Universität Stuttgart

Schriftenreihe der Institute des Studiengangs Geodäsie und Geoinformatik

Technical Reports
Department of Geodesy and Geoinformatics

Quo vadis geodesia ... ?

Festschrift for Erik W. Grafarend

on the occasion of his 60th birthday

Part 2

Friedhelm Krumm,
Volker S. Schwarze
(Eds.)

ISSN 0933-2839

Report Nr. 1999.6-2

Friedhelm Krumm / Volker Siegfried Schwarze (Eds.)

Quo vadis geodesia...?

Festschrift for Erik W. Grafarend

on the occasion of his 60th birthday

Part 2



Geodätisches Institut der Universität Stuttgart 1999



Foreword

This „Festschrift" is dedicated to

Prof. Dr.-Ing. habil. Tekn. Dr. h.c. Dr.-Ing. E.h. Dr. h.c. Erik W. Grafarend

in honour of his 60th birthday. The two volumes of this „Festschrift" cover a broad spectrum of geodetic research and mirror his scientific contributions and widespread scientific interests. Prof. Grafarend's scientific work spans from the rotation of a gyroscope to the rotation of the Earth as described by a deformable, viscoelastic body. It also covers geodetic sensors, from theodolites to inertial systems, from photogrammetric cameras to artificial Earth satellites.

Prof. Grafarend's major scientific contribution is towards a unified theory of Geodesy. In working with the gravitational field of the Earth he used the disciplines of Geodesy and Physics, of geometry space and gravity space, and beautifully showed the interrelationship between these models. This enjoyable interplay between these disciplines lead him to a unified approach towards geodesy: For example, the motion of a satellite, the trajectory of a light ray and the course of a plumbline can all be interpreted as geodesics - curves of minimal distance on some properly chosen space-time manifold. Prof. Grafarend also looked at the statistical nature of the geodetic measurement processes. It is therefore not surprising that he has written many important papers on adjustment theory and statistical inference.

No area of geodetic interest has been neglected by Prof. Grafarend to achieve a unified theory of geodesy. His scientific work in totality looks like the geodetic version of the „Glasperlenspiel" by Hermann Hesse with its never resting „magister geodesiae" Prof. Grafarend in its centre. In that sense these two volumes of papers written by internationally acknowledged scientists give a timely and representative overview over geodesy, as Prof. Grafarend sees it. It also becomes obvious how many people have been stimulated by his ideas and visions, and wish to express their respect and gratitude on the occasion of his 60th birthday.

The editors appreciate all the work by the authors of „Festschrift".

Friedhelm Krumm

Volker Schwarze

Stuttgart, October 1999

Welcome Address

On behalf of the directorate of the Universität Stuttgart it is my great pleasure to welcome all of you to the two-day symposium *Quo vadis geodesia...?* being held on occasion of the 60th Birthday of Prof. Dr.-Ing. habil. Tekn. Dr. h.c. Dr.-Ing. E.h. Dr. h.c. Erik W. Grafarend. Nobody will believe that Erik Grafarend has reached the age of *wise men*, who normally show silver hair and behave according to their age. He seems *ageless*, he stimulates, researches and lectures the basics of geodetic science at our university since 20 years without showing any loss of his young spirit. His always straightforward interests to introduce new theoretical concepts into the disciplines of statistical inference, mathematical and astronomical geodesy, cartographic map projections, and basic surveying computations have made him an national and international well-acknowledged and awarded scientist. We feel honored to have him as member in the professors' board, we feel also honored today by your participation in this symposium.

Erik Grafarend was born on Oct. 30, 1939 in Essen, Germany. After primary and secondary education he started his academic career studying Mining and Surveying at the Technical University Clausthal-Zellerfeld, Germany (1959-1964). He graduated 1964 as Diplom-Ingenieur (MSc) in Mining and Surveying. Spending three years as research associate in the Institute of Mining and Surveying of this university he obtained his doctoral degree (PhD) by submitting a thesis on *Azimuth Gyroscopes*. He realized very soon that natural sciences might be helpful to provide him deeper knowlegde and roots for his theory-oriented engineering concepts. Thus, he became again student studying Physics and graduated 1968 with the Diplom-Physiker (MSc in Physics). Prof. Helmut Wolf at Bonn university became his mentor in 1969, when Erik Grafarend was promoted being lecturer for geodesy and geophysics at the Institute of Theoretical Geodesy. After submitting a second thesis for qualifying himself to become professor (Habilitation) he was appointed 1970 Associate Professor at Bonn University. This thesis is well-known, here he wrote down the *Accuracy measures of a point manifold in multidimensional Euclidean spaces*. Due to his splendid publications he qualified himself very soon for a position as Full Professor - he accepted 1975 the offer of the University of the Federal Armed Forces Munich to take over the Chair of Astronomical and Physical Geodesy. He spent five very lifely and stimulating years in Munich, organised workshops and seminars to discuss and to deep statistical inference concepts, for example to prove robust statistics and to fully explore the potential of least squares prediction and collocation. Parallel to his Munich activities he became Adjunct Professor of the Technical University of Darmstadt lecturing the basics of satellite orbit computations.

Since 1980 he is with the Universität Stuttgart, he accepted the position being Full Professor and Head of the Geodetic Institute embedded in the Curriculum of Geodesy and Surveying. Besides his strong scientific interests he took over also the responsibility being Dean of the Faculty of Civil Engineering and Surveying (1985-1986) and served from 1995-1999 as Dean for Educational Affairs within the Curriculum of Geodesy and GeoInformatics.

Since the early seventies Erik Grafarend is heavily internationally engaged, he is one of the *pillars* of the International Association of Geodesy (IAG). Here he held several positions: Special Study Group member of many SSGs, SSG President, Commission President, etc. He realised 1980 that there is a need to launch a high level scientific journal, as result the *manuscripta geodaetica* was published. This journal is today one of the internationally recognized journals of our profession, its weight in citation index we gratefully acknowledge. But besides all these activities a real measure of Erik Grafarends output are the more than 230 publications, all of high level. Very often, we the readers of his publications, have difficulties to fully understand his definitions, sets and proofs. But as many of us know him, we give him the confidence that the final results of his complex mathematical boxes are okay. The validation by examples is often done by his staff members.

It seems to us obvious that he was awarded several times. This awarding series started 1971 when he got the Award of the National Science Foundation of Germany (DFG) for research on *The Geodetic*

Reference Surface, the IAG Bomford prize followed 1975 (during the General Assembly in Grenoble/France), also to mention the Senior K. and W.A. Heiskanen Award of The Ohio State University, Department of Geodetic Science, Columbus/Ohio. The National Academy of Sciences awarded him 1978 the Senior Scientist Award. As we realized in the first sentence of this welcome address he obtained three honorary doctoral degrees: the Tekn. Dr. h.c. by the Kungl. Tekniska Högskolan, Stockholm/Sweden (1989), the Dr.-Ing. E.h. by the Technical University of Darmstadt/Germany (1996), and the Dr. h.c. by the Technical University of Budapest/Hungary (1998). Most probably, this series of awards is still open.

What have his colleagues at the Universität Stuttgart done to keep such an outstanding scientist satisfied? Well, we provided him 1997 adequate room facilities, a dream he had from the mid eighties. Many of you may have wondered whether the furniture and the room furnishings are standard of our university. They are not. It is his personal note of room design and interior architectural taste. We further offered him the opportunity to share lectures and exercises in adjustment and statistical inference, an interest he was also looking for from his beginnings at this university. Thus, we feel having done quite a lot to offer him a good atmosphere for science and education, at least during the last few years. We hope, that the next five years Erik Grafarend will further contribute high level papers to geodesy and related disciplines. Let me close with the statement: " Erik, thank you very much for your contributions. Congratulations to your 60th Birthday. We wish you and Ulrike a further interesting life. Please keep your shape and spirit young. "

Dieter Fritsch, Vice President Education
Universität Stuttgart

Contents Part 1

Ádám József: **Spaceborne VLBI beyond 2000** 1

Aduol Francis W O: **Robust geodetic parameter estimation under least squares through weighting on the basis of the mean square error** 5

Ardalan Alireza A and Awange Joseph L: **Care while using the NMEA 0183!** 17

Ardalan Alireza A: **Somigliana-Pizetti minimum distance telluroid mapping** 27

Awange Joseph L: **Partial Procrustes solution of the threedimensional orientation problem from GPS/LPS observations** 41

Bähr Hans-Peter: **Geodesy and Semantics – Progress by Graphs** 53

Caputo Michele and Plastino Wolfango: **Diffusion with space memory** 59

Crosilla Fabio: **Procrustes analysis and geodetic sciences** 69

Dermanis Athanasios: **On the maintenance of a proper reference frame for VLBI and GPS global networks** 79

Dorrer Egon: **From Elliptic Arc Length to Gauss-Krüger Coordinates by Analytical Continuation** 91

Featherstone Will: **Tests of two forms of Stokes's integral using a synthetic gravity field based on spherical harmonics** 101

Förstner Wolfgang and Moonen Boudewijn: **A metric for covariance matrices** 113

Groten Erwin: **Earth rotation as a geodesic flow, a challenge beyond 2000 ?** 129

Hannah Bruce M, Kubik Kurt and Walker Rodney A: **Propagation modelling of GPS signals** 137

Hartung Joachim: **A Short-Cut Method for Computing Positive Variance Component Estimates** 151

Heck Bernhard: **Integral Equation Methods in Physical Geodesy** 155

Hehl Friedrich W, Obukhov Yuri N and Rubilar Guillermo F: **Classical Electrodynamics: A Tutorial on its Foundations** 171

Hils Alfred: **Grußwort** 185

Hofmann-Wellenhof Bernhard: **Erik W. Grafarend – Ist Größe messbar ?** 187

Ilk Karl Heinz: **Energiebetrachtungen für die Bewegung zweier Satelliten im Gravitationsfeld der Erde** 191

Jurisch Ronald, Kampmann Georg and Linke Janette: **Über die Analyse von Beobachtungen in der Ausgleichungsrechnung – Äußere und innere Restriktionen** 207

Kakkuri Juhani: **The challenge of the crustal gravity field** 231

Keller Wolfgang: **Geodetic pseudodifferential operators and the Meissl scheme** 237

Kleusberg Alfred: **Analytical GPS navigation solution** 247

Koch Karl-Rudolf: **Grundprinzipien der Bayes-Statistik** 253

Leick Alfred: **GLONASS Carrier Phases** 261

Lelgemann Dieter and Cui Chunfang: **Analytical versus numerical integration in satellite geodesy** 269

Linkwitz Klaus: **About the generalised analysis of network-type entities** 279

Livieratos Evangelos: **Intrinsic Parameters and Satellite orbital elements** 295

Contents Part 2

Marchenko Alexander N: **Simplest solutions of Clairaut's equation and the Earth's density model** 303

Martinec Zdeněk: **Stokes's two-boundary-value-problem** 313

Meier Siegfried: **Vom Punktlagefehler zum Qualitätsmodell** 323

Mira Sjamsir: **Hydrographic surveying and its education** 331

Moritz Helmut: **The strange behavior of asymptotic series in Mathematics, Celestial Mechanics and Physical Geodesy** 335

