2}{b^2}\frac{z_h^2}{a^2}\end{aligned} \tag{26}$$

The discriminant of Eq. (24) is

$$\Delta=E^3+F^2=\varepsilon(\varepsilon+2G^3) \tag{27}$$

If $\Delta\geq 0$, the real-valued root of Eq. (24) is

$$\begin{aligned}m&=\sqrt[3]{F+\sqrt{E^3+F^2}}+\sqrt[3]{F-\sqrt{E^3+F^2}}\\&=\sqrt[3]{F+\sqrt{E^3+F^2}}-\frac{E}{\sqrt[3]{F+\sqrt{E^3+F^2}}}\end{aligned} \tag{28}$$

Finally, we find that $s'_0=-m-\frac{p}{6}$ is a positive real-valued root and $t_0=w+(u+v)\to\sqrt{-p}\to\sqrt{\alpha}$ or $\lambda=t_0-\frac{\gamma}{2}\to\sqrt{\alpha}-1$ is the root that we need.

4 Convergence region

The relation between a Cartesian, cylindrical or spherical coordinate system is one-to-one. However, the relation between a Cartesian and a geodetic coordinate system is not one-to-one. A point in some region can have many geodetic coordinates. For example, if a point is located at the origin, its geodetic coordinates may be $(0, L, -a)$ or $(\pm\pi/2, L, -b)$, where $L\in[0,2\pi)$. This means that the algorithm that we will propose is effective, if and only if a point has its unique geodetic coordinates.

The above can be explained geometrically as follows. From the general knowledge of ellipse in analytical geometry, if two normals within an ellipse, say P_1K_1 and P_2K_2, intersect at I (Fig. 2), then the coordinates of the point of intersection can be determined using their equations. In Fig. 2, B_1 and B_2 are inclinations of the two normals, which are also the latitudes.

From our knowledge of ellipsoidal geometry, the coordinates of the point of intersection I are

$$\left(\frac{N_1\sin B_1-N_2\sin B_2}{\tan B_1-\tan B_2}e^2,\ \frac{N_1\sin B_1\tan B_2-N_2\sin B_2\tan B_1}{\tan B_1-\tan B_2}e^2\right) \tag{29}$$

in which N_1 and N_2 stand for the length of P_1K_1 and P_2K_2 respectively, and can be computed using:

$$N=\frac{a}{\sqrt{1-e^2\sin^2 B}} \tag{30}$$

Within the ellipsoid, the region that contains the points of intersections can be further determined numerically, just as Fig. 3 shows.