Papo Haim: **Datum accuracy and its dependence on network geometry** 343

Reilly Ian W: **Strain in the Earth – a geodetic perspective** 355

Rizos Chris, Han Shaowei, Chen Horng-Yue and Chai Goh Pong: **Continuously operating GPS reference station networks: New algorithms and applications of carrier-phase-based, medium range, static and kinematic positioning** 367

Rozanov Youri and Sansò Fernando: **The analysis of the Neumann and oblique derivative problem. Weak theory.** 379

Rummel Reiner and van Gelderen Martin: **From the Generalized Bruns Transformation to Variations of the Solution of the Geodetic Boundary Value Problem** 397

Schäfer Volker: **Quo vadis geodesia? ... Sic erit pars publica** 413

Schaffrin Burkhard: **Reproducing estimators via least squares: An optimal alternative to the Helmert transformation** 419

Schmitz-Hübsch Harald and Schuh Harald: **Seasonal and short-period fluctuations of Earth rotation investigated by wavelet analysis** 421

Schreiber Ulrich, Schneider Manfred, Stedman Geoffrey E, Rowe Clive H and Schlüter Wolfgang: **Charakterisierung des C-II Ringlasers** 433

Seitz Kurt: **Ellipsoidal and topographical effects in the scalar free geodetic boundary value problem** 439

Sideris Michael G, Fei Zhiling L and Blais J A R.: **Ellipsoidal corrections for the inverse Hotine/Stokes formulas** 453

Sjöberg Lars E: **Triple frequency GPS for precise positioning** 467

Svensson S Leif: **Map projections and the boundary problems of Physical Geodesy** 473

Teunissen Peter J G: **GPS, Integers, Adjustment and Probability** 479

Varga Péter: **Geophysical Geodesy beyond 2000** 487

Vermeer Martin: **Geodetic science and the tools of the trade** 497

Welsch Walter: **Fortgeschrittene geodätische Deformationsanalyse** 505

Wittenburg Rüdiger: **Zur geodätischen Beschreibung von Deformationsprozessen im 3D-Bereich** 515

Xu Peiliang: **Random stress/strain tensors and beyond** 525

You Rey-Jer: **Geodesy beyond 2000: An attempt to unify Geodesy by the geodesic flow in all branches** 549

Yurkina M I: **A solution of Stokes' problem for the ellipsoidal Earth by means of Green's function** 557

List of Authors 565

Simplest solutions of Clairaut's equation and the Earth's density model

Alexander N. Marchenko

Introduction

Starting from the first investigations of the Earth's density distribution some remarkable and simple density laws were constructed by Legendre, Laplace, G.Darvin, Roche, etc. These laws have a spherically symmetric density distribution with the volume density $\rho(\ell)$ that depends on the radial distance ℓ. At the Geodetic Week97 (Berlin,1977) the author had several interesting discussions with Prof. E. Grafarend about the exponential nature of the flattening distribution according to the Clairaut's equation. Maybe these meetings and discussions yielded now the presented consideration of the famous classic law of density in view of their mathematical descriptions. As a result, the latter is the main goal of this paper. On the other hand, we shall try to illustrate some our results by the numerical investigations on the ground of the fundamental constants of geodesy and astronomy together with global data of the seismic tomography of the Earth's interior.

1. Some basic relationships

In view of a mathematical formulation the traditional representation of the Earth's radial density can be treated as a function $\rho(\ell)$ (continuos or piecewise in form of shells) of one variable ℓ only, which is defined on the finite segment $(0 \le \ell \le R)$ only if we assume that the figure of the planet is spherical, (R is the mean Earth's radius (R = 6371 km)). It is well-known also (Moritz, 1990) that in this case the gravitational potential V is equal to the gravity potential W, since we use such simplest approximation of the ellipsoid by the sphere when the flattening $f = 0$.
First of all our initial (observed) information will be the Earth's mass M and the mean moment of inertia I. For latter use we shall write some well-known formulae within the sphere of the radius ℓ (the part of the Earth's mass which is restricted by this radius) for the *mass*

$$M(\ell) = 4\pi \int_0^{\ell} \rho(x) x^2 dx \quad , \tag{1}$$

where dx is the element of a line and the *mean density* $D(\ell)$:

$$D(\ell) = \frac{3}{4 \cdot \pi \cdot \ell^3} M(\ell) \quad . \tag{2}$$

The value $D(\ell)$ in the form of (2) leads to the following representations

$$g(\ell) = \frac{4 \cdot \pi \cdot G}{3} \ell \cdot D(\ell) \quad \Leftrightarrow \quad g(\ell) = \frac{GM}{\ell^2} \quad , \tag{3}$$

of the *gravity* $g(\ell)$ inside the Earth, where G=6.673·10^{-8} [cm^{3}s^{-2}g^{-1}] is the gravitational constant. The *mean moment of inertia* is

$$I(\ell)=\frac{8\pi}{3}\int_0^{\ell}\rho(x)x^4dx \quad . \tag{4}$$

We shall use also the seismic velocities V_p and V_s in the form of the function

$$\Phi=\Phi(\ell)=V(\ell)_P^2-\frac{4}{3}V(\ell)_S^2 \quad , \tag{5}$$

by applying their grid values in accordance with (Dziewonski and Anderson, 1981), which practically represents the results of seismic tomography of the Earth interior.

2. The simplest solutions of Clairaut's equation

Now we recollect that the famous oldest hypothesis for the Earth's density distribution were proposed after solutions of Clairaut's equation for the flattening inside the Earth (see, for instance, Bullen, 1975; Moritz, 1990).
There exist (Bullen, 1975) three famous solutions of this equation for the density ρ. First one is *Legandre - Laplace law*

$$\rho(x)=\rho_0\frac{\sin(\beta x)}{\beta x}=\rho_0\frac{\exp(\sqrt{-1}\beta x)-\exp(-\sqrt{-1}\beta x)}{2\sqrt{-1}\beta x}, \qquad \beta=\text{const}, \tag{6}$$

where we apply the dimensionless „radius-vector"

$$x=\frac{\ell}{R} \quad , \tag{7}$$

regarding to R; ρ_0 = const and may be considered here as the density at the origin.
The second one is *Roche's law*

$$\rho(x)=\rho_0\left(1-Kx^2\right)=a+bx^2 \quad , \tag{8}$$

where

$$a=\rho_0>0 \quad \text{and} \quad b=\rho_0K<0 \quad . \tag{9}$$

Note now that Taylor series expansion of (6) (disregarding other higher powers of x) in view of mathematics leads to the similar expression:

$$\rho(x)=\rho_0\left(1-\frac{\beta^2}{6}x^2\right) \quad . \tag{10}$$

The third one is *G. Darwin law*

$$\rho(x)=C\cdot x^{-n} \quad , \tag{11}$$

where C is a constant. His solution involves an „assumption of the form for the law of the internal density of the planet and subsequent determination of the law of compressibility„ (Darwin, 1884). Clearly, the expression (11) represents the density with a singularity at the origin. G. Darwin noted

already that case $n = 0$ for the model (11) corresponds to the case of homogeneous density; for $n = 3$ the Earth's mass M will become infinite; for $n > 3$ the mass M must be assumed to be negative. As a result, we get the inequality $0 < n < 3$ which agrees with the determination $n = 1.011$ (Bullen, 1975). Thus the expression (11) represents a power function.

3. Williamson-Adams equation

The density ρ may fulfil the so-called Williamson-Adams equation for each shell of the stratified Earth under the following assumptions: the Earth is globally in hydrostatic equilibrium; chemical composition and phase transformation are homogeneous in every shell; the temperature is adiabatic in each shell. Thus, if we have the observable seismic velocity (5), in view of the gravitational (3) and hydrostatic relationships

$$gradp(\ell) = \rho(\ell) \cdot gradV(\ell) \qquad \Rightarrow \qquad \frac{dp(\ell)}{d\ell} = -\rho(\ell) \cdot g(\ell) \quad , \tag{12}$$

finally the Williamson-Adams equation can be written as

$$\frac{d \ln \rho(\ell)}{d\ell} = -\frac{g(\ell)}{\Phi(\ell)} \quad , \tag{13}$$

where p is the pressure inside the Earth. Thus (13) is a formula to derive the radial density distribution from the seismic velocity data, fulfilled under the assumptions listed above.
In order to use (13) we must first try to solve this equation and to express the observed seismic data by a suitable function of depth, separating the Earth into convenient shells. Traditionally we shall assume that the separation into shells has to be choice at those spheres, where discontinuities in the parameter Φ or in its derivative can be observed.
It is evident that the formal solution of (13) may be obtained after the integration of Williamson-Adams equation. The result is

$$\rho(\ell) = \rho_0 \exp\left(-\int_0^{\ell} \frac{g(x)}{\Phi(x)} dx \right) \quad , \tag{14}$$

and we get the functional dependence for radial density as an exponential function. The right hand side of the expression (14) is unknown. For this reason, we shall apply instead of (14) the simplest approximating function

$$\rho(\ell) = \rho_0 \exp\left(-\gamma^2 x^2\right) \quad , \qquad \gamma = \text{const} \quad , \tag{15}$$

where the power 2 is the lowest power for which we may get a non-zero value Φ at the origin. Taylor expansion of (15) leads again to the Roche's model

$$\rho(x) = \rho_0 \left(1 - \gamma^2 x^2\right) = a + bx^2 \quad , \tag{16}$$

if we disregard other higher powers of x.

4. Poisson's equation

The density ρ must fulfil the Poisson's equation for the gravity potential $W = V$ of the Earth. Using the spherical coordinates after simple manipulations we get for a radial layered Earth, that is for $\rho = \rho(\ell)$, in spherical approximation

$$-\Delta V = 4\pi G\rho = \frac{dg}{d\ell} + \frac{2g}{\ell} = M[g] \quad . \tag{17}$$

The operator

$$M[\ ] = \frac{d}{d\ell} + \frac{2}{\ell} \quad , \tag{18}$$

is well-known in geodesy as Molodensky operator (see, for instance, Neyman, 1979) and it was introduced first for the basic boundary problem of geodesy in the next form

$$M[T] = \frac{dT}{dr} + \frac{2T}{r} = -\Delta g \quad , \tag{19}$$

where T is the anomalous potential, Δg is the gravity anomaly, r is the radius-vector of an *external* point (the parameter ℓ represents the radius-vector of an *internal* point).
The expression (19) is used for the determination of T on the ground of known gravity anomalies. In the expression (17) we have as unknown values both the density and gravity inside the Earth. Nevertheless, if the gravity g is known we get a simple rule for the computation of radial density profile in accordance with Poisson's equation. If gravity is known approximately, we get one of the most important additional information for a stable creation of the density models. So, one of our next steps will connected with the gravity distribution inside the Earth.

5. Some remarks on the regular Darwin's law

If we want to avoid a singularity at the origin in (11), this function may be transform to the expression

$$\rho(x) = C \cdot x^{-f(x)} = C \cdot \exp(-f(x)\ln x) \quad , \tag{20}$$

where $f(x)$ is any suitable function. Such a function can represent a regular form of Darwin's law without a singularity at the origin (Marchenko and Lelgemann, 1997). The expression (20) may be considered as an exponential function.
Taking into account the relationships (14), (15) we may try to insert into (20) another function $f(x) = F(x)/\ln(x)$ (in particular, $F(x) = \gamma^2 x^2$) that leads on the whole again to

$$\rho(x) = C \cdot \exp(-F(x)) \quad , \tag{21}$$

the solution (14) of Williamson-Adams equation and to the considered case (15) in particular.
Note that the direct integration of (20) is impossible for mass (1), for moment of inertia (4), etc. The expression (21) in the form of (15):

$$\rho(\ell) = \rho_0 \exp\left(-\gamma^2 x^2\right) \quad , \tag{22}$$

admits according to (1) and (4) the next remarkable expressions for the mass

$$M(\ell)=\frac{4\pi\rho_0 R^3}{\gamma^2}\left[\frac{\sqrt{\pi}\cdot\mathrm{erf}(\gamma\cdot x)}{4\gamma}-\frac{x}{2\rho_0}\rho(\ell)\right]\ , \tag{23}$$

and for the mean moment of inertia

$$I(\ell)=\frac{8\pi\rho_0 R^5}{3\gamma^4}\left[\frac{3\sqrt{\pi}\cdot\mathrm{erf}(\sqrt{\gamma}\cdot x)}{8\gamma}-\frac{x}{4\rho_0}\rho(\ell)\cdot\left(2\gamma^2x^2+3\right)\right]=\frac{R^2}{\gamma^2}\left[M(\ell)-\frac{4\pi\ell^3}{3}\rho(\ell)\right]\ , \tag{24}$$

where erf(z) is the integral of the Gaussian distribution from 0 to z or the probability integral with the density distribution according to (22).
Thus we come to a remarkable result: one of solutions of Williamson-Adams equation in the regular Darwin's form is nothing else but the famous Gaussian distribution, which may be approximated by the Roche's model, represented the possible solution of the Clairaut's equation.
In spite of the difference between considered above various expressions for density, we come to their exponential nature on the whole. Roche's model we may treat now as a truncated Taylor series of them.

6. Saigey's theorem and the Roche's model

According to the so-called Saigey theorem the gravity $g(\ell)$ has a maximum inside the Earth. We shall use the Roche's model as a basic tool for next study. So that, it is necessary to find such a point(s), where the radial derivative $\frac{dg(\ell)}{d\ell}$ is equal to zero. As a result, for the stationary point(s) we get the well-known expression

$$\frac{dg(\ell)}{d\ell}=\frac{4\pi G}{3}\left(D(\ell)+\ell\frac{dD(\ell)}{d\ell}\right)=4\pi G\left(\rho(\ell)-\frac{2}{3}D(\ell)\right)=0 \quad\Rightarrow\quad \rho(\ell)=\frac{2}{3}D(\ell)\ . \tag{25}$$

Now applying the Roche's model (8) or (16) to (25) we get immediately

$$D(\ell)=a+\frac{3\cdot b}{5}\left(\frac{\ell}{R}\right)^2=a+\frac{3\cdot b}{5}x^2\ , \tag{26}$$

and the solution of (25) for the parameter x

$$x=\frac{\ell}{R}=\frac{\sqrt{5}\cdot\sqrt{a}}{3\cdot\sqrt{-b}}\ . \tag{27}$$

Note that this root of (25) corresponds to (9) and $a>0$. In this case the sign of b must be negative: $b<0$. Moreover applying such dimensionless $x\in[0,1]$ and (27) the following inequality

$$\frac{a}{-b}\le\frac{9}{5}\ , \tag{28}$$

may be found for the coefficients of the Roche's model. Note only that the sign of the second radial derivative follows from the coefficient b. For this reason $\frac{d^2 g(\ell)}{d\ell^2} < 0$ in the point (27) and our function $g(\ell)$ has a maximum only at this point.

7. Piecewise Roche's model

If a suitable stratification of the Earth leads to its division into m shells, first we shall represent the density distribution by own Roche's model within every shell separately

$$\rho_i(x) = a_i + b_i x^2 \quad , \qquad i = 1,2,\ldots m \, . \tag{29}$$

Inserting (29) into the expressions (1), (2), and (4) we get finally the recurrence formulae for the mass, the mean density and mean moment of inertia, respectively:

$$M_{1,m}(\ell) = M_{1,m-1}(\ell_{m-1}) + [M_m(\ell) - M_m(\ell_{m-1})] \quad , \qquad (\ell_{m-1} \le \ell \le R) \, , \tag{30}$$

$$D_{1,m}(\ell) = \left(\frac{\ell_{m-1}}{\ell}\right)^3 D_{1,m-1}(\ell_{m-1}) + \left[D_m(\ell) - \left(\frac{\ell_{m-1}}{\ell}\right)^3 D_m(\ell_{m-1})\right] \, , \tag{31}$$

$$I_{1,m}(\ell) = I_{1,m-1}(\ell_{m-1}) + [I_m(\ell) - I_m(\ell_{m-1})] \quad , \tag{32}$$

where for the piecewise Roche's model

$$M_i(\ell) = \frac{4\pi}{3}\ell^3\left[a_i + \frac{3}{5}b_i x^2\right] \quad , \qquad M_{1,1}(\ell) = M_1(\ell) \, , \tag{33}$$

$$D_i(\ell) = \left[a_i + \frac{3}{5}b_i x^2\right] \, , \qquad D_{1,1}(\ell) = D_1(\ell) \, , \tag{34}$$

$$I_i(\ell) = \frac{8\pi}{3}\ell^5\left[\frac{a_i}{5} + \frac{b_i}{7}x^2\right] \, , \qquad I_{1,1}(\ell) = I_1(\ell) \, , \tag{35}$$

starting from the first shell $(0 \le \ell \le \ell_1)$. In these formulae ℓ_j (j =1,2,...m-1) are the fixed radius-vectors, where jumps of radial density are presupposed. The recurrence formulae for gravity is based on the expressions (3) and (31):

$$g_{1,m}(\ell) = \frac{4 \cdot \pi \cdot G}{3}\ell \cdot D_{1,m}(\ell) \quad , \qquad (\ell_{m-1} \le \ell \le R) \, , \tag{36}$$

again starting from the first shell.
For the recurrence formulae of the seismic parameter Φ and it jumps first we shall find

$$\frac{d \ln \rho_i(\ell)}{d\ell} = \frac{2b_i\ell}{R^2\rho_i(\ell)} \quad . \tag{37}$$

Further by applying the Williamson-Adams equation (13) for the piecewise model (29) in view of (38) after some manipulations we get

$$\Phi_{1,m}(\ell) = -\frac{2\cdot\pi\cdot G\cdot R^2}{3\cdot b_m}\rho_m(\ell)\cdot D_{1,m}(\ell) \quad , \qquad (\ell_{m-1} \le \ell \le R) \, , \tag{38}$$

$$\Phi_i(\ell) = -\frac{2\cdot\pi\cdot G\cdot R^2}{3\cdot b_i}\rho_i(\ell)\cdot D_i(\ell) \quad , \qquad \Phi_{1,1}(\ell) = \Phi_1(\ell) \, , \quad (0 \le \ell \le \ell_1) \, . \tag{39}$$

By the definition (5) the parameter Φ must be always positive and we shall consider the ratio

$$\frac{\Phi_{1,i}(\ell_{j-1})}{\Phi_{1,i-1}(\ell_{j-1})} = \frac{b_i}{b_{i-1}}\frac{\rho_i(\ell_{j-1})}{\rho_{i-1}(\ell_{j-1})} > 0 \, , \tag{40}$$

which must be positive for each boundary of two shells. From this inequality together with (9), (28) (for one shell) we come to a remarkable results: *all coefficients a_i will be positive and all coefficients b_i will be negative for the piecewise Roche's model of density.*
Finally we may compute the seismic jump of Φ at the j - boundary

$$\Delta\Phi = \Delta\Phi_{i,i+1} = \Phi_i(\ell_j) - \Phi_{i+1}(\ell_j) = -\frac{2\cdot\pi\cdot G\cdot R^2}{3} D_{1,i}(\ell_j)\left[\frac{a_i}{b_i} - \frac{a_{i+1}}{b_{i+1}}\right] \, . \tag{41}$$

This formula may use as the additional condition between the coefficients of every shell, because the left hand side of (41) is known from seismic data.

8. First iteration for piecewise density distribution

Now we recollect (see, for instance, Moritz, 1990) that „any global density law must satisfy three basic conditions:

1. It must provide the correct total mass or, equivalently, the mean density;
2. It must give the value for the mean moment of inertia;
3. It must reproduce the density at the base of continental layers, which may be taken as about 3.2 to 3.3 g/cm^3, e.g. the conventional density just below Mohorovichich discontinuity much used in isostasy ρ_1 =3.27 g/cm^3 „.

These three conditions may lead to the construction of the continuos radial density distribution. First two conditions can apply for the determination of the continuos Roche's model. In this case we get a remarkable expression for the coefficient b of such a model

$$\left.\begin{aligned} b &= \frac{5}{3}[D - \rho_0] \\ a &= \rho_0 \end{aligned}\right\} \tag{42}$$

Nevertheless, we may add according to (39) the additional condition for density at the origin, which will depend on the observe value of Φ:

$$\Phi(0) = -\frac{2\cdot\pi\cdot G\cdot R^2}{3\cdot b_1} a_1^2 \, . \tag{43}$$

and use then forth conditions for determinations of the coefficients of two (m=2) models (29). We presuppose also that the first model will describe the density on the interval [0, 3480km] and the second model is valid for the interval [3480km, 6371km]. Now according to three condition listed above we get

$$a_1\left(\frac{\ell_1}{R}\right)^3+a_2\left[1-\left(\frac{\ell_1}{R}\right)^3\right]+\frac{3b_1}{5}\left(\frac{\ell_1}{R}\right)^5+\frac{3b_2}{5}\left[1-\left(\frac{\ell_1}{R}\right)^5\right]=D \quad , \tag{44}$$

$$\frac{2}{D}\left\{\frac{a_1}{5}\left(\frac{\ell_1}{R}\right)^5+\frac{a_2}{5}\left[1-\left(\frac{\ell_1}{R}\right)^5\right]+\frac{b_1}{7}\left(\frac{\ell_1}{R}\right)^7+\frac{b_2}{7}\left[1-\left(\frac{\ell_1}{R}\right)^7\right]\right\}=I_d \quad , \tag{45}$$

$$a_2+b_2=\rho_s \quad , \tag{46}$$

where $I_d = I/MR^2$ is the dimensionless Earth's moment of inertia, ρ_s is the surface density, $D=D(R)$ is the Earth's mean density. Because the equation (43) is non-linear, on the first step we shall add the following linear equation

$$a_1=\rho_0 \quad , \tag{47}$$

and will solve this system (44)-(47) with respect to the density (47) at the origin. On the second step the non-linear equation (43) may be solved numerically in a traditional way. After iterations, we can get these four coefficients and compute now the basic jump of the Earth's density. To our own surprise such solution of the equations (43)-(46) together with the seismic data alone provided finally (in this step) the density jump at the core/mantle boundary $\Delta\rho = 4.454$ g/cm^3, the density at the centre mass of the Earth $\rho = 12.953$ g/cm^3, and the remarkable restoring of the main behaviour of the gravity distribution according to PREM model.