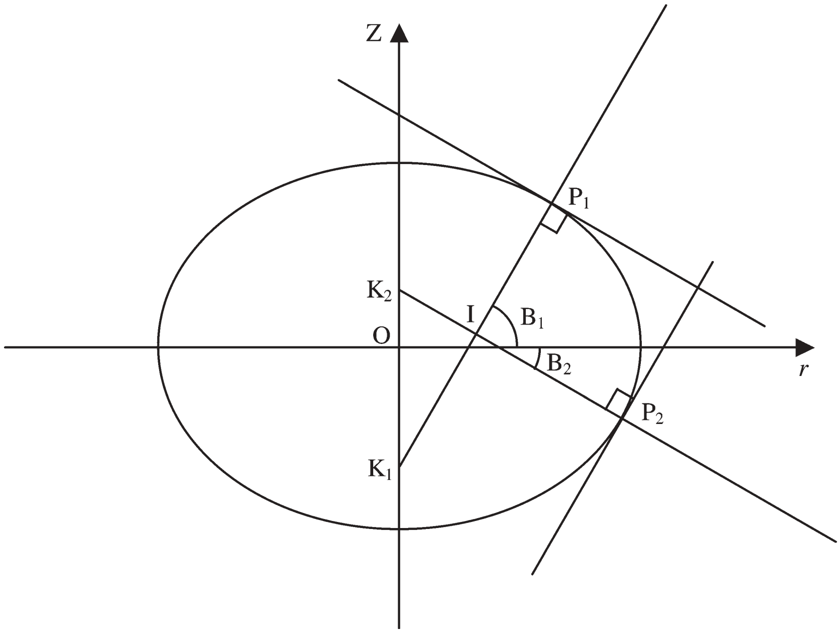


Fig. 2 Point of intersection of two normals within an ellipse

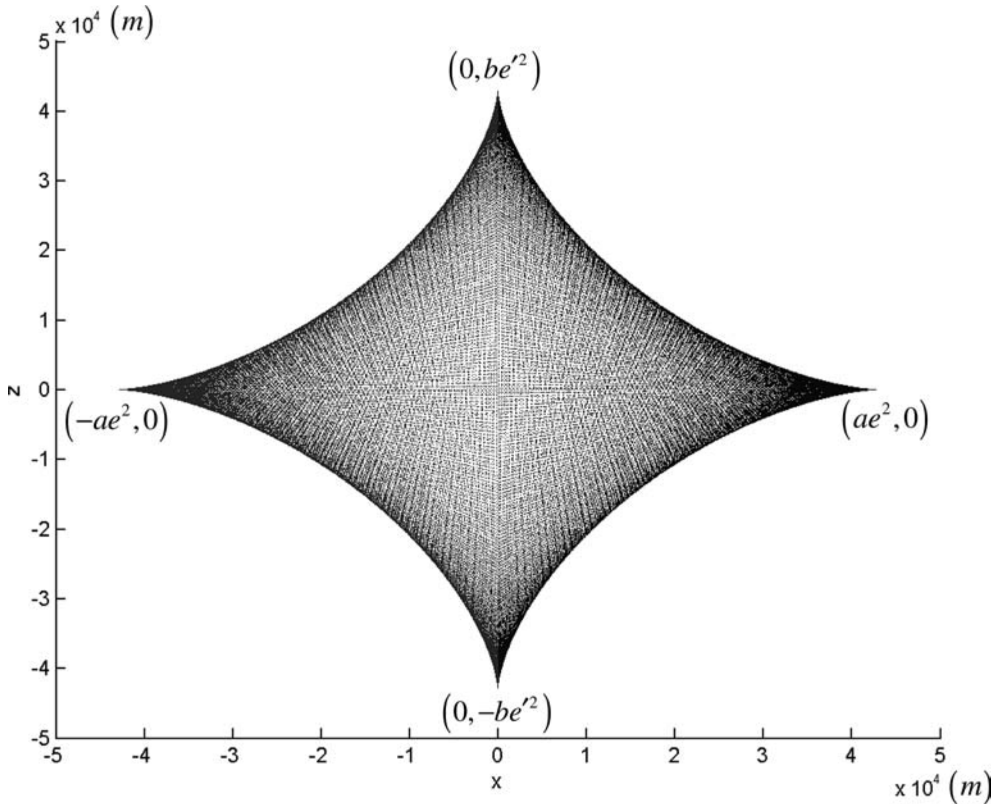


Fig. 3 Unique and multivalue regions of geodetic coordinates. The parameters of reference ellipsoid are set as those of GRS80, namely $a = 6378137$ m and $f = 1/298.257222101$, where the flattening f is defined as $f = 1 - b/a$, e is the first eccentricity, e' is the second eccentricity