For this reason after the creation of these two models we may continue such approach for the further division of the Earth and determination of the set of the models (29) which should be agreed with the whole initial information about the seismic data. Thus on the first step we may get a preliminary solution for every shell separately by the „golden section" technique (in view of the necessity of the Earth's stratification and solution of the non-linear equation (43)). The second step consists of the readjustment of these independent pieces of density to the piecewise density distribution which agrees with the set of the seismic data and other additional information about fundamental constants.

Regarding the discontinuities in the seismic velocities as sampled for PREM, we are led to the following separation into shells (Table 1) as a particular case. Based on this separation a mathematical description of the Earth's density based on the piecewise Roche's model was derived and presented in Table1. This model (see, Figure 1) can be used further for an improvement as a starting model using another – exponential solution of Williamson-Adams equation. Figure 1 reflects its good agreement with the PREM-density model, with the exception of the crust shells: we try to create on the final step a „geodetic version" of the Earth density profile with surface density ρ_s=2.67 g/cm^3.

Table 1. Piecewise Roche's density model (m=7)

Shell	a_i	b_i	ℓ_j, km	Density jump
1	13.061	-8.891	1221.5	0.558
2	12.483	-8.343	3480.0	4.392
3	6.370	-2.574	5701.0	0.314
4	6.058	-2.577	5971.0	0.228
5	5.784	-2.524	6151.0	0.080
6	6.057	-2.903	6346.6	0.476
7	6.622	-3.952		

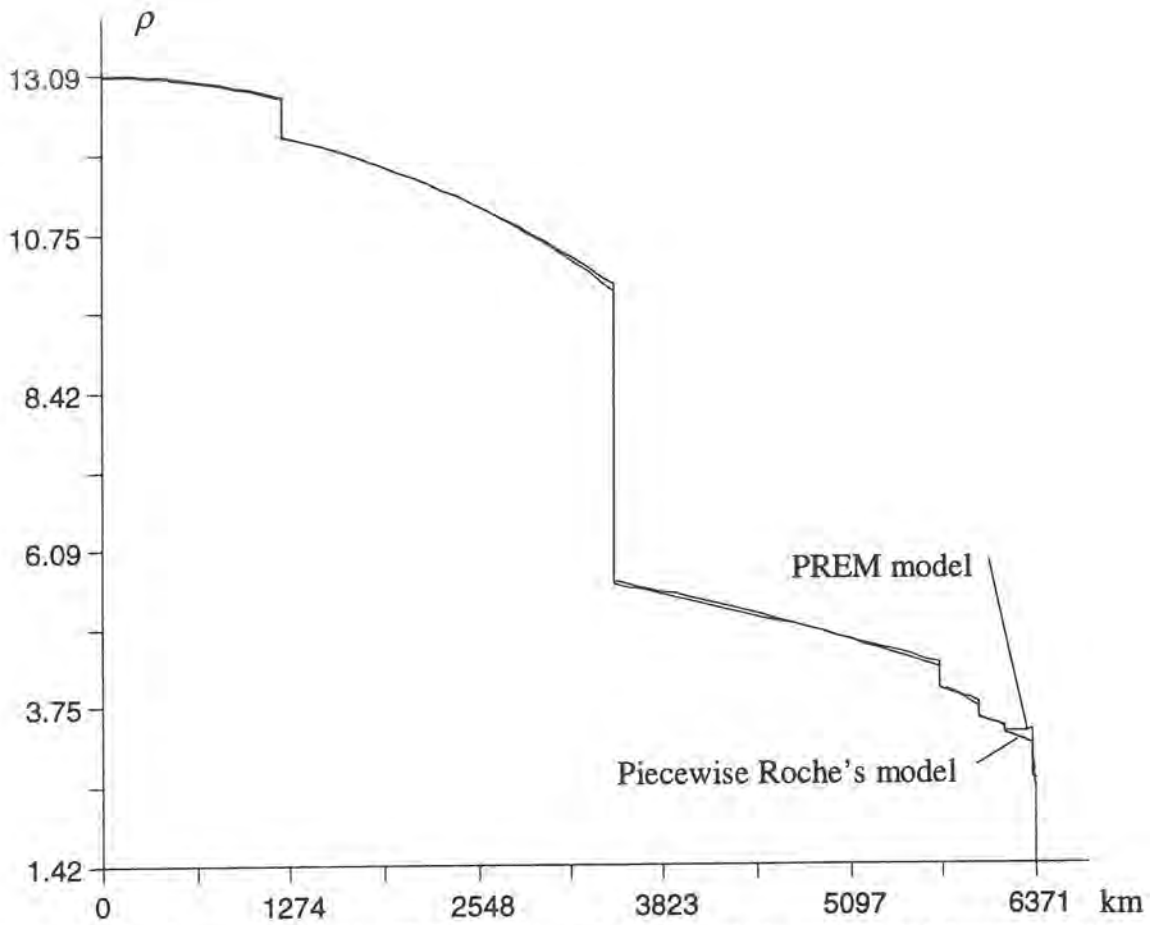


Figure 1. Comparison of the PREM-density ρ [g/cm^3], with the constructed density model.

References

Bullen, K. E. (1975) *The Earth's Density*. Chapman and Hall, London.

Darwin, G. (1884) On the figure of equilibrium of a planet of heterogeneous density. *Proceeding of the Royal Society*, Vol. XXXVI, 158-166.

Dziewonski, A.M. and Anderson, D.L. (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, Vol. 25, 297-356.

Neyman, Yr. (1979) *Variational method of Physical Geodesy*, Nedra, Moscow, (in Russian).

Marchenko, A.N., and Lelgemann, D. (1997). On the parametrization of global Earth's models. *Paper presented at the "Geodetic Week"*, Berlin, October, 1997.

Moritz, H. (1990) *The Figure of the Earth. Theoretical Geodesy and Earth's Interior*, Wichmann, Karlsruhe, 1990.

Stokes's two-boundary-value-problem

Zdeněk Martinec

It is a common belief that, after removing the first-degree spherical harmonics from the gravitational potential, only a regularization of the downward continuation of a high frequency part of the gravity is necessary to guarantee the existence of a unique solution to the Stokes boundary-value problem for gravimetric determination of the geoid. In this paper, we will deal with the original formulation of the problem prior to the downward continuation of gravity. We intend to demonstrate numerically that, besides the spherical harmonics of degree one, the existence of the solution is not also guaranteed for higher-degree harmonics. This lack of guaranty is due to the fact that the input data – the surface gravity and the potential of the geoid – are prescribed on different boundaries.

1. Formulation of the Stokes two-boundary-value problem

Let the geocentric radius of the geoid S_g be described by an angularly dependent function $r = r_g(\Omega)$, where (r, Ω) are the geocentric spherical coordinates, i.e., $(r_g(\Omega), \Omega)$ are points lying on the geoid. We will assume that the function $r_g(\Omega)$ is not known. Let $H(\Omega)$ be the height of the Earth's surface above the geoid reckoned along the geocentric radius. Unlike the geocentric radius of the geoid, we will assume that $H(\Omega)$ is a known function. Finally, let the following quantities be given: the gravity $g_S(\Omega)$ measured on the Earth's surface, the density $\varrho(r, \Omega)$ of the topographical masses (the masses between the geoid and the Earth's surface), and the gauge value W_0 of the gravity potential on the geoid.

The question we pose is: how to determine the gravity potential $W(r, \Omega)$ **inside** and **outside** the topographical masses and the radius $r_g(\Omega)$ of the geoid? The problem is governed by the Poisson equation with the boundary conditions given on the free boundaries S_t and S_g coupled by means of height $H(\Omega)$:

$$\nabla^2 W = -4\pi G\varrho + 2\omega^2 \qquad \text{outside } S_g\ , \qquad (1)$$

$$|\text{grad}\, W| = g_S \qquad \text{on } S_t\ , \qquad (2)$$

$$W = W_0 \qquad \text{on } S_g\ , \qquad (3)$$

$$W = \frac{1}{2}\omega^2 r^2 \sin^2\vartheta + \frac{GM}{r} + O\left(\frac{1}{r^3}\right) \qquad r \to \infty\ , \qquad (4)$$

where G is the gravitational constant, M is the mass of the Earth, and ϱ is equal to zero outside the Earth. The first-degree harmonics are left out from the potential W because of the geocentric coordinate system.

Martinec and Matyska (1997) have shown that the boundary-value problem (1)–(4) can be linearized with respect to the anomalous potential T^h such that

$$\nabla^2 T^h = 0 \qquad \text{outside } S_g, \tag{5}$$

$$\left.\frac{\partial T^h}{\partial r}\right|_P + \frac{2}{r_Q} T^h_{P_g} - \epsilon_h(T^h_P) - \epsilon_\gamma(T^h_{P_g}) = -\Delta g^h + \sum_{m=-1}^{1} a_{1m} Y_{1m}(\Omega)\ , \tag{6}$$

$$T^h = \frac{c}{r} + O\left(\frac{1}{r^3}\right) \qquad \text{for } r \to \infty\ , \tag{7}$$

where P, P_g and Q are the points on the Earth's surface, the geoid and the level ellipsoid, respectively, ϵ_h and ϵ_γ are ellipsoidal corrections (e.g., Jekeli, 1981), Δg^h is the Helmert gravity anomaly and a_{1m} are constants to be determined.

2. Numerical investigations

The original problem (1)–(4) as well as the problem described by eqns.(5)–(7) are scalar nonlinear free boundary-value problems since the radial coordinate of the geoid is one of the unknowns to be determined. Having some approximation of geoid, it is easy to transform the latter free boundary-value problem to a problem with fixed boundaries. For example, replacing P_g by r_Q, r_Q being the radius of the normal point Q, and P by $r_Q + H(\Omega)$ in eqn.(6) yields the ellipsoidal approximation of the Stokes two-boundary-value problem, where eqns.(5)–(7) serve to determine T^h; Bruns's formula then gives the geoidal height N. Another possibility, most often used in geoid height computations, is to approximate the geoid in the boundary condition (6) by a mean sphere with radius $R = 6371$ km. This means the radius of the point P_g is replaced by R and radius of the point P by $R + H(\Omega)$. The relative error introduced by this spherical approximation is of the order of 3×10^{-3} in the classical problems (Heiskanen and Moritz, 1967, sect.2-14), which then causes a long-wavelength error of at most 0.5 metres in geoidal heights. In regional problems, where only shorter wavelengths are to be determined, this approximation is often reasonable. In the following numerical tests we will employ the spherical approximation of boundary condition (6) for its simplicity. We intend to concentrate on the effects connected with the 'two-boundary nature' of this condition that appear only in a very short wavelength part of the solution.