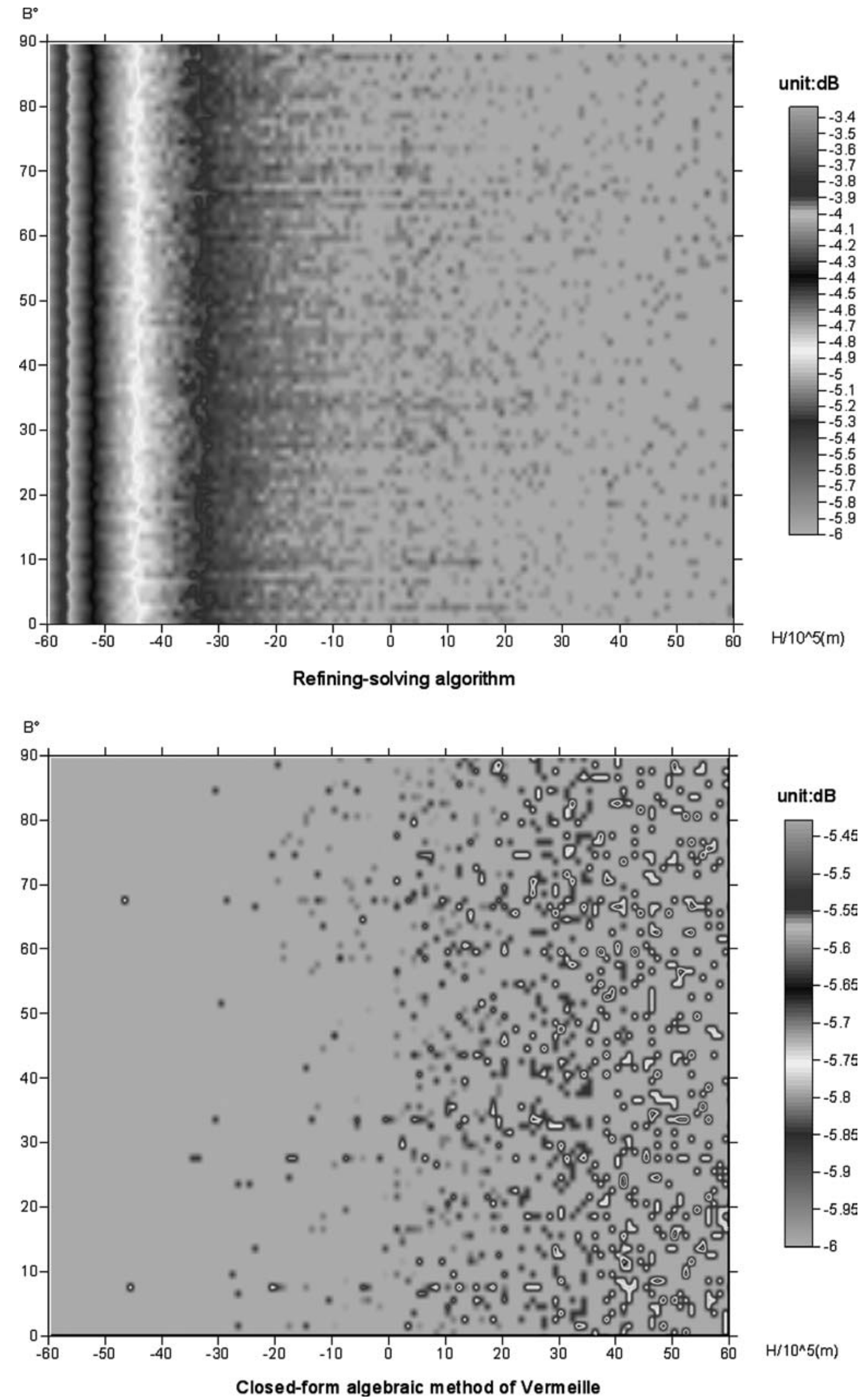
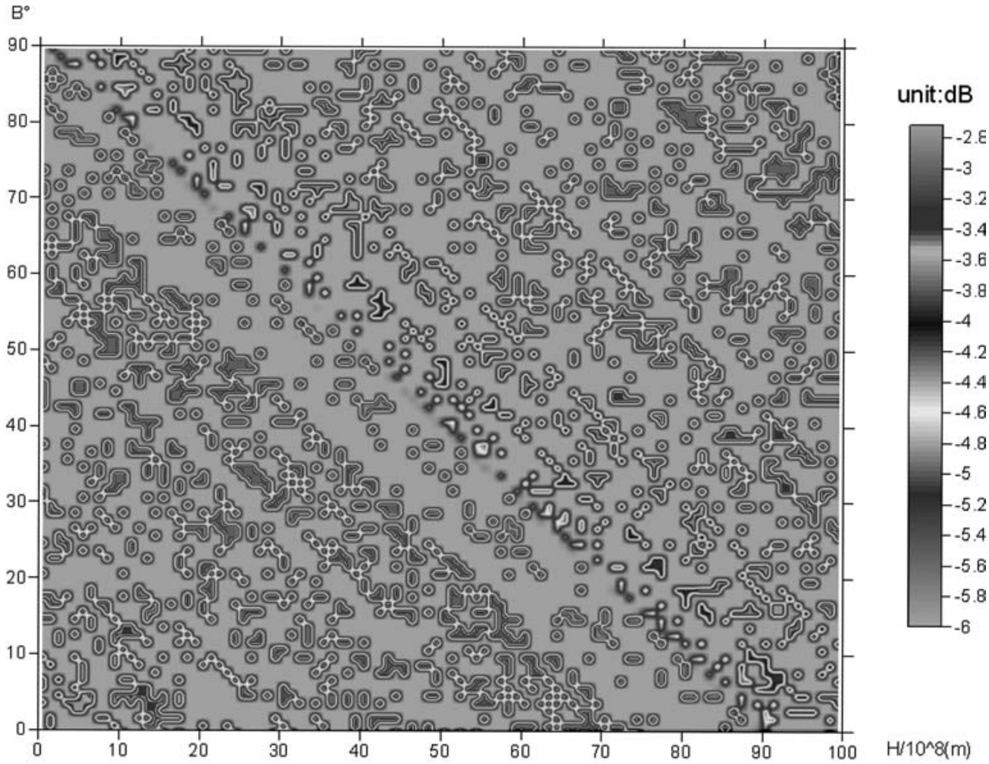
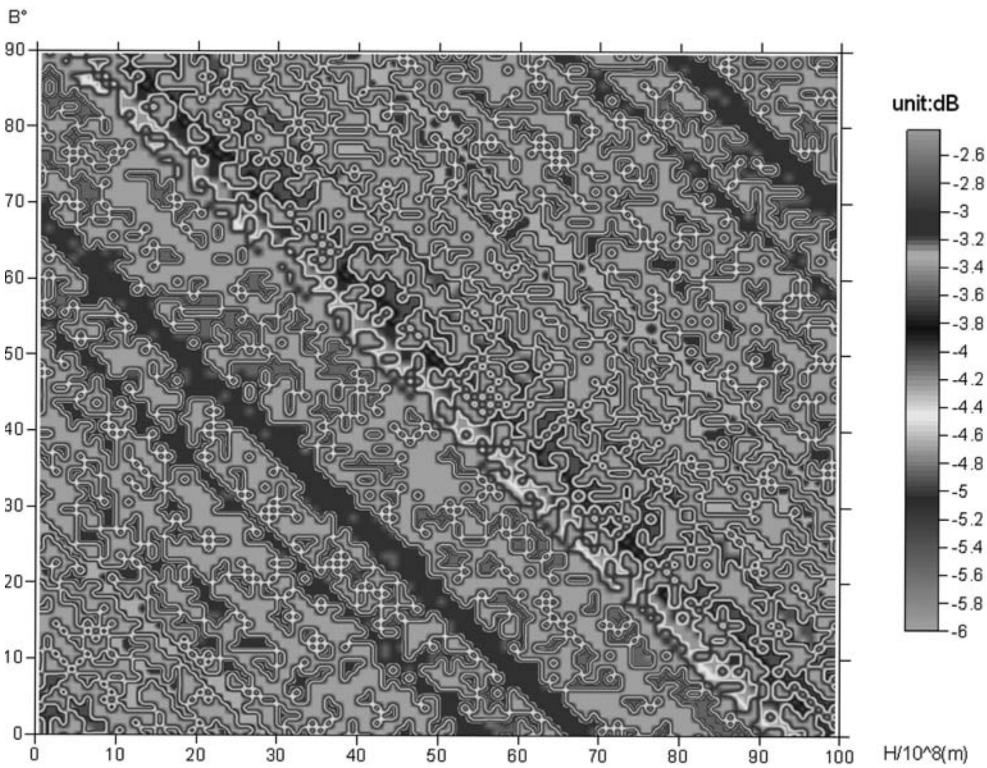


Fig. 4 Height error distribution $\left(\text{units in dB, i.e., } \log_{10}\left(\frac{|\Delta H|(\mathrm{m})}{0.001(\mathrm{m})}\right)\right)$ in area with latitude ranging from 0 to 90° and height from -6×10^6 to 6×10^6 m. The parameters are same as Fig. 2. If height errors are less than -6 dB, they are set to -6 dB in the plots



Refining-solving algorithm



Closed-form algebraic method of Vermeille

Fig. 5 Height error distribution $\left(\text{units in dB, i.e., } \log_{10}\left(\frac{|\Delta H|(\mathrm{m})}{0.001(\mathrm{m})}\right)\right)$ in area with latitude ranging from 0 to 90° and height from 10^0 to 10^{10} m. The parameters are same as Fig. 2. If height errors are less than -6 dB, they are set to -6 dB in the plots

According to Zwillinger (1996), we find that if and only if the discriminant $\Delta < 0$ (Eq. 21), are there three unequal roots. This means that the Lagrange parameter λ (in Eqs. 5 and 6) has many real solutions. Therefore, any point located in the region $\Delta < 0$ has many geodetic coordinates to which it is related. That is to say, the point in $\Delta > 0$ has its unique geodetic coordinates. The convergence region and multivalue region are shown in Fig. 3.

From the closed-form algebraic method of Vermeille (2002), $t = \sqrt[3]{1 + s + \sqrt{s(2 + s)}}$, and letting $s(2 + s) > 0$, we can also arrive at Eq. (27). This means the above view of the convergence region is correct.