The solution to the Laplace equation (5) with the condition (7) can be represented as a series of solid spherical harmonics $r^{-j-1}Y_{jm}(\Omega)$,

$$T^h(r,\Omega) = \sum_{\substack{j=j_{min}\\ j\neq 1}}^{j_{max}} \sum_{m=-j}^{j} T_{jm} \left(\frac{R}{r}\right)^{j+1} Y_{jm}(\Omega)\ , \tag{8}$$

where $j_{min}(\geq 0)$ and j_{max} are the respective minimum and maximum cut-off degrees, $Y_{jm}(\Omega)$ are spherical harmonics of degree j and order m, and T_{jm} are the coefficients of potential T^h to be determined. In order to normalize the potential coefficients T_{jm}, we have introduced the mean Earth's radius R into the expansion (8). Equation (6) in the spherical approximation then becomes

$$\frac{1}{R} \sum_{\substack{j=j_{min}\\ j\neq 1}}^{j_{max}} \sum_{m=-j}^{j} \left[(j+1)\left(\frac{R}{R+H(\Omega)}\right)^{j+2} - 2 + e_0^2(3\cos^2\vartheta - 2)\right] Y_{jm}(\Omega) T_{jm} +$$

$$+\frac{e_0^2}{R} \sum_{\substack{j=j_{min}\\ j\neq 1}}^{j_{max}} \sum_{m=-j}^{j} \sin\vartheta\cos\vartheta \frac{\partial Y_{jm}(\Omega)}{\partial\vartheta} T_{jm} = \Delta g^h - \sum_{m=-1}^{1} a_{1m} Y_{1m}(\Omega)\ . \tag{9}$$

This boundary condition must hold in any direction Ω. In order to ensure it, we will employ the Galerkin method in which eqn.(9) can be rewritten as a system of linear algebraic equations for coefficients T_{jm}:

$$\boldsymbol{A}\boldsymbol{m} = \boldsymbol{d} \,, \tag{10}$$

where $\boldsymbol{m}$ is a column vector composed of potential coefficients T_{jm}, i.e.,

$$\boldsymbol{m} := \{T_{jm} | j = j_{min}, ..., j_{max}, j \neq 1, m = -j, ..., j\} \,, \tag{11}$$

$\boldsymbol{A}$ is the matrix composed of the weighted left-hand side of eqn.(9),

$$A_{j_1 m_1, jm} :=$$

$$:= \int_{\Omega_0} \left[(j+1) \left(\frac{R}{R + H(\Omega)} \right)^{j+2} - 2 + e_0^2 (3\cos^2\vartheta - 2) \right] Y_{jm}(\Omega) Y^*_{j_1 m_1}(\Omega) d\Omega + \tag{12}$$

$$+ e_0^2 \int_{\Omega_0} \sin\vartheta \cos\vartheta \frac{\partial Y_{jm}(\Omega)}{\partial \vartheta} Y^*_{j_1 m_1}(\Omega) d\Omega \,,$$

and $\boldsymbol{d}$ is a column vector of weighted right-hand side of eqn.(9),

$$d_{j_1 m_1} := R \int_{\Omega_0} \Delta g^h(\Omega) Y^*_{j_1 m_1}(\Omega) d\Omega \,, \tag{13}$$

where $j_1 = j_{min}, ..., j_{max}, j_1 \neq 1$, and $m_1 = -j_1, ..., j_1$.

2.1. An example: constant height

Let us first consider a simple, but illustrative, case when $H = H_0 = const.$ over the Earth, and $e_0^2 = 0$. Introducing function

$$K_j(H_0) := (j+1) \left(\frac{R}{R + H_0} \right)^{j+2} - 2, \quad \text{for } j \geq 2 \,, \tag{14}$$

the transfer matrix $A_{j_1 m_1, jm}$ between unknown parameters T_{jm} and the gravity anomalies Δg^h on the right-hand side of eqn.(9) becomes $A_{j_1 m_1, jm} = K_j(H_0)\, \delta_{j j_1} \delta_{m m_1}$ and thus

$$T_{jm} = \frac{R}{K_j(H_0)} \int_{\Omega_0} \Delta g^h(\Omega) Y^*_{jm}(\Omega) d\Omega \,. \tag{15}$$

Since $0.998 < R/(R + H_0) < 1$ for the Earth, it is clear that $\lim_{j\to\infty} K_j = -2$ for any fixed $H_0 > 0$. On the other hand, $K_j > 0$ for low degrees j because $0.976 < K_2 < 1$. This means that there is a range of j's in which K_j is zero or near zero. For those j's the solution of eqn.(10) is unstable or even does not exist once $K_j = 0$.

Let us estimate the range of j's for which the solution of eqs.(10) becomes unstable for this simple example. Figure 1 plots the values of K_j for height H_0 equal to 1 km, 5 km and 10 km. We can see that the increase of K_j with increasing j is confined to low degrees j and then K_j starts to decrease to its limiting value -2. That is why, the determination of disturbing potential T^h is stable only in some part of the spectral domain. The width of the stable part grows with decreasing H_0.

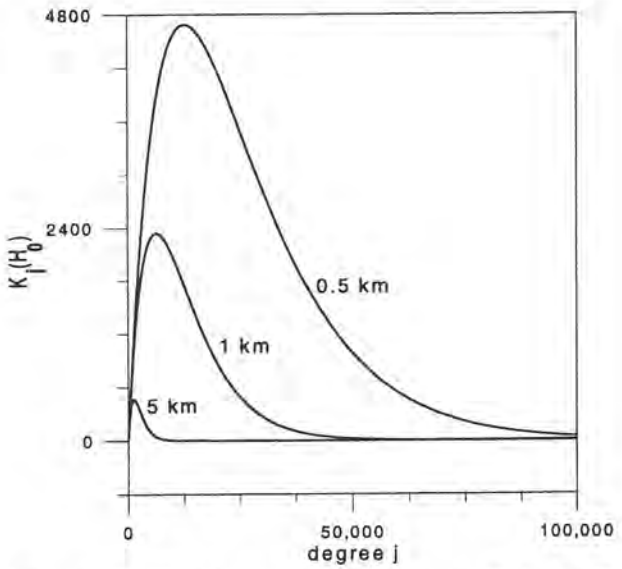


Figure 1: Transfer function $K_j(H_0)$ between unknown coefficients T_{jm} and gravity anomalies Δg^h for $H_0 = 1$ km, 5 km, and 10 km.

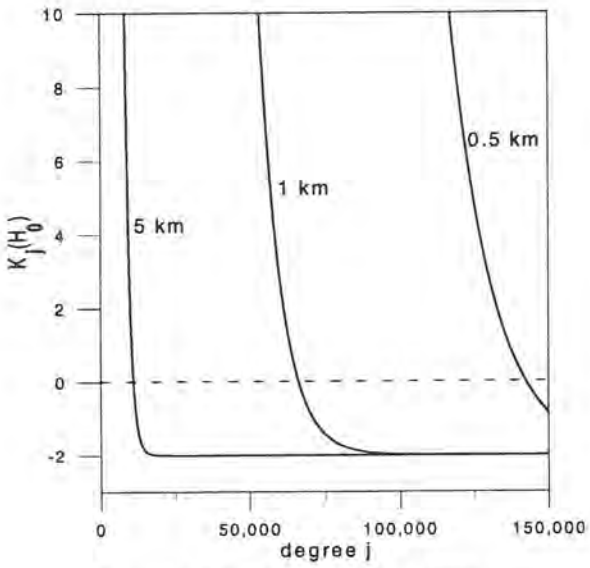


Figure 1a: A detail of Figure 1.

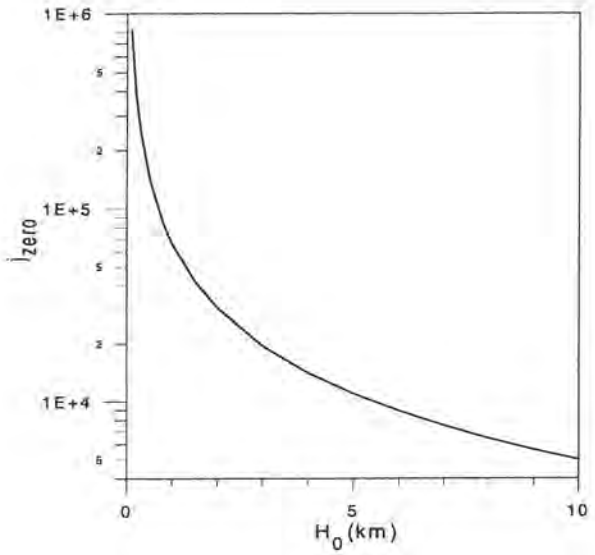


Figure 2: The roots j_{zero} of function $K_j(H_0)$ for $H_0 \in (100\,\mathrm{m}, 10^4\,\mathrm{m})$.

Figure 2 plots those j_{zero} for which function $K_j(H_0)$ vanishes. For such degrees matrix $\boldsymbol{A}$ is singular and the solution of system of equations (10) does not exist. Since spherical degree j corresponds to a given resolution $\Delta\Omega$ in a spatial domain, $\Delta\Omega = \pi/j$, we may also convert critical degree j_{zero} to a critical spatial resolution size $\Delta\Omega_{zero}$, $\Delta\Omega_{zero} = \pi/j_{zero}$, for which the solution to our problem does not exist. Figure 2 shows that, for instance, $j_{zero} = 10980$, for $H_0 = 5$ km, and the critical spatial resolution size is $\Delta\Omega_{zero} \doteq 1$ arcmin. To interpret the result in other words, let us imagine that the Earth's topography is a Bouguer spherical shell with a constant height of 5 km above the geoid and the Stokes two-boundary-value problem is solved in a spatial domain such that the potential $T^h(R, \Omega)$ is parameterized by discrete values $T^h(R, \Omega_i)$ in a regular angular grid with grid step size $\Delta\Omega$. Then the solution to the Stokes two-boundary-value problem will not exist if the grid step size $\Delta\Omega$ of the parameterization of T^h is less than or equal to the critical step size $\Delta\Omega_{zero}$, i.e., of about 1 arcmin in our example, even though the surface gravity data would be known continuously on the Earth's surface.

To map the non-existence of the solution for regional geoid determination and for a more realistic model of the Earth's topography, we need to set up and to solve the system of eqn.(10) for high degrees and orders ($j_{max} = 10^4 - 10^5$). This leads to computational difficulties because of huge consummation of computational time and memory; with today's computer equipment it is impossible to carry out the analysis of the existence for such a general case. Thus, we are forced to approximate the Earth's surface by a simplified model of axisymmetric geometry. By making use of the analysis of this simplified case, we will attempt to estimate the range of critical spectral degrees j_{zero} for the actual case.