5 Algorithmic flow

Theoretically, the expressions in radical sign are all real non-negative numbers. However, their values via computer are sometimes negative, because of computer rounding errors. Therefore, the applicable algebraic algorithm becomes the following refining-solving (RS) algorithm:

Refining part:

1. Calculate (α, β, γ) from Eq. (9).
2. Calculate (p, q, r) from Eq. (13).
3. Calculate (E, F) from Eq. (25) or/and Eq. (26).
4. Calculate $m = \sqrt[3]{\left|F + \sqrt{\left|E^3 + F^2\right|}\right|} - \dfrac{E}{\sqrt[3]{\left|F + \sqrt{\left|E^3 + F^2\right|}\right|}}$
5. Calculate $s_0' = |-m - p/6|$. (Note: here we take the absolute value to avoid negative numbers caused by computer roundoff because we have stated that it is a positive real root at Eq. (28))
6. Calculate $u + v = \sqrt{\left|\sqrt{\left|\left(\frac{p}{2} + 2s_0'\right)^2 - r\right|} - \frac{p}{2} - s_0'\right|}$.
7. Calculate $t_0 = -\mathrm{sign}(q)\sqrt{s_0'} + (u + v)$.
8. Update $t_0 = \sqrt{-\dfrac{p}{2} + \dfrac{\sqrt{p^2 - 4(qt_0 + r)}}{2}}$ from Eq. (12) (only one pass).

Solving part:

9. Calculate $\lambda = t_0 - \frac{\gamma}{2}$ and (x, y, z) from Eq. (5).
10. Calculate (B, h) from Eq. (6).

This RS algorithm is an exact algebraic one.

6 Numerical experiments

The main objectives of the numerical experiments are (1) to validate the RS algorithm, (2) to estimate the error of the algorithm, (3) to demonstrate that our RS algorithm is suitable for solving the transformation problem of Cartesian to geodetic coordinates.

In the first test, we ran a series of conversion tests from Cartesian to geodetic coordinates for millions of grid points covering the area with latitude φ ranging from 0° to 90° and height h from -6×10^6 to 6×10^6 m with both the above RS algorithm and the closed-form algebraic method of Vermeille (2002). In the second test, we used grid points covering the area with latitude φ ranging from 0° to 90° and height h, from 10^0 to 10^{10} m.

The test results were obtained on a Sony PCG-GRV7CP computer. They show that our RS algorithm has good numerical stability and ultra-high precision. In the two test cases, the errors are $< 10^{-5}$ arc-second in latitude and < 0.5 mm error in geodetic height. The geodetic height errors in the two different cases are shown in Figs. 4 and 5, respectively. The reference values (0 dB) in Figs. 4 and 5 are 1 mm for height. Satisfactory performance is indicated by the plane at -3 dB. Errors above this plane are not applicable to geodetic applications.

Further tests show that the errors are mainly due to numerical effects. If we use more that 16-digit numerical computation with MATLAB, the error decreases and the numeric convergence region increases.

7 Conclusion

An alternative algebra solution for transformation of geocentric rectangular coordinates to geodetic coordinates has been proposed. It is sufficiently precise because its error is less than 0.5 mm in height and 10^{-5} arc-second in latitude over the height ranging from -6×10^6 to 10^{10} m, and it is stable since it applies to all given coordinates, including those near to the geocenter, the polar axis and the equator. A MATLAB routine of our RS algorithm is available as ESM with this article. The RS algorithm agrees well with the closed-form algebraic method of Vermeille (2002).

Acknowledgements We thank B.-Y. Ma, G. Strang van Hees, S.A. Holmes and an anonymous person for their thoughtful reviews of the preprint of this paper. This work has been supported by National outstanding youth foundation of China (40125013) and A Foundation for the Author of National Excellent Doctoral Dissertation of PR China (200344).

References

Borkowski KM (1989) Accurate algorithms to transform geocentric to geodetic coordinates. Bull Geod 63:50–56
Bowring BR (1976) Transformation from spatial to geographical coordinates. Surv Rev 181:323–327
Zwillinger D (ed) (1996) CRC standard mathematical tables and formulae, 30th edition, CRC Press, Boca Raton
Fukushima T (1999) Fast transform from geocentric to geodetic coordinates. J Geod 73:603–610
Lin KC, Wang J (1995) Transformation from geocentric coordinates using Newton's iteration. Bull Geod 69:14–17
Pollard J (2002) Iterative vector methods for computing geodetic latitude and height from rectangular coordinates. J Geod 76:36–40
Torge W (1980) Geodesy, Walter de Gruyter, Berlin

Heiskanen WA, Moritz H (1967) Physical Geodesy, W H Freeman, San Francisco

Laskowski P (1991) Is Newton's iteration faster than simple iteration for transformation between geocentric and geodetic coordinates? Bull Geod 65:14–17

Vaníček P, Krakiwski EJ (1982) Geodesy: the concepts. North Holland, Amsterdam

Vermeille H (2002) Direct transformation from geocentric coordinates to geodetic coordinates. J Geod 76:451–454