2.2. Axisymmetric geometry

Let the height $H(\vartheta, \lambda)$ of the Earth's surface above the geoid is modelled by zonal as well as tesseral and sectoral spherical harmonics of the global digital terrain model TUG87 (Wieser, 1987) cut at degree 180. To create a rotational symmetric body, axisymmetric height $H(\vartheta)$ will be generated by height $H(\vartheta, \lambda)$ taken along a fixed meridian $\lambda = \lambda_0$. In the case of an axisymmetric surface, the elements $A_{j_1 m_1, jm}$ of matrix $\boldsymbol{A}$ do not depend on angular orders m and m_1; they can be written as

$$A_{j_1 j} = \int_{\vartheta=0}^{\pi} \left[(j+1) \left(\frac{R}{R + H(\vartheta)} \right)^{j+2} - 2 + \right.$$

$$+e_0^2(3\cos^2\vartheta - 2)\Big] P_j(\cos\vartheta)P_{j_1}(\cos\vartheta)\sin\vartheta d\vartheta + \tag{16}$$

$$+e_0^2\int_{\vartheta=0}^{\pi}\sin\vartheta\cos\vartheta\,\frac{dP_j(\cos\vartheta)}{d\vartheta}P_{j_1}(\cos\vartheta)\sin\vartheta d\vartheta\ .$$

Note that the elements $A_{j_1 j}$ can only be evaluated by a method of numerical quadrature.
To analyse the posedness of the Stokes two-boundary-value problem, we will employ the eigenvalue analysis of matrix $\boldsymbol{A}$. According to this method, a non-symmetric matrix $\boldsymbol{A}$ can be decomposed to the product of three matrices,

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}\ , \tag{17}$$

where the columns of matrix $\boldsymbol{U}$ are formed from the right eigenvectors of $\boldsymbol{A}$, the rows of $\boldsymbol{U}^{-1}$ are formed from the left eigenvectors of $\boldsymbol{A}$, and the diagonal matrix $\boldsymbol{\Lambda}$ consists of eigenvalues of $\boldsymbol{A}$. We have employed subroutines BALANC, ELMHES and HQR (Press et al., 1992, sect.11.5 and 11.6) to find the eigenvalues of a non-symmetric matrix $\boldsymbol{A}$.

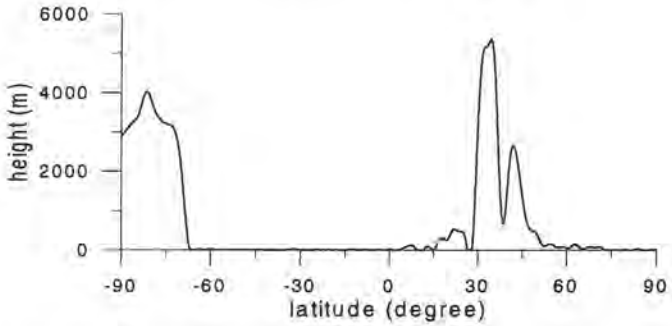


Figure 3: The meridian profile $\lambda = 80^o$ of topographical height $H(\vartheta, \lambda)$ generated by the global digital terrain model TUG87 (Wieser, 1987) cut at degree 180. This profile is used to create a body with the axisymmetric geometry of external surface.

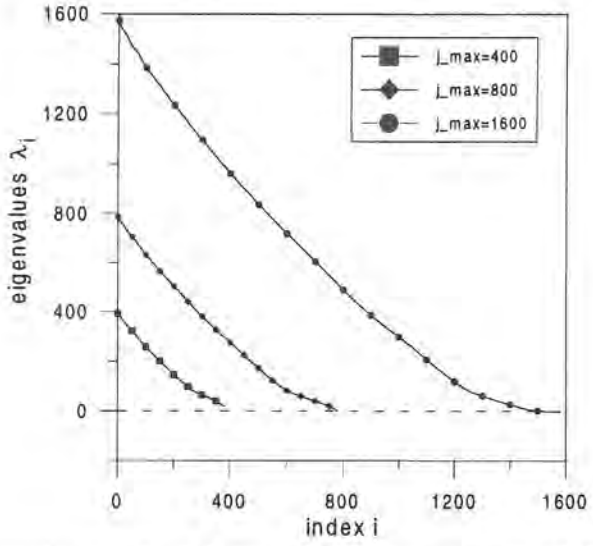


Figure 4: The eigenvalue spectra of matrix $\boldsymbol{A}$ for various cut-off degrees j_{max} and a body with axisymmetric surface generated by height $H(\vartheta, \lambda = 80^o)$ multiplied by 10 ($j_{min} = 21$). The ellipsoidal corrections ϵ_h and ϵ_γ are equal to zero.

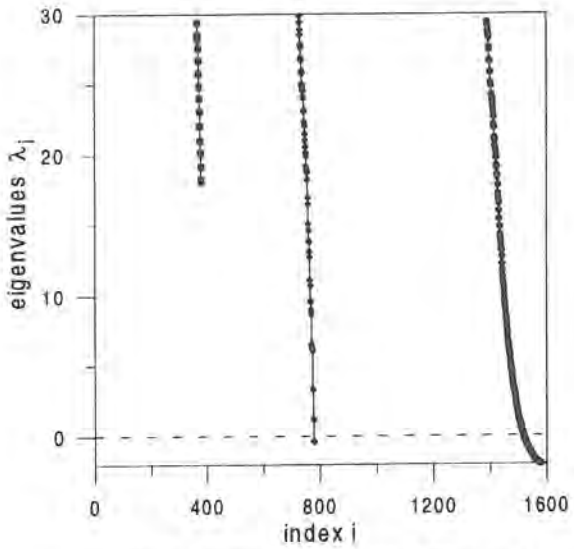


Figure 4a: A detail of Figure 4.

Figure 3 shows the topographical height $H(\vartheta, \lambda_0)$ along the meridian profiles $\lambda_0 = 80^o$ reaching value $H_{max} = 5353$ metres. The consequent Figure 4 shows a plot of the eigenvalues of matrix $\boldsymbol{A}$ for an axisymmetric body with the outer surface generated by this meridian profile. In order to avoid high degrees j, and thus, be able to perform the eigenvalue analysis in real CPU time, we multiply function $H(\vartheta)$ by a factor of 10. The minimum spherical degrees j_{min} of the potential series (8) is $j_{min} = 21$, which models the situation when low-degree harmonics of potential T^h are determined by another approach, e.g., when considering a satellite gravitational model. In Figure 4, where we further put the eccentricity of the level ellipsoid equal to zero, $e_0 = 0$, we change the maximum cut-off degree j_{max} of the disturbing potential T^h and plot eigenvalues of matrix $\boldsymbol{A}$ ordered according to their size (note that the eigenvalues are real numbers in this particular case). Inspecting Figure 4 we can observe that the eigenvalue spectrum of matrix $\boldsymbol{A}$ intersects the zero level starting from degree $j_{zero} \doteq 800$. Once the cut-off degree j_{max} of the spherical harmonic expansion (8) of potential T^h is greater or equal to j_{zero}, the eigenvalue spectrum of $\boldsymbol{A}$ contains a null eigenvalue or an eigenvalues of a very small size. The matrix $\boldsymbol{A}$ becomes ill-conditioned or even singular and the inverse $\boldsymbol{A}^{-1}$ may be distorted by large round-off errors or may not exist at all; in such a case the Stokes two-boundary-value problem does not have a unique and stable solution. As in the preceeding section, the critical degree j_{zero} can again be converted to the critical spatial discretization size $\Delta\Omega_{zero}$ for a case when the Stokes two-boundary-value problem is solved in a spatial domain.

The next test investigates the influence of the ellipsoidal corrections terms ϵ_h and ϵ_γ on the posedness of matrix $\boldsymbol{A}$. We choose the same body as in the preceeding example together with $j_{min} = 21$ and $j_{max} = 1600$ and compute the eigenvalues of matrix $\boldsymbol{A}$ putting $e_0^2 = 0$ and $e_0^2 = 0.006694$, respectively. Figure 5 shows those eigenvalues the magnitudes of which are smaller than 3. (Note that the eigenvalues of $\boldsymbol{A}$ for the case $e_0^2 = 0.006694$ are complex numbers.) We can observe that the eigenvalue spectrum of $\boldsymbol{A}$ changes significantly when e_0^2 differs from zero: there is no null eigenvalue and the magnitude of the smallest eigenvalue is larger than 1. In other words, the ellipsoidal corrections ϵ_γ and ϵ_h

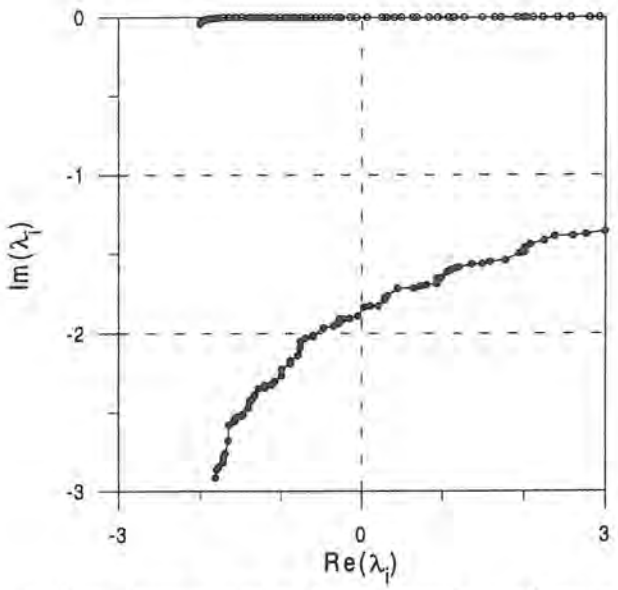


Figure 5: The real vs. imaginary parts of the eigenvalues of matrix $\boldsymbol{A}$ with (lower branch) and without (upper branch) the ellipsoidal corrections ϵ_h and ϵ_γ. The axisymmetric body is the same as that considered in Figure 4 ($j_{min} = 21$, and $j_{max} = 1600$).

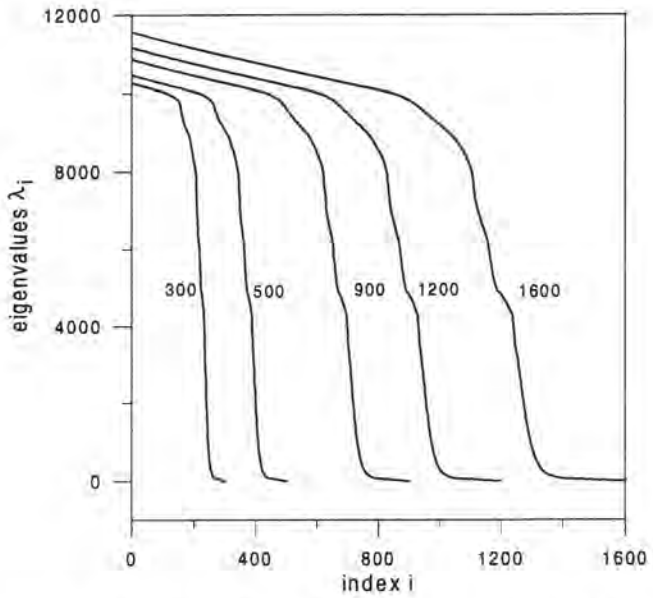


Figure 6: The eigenvalue spectra of matrix $\boldsymbol{A}$ for various cut-off degrees $j_{max} = j_{min} + \Delta j$, $\Delta j = 300, 500, ..., 1600$, and a body with axisymmetric surface generated by height $H(\vartheta, \lambda = 80^o)$ ($e_0^2 = 0$ and $j_{min} = 10000$).

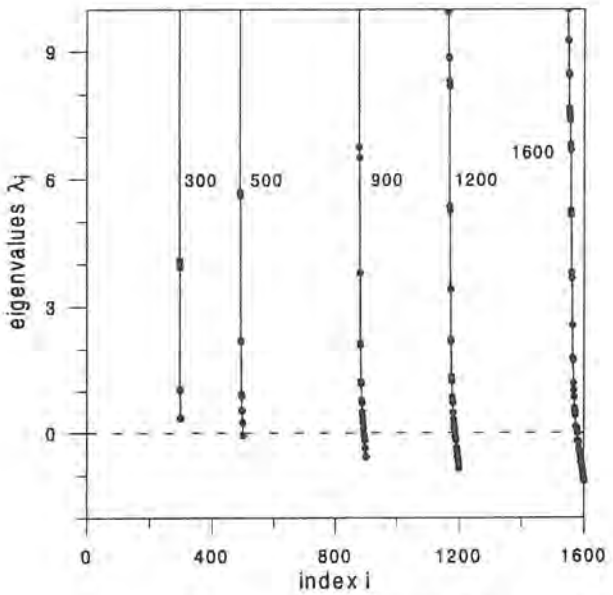


Figure 6a: A detail of Figure 6.

act as regularization factors removing the ill-posedness of matrix $\boldsymbol{A}$. It also means that, in this particular case, ϵ_γ and ϵ_h cannot be subtracted from the right-hand side of eqn.(10) as known quantities determined a priorily by using a known global gravitational model of the Earth; such usage of ellipsoidal corrections is often recommended in real geoid computations.

In order to create a more realistic example, we use the same profile of topographical height as plotted in Figure 3, but now, in contrast with preceeding example, we will not multiplied height $H(\vartheta)$ by 10. In this case, it is not possible to carry out the eigenvalue analysis of matrix $\boldsymbol{A}$ starting from degree $j_{min} = 21$ and going up to degrees $j_{max} \approx 10^4 - 10^5$ due to a huge consummation of computer time and memory. We have to confine ourselves to a smaller range of sought spherical harmonics. That is why we choose $j_{min} = 10000$ and j_{max} in the range between 10300 and 11600. The results for the case $e_0^2 = 0$ are shown in Figure 6. We can again observe that eigenvalue spectra intersect the zero-level starting at degree $j_{zero} \doteq 10500$. It means that whenever $j_{max} \geq j_{zero}$, the spectrum of matrix $\boldsymbol{A}$ contains an eigenvalue which is very close or equal to zero. Consequently, matrix $\boldsymbol{A}$ becomes ill-conditioned or even singular. Putting $e_0^2 = 0.006694$ (this case is not plotted here) has a similar stabilization effect as in the case shown in Figure 5.

To carry out the eigenvalue analysis of matrix $\boldsymbol{A}$ needs a lot of computer time. However, the critical spherical degree j_{zero} for which the existence of the solution to the Stokes two-boundary-value problem is not guaranteed can be estimated by analysing the existence of a solution for a model with a constant topographical height over the world. If we replace H_0 in the example in section 2.1. with the maximum topographical height H_{max}, then such an estimate j_{const} obviously underestimates the actual j_{zero}, i.e., it is too pessimistic, and hence it holds

$$j_{zero} \geq j_{const} \ , \tag{18}$$

where j_{const} is determined by the roots of function $K_j(H_{max})$ given by eqn.(14), i.e., j_{const} satisfies the equation

$$(j_{const} + 1) \left(\frac{R}{R + H_{max}} \right)^{j_{const}+2} - 2 = 0 \ . \tag{19}$$

For the examples in Figures 4 and 6, we obtain $j_{const} \doteq 698$ when $H_{max} = 53530$ metres, and $j_{const} \doteq 10158$ when $H_{max} = 5353$ metres. We have already learnt that the actual critical numbers are $j_{zero} \doteq 800$ and $j_{zero} \doteq 10500$, respectively. So, the criterion (18) estimates j_{zero} quite well.

3. Conclusion

This paper formulated and discussed the existence of a solution to the Stokes two-boundary-problem for geoid determination. We considered the boundary condition (6) relating to this problem without assuming that the surface gravity data had been continued from the Earth's surface to the geoid. The boundary condition (6) has not a usual form, because it contains the unknown anomalous potential referred to both the Earth's surface and the geoid coupled by the known topographical height. The numerical analysis of the 'two-boundary' condition (6) performed for a simplified model of the Earth's surface has revealed that the transfer matrix between the unknown potential on the geoid and the surface gravity anomalies may become ill-conditioned or even singular at a certain critical wavelength of a **finite** length. The existence of solution is not guaranteed for this critical geoidal wavelength. Once this ill-posed case occurs, to obtain a bounded and non-oscillating solution, the Stokes two-boundary-value problem must be regularized in such a way that this critical geoidal wavelength and its vicinity are excluded from the solution. We have given an estimate of critical geoidal wavelength; for the highest part of the Earth's surface, the critical geoidal wavelength is about 1 arcmin.

References

Heiskanen, W.H. and H. Moritz (1967). *Physical Geodesy.* W.H.Freeman and Co., San Francisco.

Jekeli, C. (1981). The downward continuation to the Earth's surface of truncated spherical and ellipsoidal harmonic series of the gravity and height anomalies. Rep. **323**, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus.

Martinec, Z. and C. Matyska (1997). On the solvability of the Stokes pseudo-boundary-value problem for geoid determination. *J. Geod.*, **71**, 103-112.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1992). *Numerical Recipes. The Art of Scientific Computing.* Cambridge Univ.Press, Cambridge.

Wieser, M. (1987). The global digital terrain model TUG87. Internal Report on Set-up, Origin and Characteristics, Inst. of Math. Geodesy, Technical Univ. of Graz, Austria.

Vom Punktlagefehler zum Qualitätsmodell

Siegfried Meier

Einleitung

In einem furiosen Auftakt Ende der Sechziger, Anfang der Siebziger hat der Jubilar innerhalb von fünf Berufsjahren gut ein Dutzend Arbeiten zur Genauigkeit eines Punktes im mehrdimensionalen euklidischen Raum vorgelegt. Mit der Punktbestimmung sowie ihrer Genauigkeit, in Polygonzügen und mittels Einschneiden, hat er sich von Anfang an zentralen Aufgaben der Geodäsie und des Vermessungswesens gewidmet. Wenn wir nun Jahrzehnte später auf anderer Stufe, abgeklärter und aus verschiedenen Blickwinkeln die Frage *Quo vadis geodesia...?* stellen, so ist es gewiß nicht abwegig zu fragen, wohin sich unsere, meist aufgabenbezogenen Vorstellungen von Genauigkeit, Zuverlässigkeit und Qualität entwickelt haben und wahrscheinlich entwickeln werden, und ob im speziellen das Punktfehlerkonzept, seinerzeit vom Jubilar verallgemeinernd behandelt, auch weiterhin eine tragfähige Grundlage sein kann oder ob wir, im Zeitalter der Massendatenverarbeitung mit noch ganz anderen Anforderungen als in der reinen Geometrie konfrontiert, auch ganz andere Qualitätsmodelle - so der aktuell gebrauchte terminus - brauchen, und ob der "'gute, alte"' Punktfehler pass ist. Das sicher nicht an Punkten, die als geometrische Abstraktion *geodätischen Wert an sich* haben, wie Netzpunkte, Paßpunkte, Grenzpunkte. An Punkten von kontinuierlichen Objekten wie Linien- und Flächenobjekten mit ausgeprägter Semantik, abgelegt etwa in objektstrukturierten Datenspeichern, dürfte dagegen das Punktfehlerkonzept allein nicht ausreichen.

Die Genauigkeit eines Punktes im mehrdimensionalen Euklidischen Raum

heißt die Habilitationsschrift von E. GRAFAREND (1969; Druck DGK 1970). Die bis dahin in Teilen, vor allem in ZfV und AVN veröffentlichten Arbeitsergebnisse wurden in dieser Schrift in einem dualen geometrisch-stochastischen Konzept mit Analogien zur Mechanik und Thermodynamik dargestellt.

Bereits die frühen Abhandlungen GRAFARENDs zeichnen sich - in der Tradition von H. WOLF stehend - durch gründliche Recherchen der Vorgänger-Arbeiten aus; man erfährt Wissenschaftsgeschichtliches aus der Geodäsie, Geometrie, Statistik und Physik. Darin ist GRAFAREND von unübertroffener Akribie, und daher können wir im Literaturverzeichnis auf eine Bibliographie der älteren Arbeiten, wie z. B. die in Tabelle 1 genannten, verzichten. Ferner sind GRAFARENDs Abhandlungen verallgemeinernd, manche universell, die benutzten Werkzeuge wirksam, die Notation rationell, die Darstellung häufig kompakt. Letzterem nachempfunden sollen Schema 1 und Tabelle 1 *kurz und bündig* an die einschlägigen Arbeiten des Jubilars zum (mittleren) Punktlagefehler, kurz auch Punktfehler genannt, erinnern.

GRAFAREND untersuchte die tensoriellen Eigenschaften des Schemas der zweiten Momente der Koordinatenfehler. Die Invarianten gegenüber Koordinatentransformationen stehen implizit als Koeffizienten in der charakteristischen Gleichung der Kovarianzmatrix $\underline{C}$ der Koordinatenfehler, sind also mit den Eigenwerten von $\underline{C}$ verknüpft (Schema 1). In den Eigenschaften der Lösungen der quadratischen bzw. der kubischen Gleichung im $I\!R^2$ bzw. im $I\!R^3$ kommt die Dualität von Invarianten und Eigenwerten ideal-symmetrisch zur Anschauung. Die Eigenwerte der Kovarianzmatrix $\underline{C}$ bestimmen die Halbachsen der Konfidenzellipsen bzw. -ellipsoide, die Invarianten des Tensors $\underline{C}$ die möglichen (skalaren) Punktfehler lt. Tabelle 1 mit unterschiedlichen Überdeckungswahrscheinlichkeiten (Österr. ZfV 1970). Die Punktfehler können neben der Genauigkeitsbeurteilung auch als Zielfunktionen der linearen Optimierung von Meßanordnungen (ZfV 1968, 1971) und der linearen Prädiktion dienen (AVN 1970). Ihre Unterschiede werden speziell im Sonderfall vollständig unkorrelierter Koordinatenfehler von gleicher Varianz σ^2 deutlich (Tabelle 1); z. B. sind im 2D-Fall die Punktfehler nach MÖHLE und WERKMEISTER gleich groß und entsprechen dem wahrscheinlichsten Wert der rayleigh-verteilten Zufallsgröße "'ebener Abstand vom Ursprung"' (vgl. auch Bull. Godsique 1970). Der um den Faktor $\sqrt{2}$ abweichende Punktfehler nach HELMERT läßt sich ebenfalls wahrscheinlichkeitstheoretisch erklären. Sei r der ebene und R der räumliche Abstand vom Ursprung, dann sind ihre Quadrate unter den o. a. Voraussetzungen chiquadratverteilt mit zwei bzw. drei Freiheitsgraden und die Erwartungswerte

$$E\{r^2\} = 2\sigma^2,\ E\{R^2\} = 3\sigma^2.$$

Mithin ist der HELMERTsche Punktfehler auch ein *mittlerer*, was gelegentlich konträr diskutiert wurde.

Die für den Geodäten wohl wichtigste Folgerung aus dem tensoriellen Transformationsverhalten des Schemas der Koordinatenfehler besteht darin, daß der Fehlertensor geometrisch als Ellipsoid interpretiert werden kann und daß mit den Mohrschen Kreisen eine Möglichkeit gegeben ist, auf einfachem Wege Fehler in beliebigen Richtungen im räumlichen oder ebenen Fehlerzustand zu bestimmen (GRAFAREND in ZfV 1967, S. 165; vgl. auch AVN 1969).

Genauigkeit, Unbestimmtheit, Information

Messungen sind nicht fehlerfrei, Meßwerte mehr oder weniger genau bzw. ungenau, vor dem Meßexperiment unbestimmt; selbst *nach* dem Experiment ist die gewonnene *Information als beseitigte Unbestimmtheit* nicht vollkommen. Faßt man wie üblich die Meßfehler als Realisierungen einer Zufallsgröße mit zugehöriger Wahrscheinlichkeitsverteilung auf, so kann man die Unsicherheit nicht nur durch gewisse Momente, sondern auch vermöge der Entropie dieser Verteilung charakterisieren. Mit dem Aufkommen neuartiger, vor allem elektronischer Meßverfahren in der Geodäsie, dem Durchgang der Daten durch verschiedenartige Kanäle von der rohen Erfassung bis zur detaillierten Auswertung, schien der Zeitpunkt gekommen zu sein, die Fehlerrechnung informationstheoretisch zu erweitern. Die Arbeiten GRAFARENDs enthalten auch dazu Fingerzeige (DGK 1970, ZfV 1971). Allerdings haben sich informationstheoretische Ansätze, abgesehen von ganz wenigen Spezialanwendungen des Maximum-Entropie-Prinzips (vgl. z. B. LEHMANN 1994), in der Fehlertheorie *nicht* einbürgern können bzw. haben sich als wenig hilfreich erwiesen. Die Gründe dafür sind offensichtlich.

Die Entropie der univariaten Normalverteilung ist dem Logarithmus der Varianz proportional. In der Entropie der multivariaten Normalverteilung sind *alle* zweiten Momente des zufälligen Vektors enthalten, und zwar als Logarithmus einer Momenten-Kombination, die gerade *det* $\underline{C}$, d. h. dem Punktfehlerquadrat nach WERKMEISTER bzw. der verallgemeinerten Varianz nach WILKS entspricht. Die Genauigkeitsmaße der Punktbestimmung haben gewohntermaßen eine

vergleichbare Maßeinheit wie die geschätzten Parameter. Es besteht kein hinlänglicher Grund, auf eine logarithmische Skala und eine andersgeartete Maßeinheit, die Basis des Logarithmus, zu wechseln. Ferner hat man, je nach Meß- und Registrierverfahren, insbesondere in der Bearbeitung hybrider Daten, sowohl mit diskreten als auch mit stetigen Verteilungen zu tun. Die Entropie einer diskreten Verteilung ist nicht negativ, jene einer stetigen Verteilung kann auch negative Werte annehmen. Schon daraus wird klar, daß beide Entropien *nicht* als Größen der gleichen Art betrachtet und *nicht* miteinander verglichen werden können.

Konstruktiver scheint der Begriff der *relativen* Information, die in einem zufälligen Objekt (zufällige Größe, zufälliger Vektor, zufällige Funktion, zufälliges Feld) über ein anderes zufälliges Objekt gleicher Art enthalten ist, zu sein; interessieren uns doch Informationen, die in einer gemessenen Größe über die zu messende, in einem ausgeglichenen Datensatz über den rohen, in einem gefilterten Signal über das ungefilterte..., enthalten ist. Doch auch in einem solchen, unseren Aufgabenstellungen durchaus angepaßten Relativkonzept, stößt man schon rein rechnerisch an Grenzen: es werden die gemeinsamen Verteilungen der statistisch verwandten Objekte gebraucht. Der Verfasser hat sich selbst an praxis-relevanten Beispielen versucht (MEIER 1990, 1991). Die Ergebnisse sind ernüchternd: die relative oder wechselseitige Information enthält als kritische Größen Kreuzkorrelationskoeffizienten bei zufälligen Größen, Vektoren und Punktfeldern oder Kreuzkorrelationsfunktionen, ggf. auch Kreuzspektren bei zufälligen Prozessen - Kenngrößen also, mit denen wir routinemäßig umgehen, so daß sich die Transformation auf logarithmische Skalen erübrigt.

Qualität von Geoinformationen

Geoinformationen werden in Bildern, speziell in signaturierten Bildern (Karten) analog präsentiert und/oder digital in vorzugsweise objektstrukturierten Datenspeichern (in Geoinformationssystemen; GIS) abgelegt. Den größten Anteil an den Datenbeständen haben die Linienobjekte, einschließlich der Grenzen (Konturen) von Flächenobjekten. Beherrscht man daher die Qualität von Linien, -netzen, -scharen, reliefbezogenen Kurven und Kanten usf., so ist schon viel, wenn auch noch nicht alles getan. Linien- und Flächenobjekte haben - generalisierend gesagt - geometrische, topologische und semantische Eigenschaften. Diese bestimmen, oft in ihren wechselseitigen Beziehungen, die (nicht notwendig statistische) relative Geoinformation, welche in den Objektmodellen über die natürlichen enthalten ist. Insofern ist die Qualitätsbeurteilung von Geodaten, GIS-Objekten und Folgeprodukten aller Art ein ganzheitliches Problem mit vielen Facetten. Als unbedingt erforderliche Qualitätsmerkmale von Geodaten werden beispielsweise von CASPARY (1993) Aussagen über die *Herkunft der Daten*, ihre *logische Konsistenz*, *Positionsgenauigkeit*, *Vollständigkeit*, *Attributgenauigkeit* und *Aktualität* angesehen.

Genauigkeitsbetrachtungen für topographische Karten (TK) haben eine lange Tradition; man denke beispielsweise an die KOPPEsche Formel zur Beurteilung der Höhengenauigkeit von Schichtenplänen oder an die der Qualitätssicherung von TK dienenden Musterblätter. Für die unterschiedlich strukturierten GIS-Objekte, ihrer Basisdaten und der Folgeprodukte, bemüht man sich seit etwa einem Jahrzehnt, allgemein anerkannte, mit DIN- und ISO-Normen konsistente Richtlinien auszuarbeiten. Die aktuellen, interdisziplinär zwischen Geodäsie, Photogrammetrie, Bildverarbeitung und Geoinformatik organisierten Forschungsarbeiten konzentrieren sich auf die o. a. Qualitätsmerkmale ebenso wie auf ihre Kombination in Qualitätsmodellen. Sie reichen von der Modellierung, der Geometrie und der Fehlerfortpflanzung (STANEK 1994, KRAUS und KRAGER 1994, BETHGE 1997) über die Kombination hybrider Daten (ILLERT 1995), ihre Konsistenz/Integrität (PLÜMER 1996, PLÜMER und GRÖGER 1997), die Visualisierung von Qualitätskenngrößen (KRAUS und HAUSTEINER 1993) bis hin zu Qualitätskontrolle/Qualitäts-

management und die Nutzeranforderungen am Geodatenmarkt (JOOS et al. 1997, SCHILCHER 1997, CASPARY und JOOS 1998).

Die Zusammenstellung der zur Qualitätsbeurteilung erforderlichen Merkmale läßt die Positionsgenauigkeit als rein geometrische Größe als ein zentrales Schlüsselmerkmal erscheinen. Einerseits wird sie von der Datenherkunft bestimmt, die im Zusammenhang mit der Vollständigkeit Aussagen über die geometrische Auflösung beim Übergang von der Realität zum Modell zuläßt, andererseits hat die Positionsgenauigkeit entscheidenden Einfluß auf Attributgenauigkeit und logische Konsistenz. So führen ungenaue Grenzen von Gebieten mit bestimmten Eigenschaften zu fehlerhafter Zuordnung von Attributen und es können topologische Beziehungen verletzt werden (BETHGE 1997, S. 9).

Die hier als zentral herausgestellte Positionsgenauigkeit ist traditionsgemäß eine Domäne der Geodäsie. In Tabelle 2 sind dazu Genauigkeitsmaße für ebene Objekte zusammengestellt. Der Punktfehler tritt nicht nur an punktförmigen Objekten auf, sondern wird auch zur Konstruktion der sog. Fehlerbänder für Linienobjekte benutzt (CASPARY und SCHEURING 1992, SCHEURING 1995). Darunter versteht man ebene Bereiche, welche die Linienlage im Sinne von Vertrauensbereichen überdecken. Da Linien digital durch geordnete Punktfolgen ersetzt sind, können die Fehlerbänder aus der Einhüllenden der Punktfehler entlang der Linie analytisch berechnet werden (DUTTON 1992, BETHGE 1997). Die Situation entlang von Linien kann fernerhin mit den zweiten Momenten der Fehler längs und quer zu ihrer Richtung beschrieben werden. Folgerichtig treten dann im Zuge der allgemeinen Fehlerfortpflanzung und in den Fehlerformeln für abgeleitete (geometrische) Größen Längs- und Querkorrelationsfunktionen auf (Beispiele in Tabelle 3). Diese rein entfernungsabhängigen Funktionen, verknüpft in der berühmten TAYLOR-KARMAN-Beziehung aus der statistischen Theorie der Turbulenz, wurden von GRAFAREND (Phys. Earth Planet. Int. 1976) für die Geodäsie erschlossen. Schließlich sind die Fehlerformeln in Tabelle 3 in Termen des Punktfehlers notiert. Mithin schließt sich der Kreis bzw. wir können vom Ausgang auf den Eingang rückkoppeln: *der Punktfehler erweist sich als ein Elementarbaustein beim Aufbau von Qualitätsmodellen für raumbezogene Geoinformationen.*

Literatur

A Einschlägige Arbeiten von E. Grafarend
(mit Numerierung nach dem autorisierten Schriftenverzeichnis des Jubilars)

3. Allgemeiner Fehlertensor bei a priori und a posteriori Korrelationen. ZfV 92 (1967) 157-165.
4. Richtungsabhängiges tensorielles Fehlerfortpflanzungsgesetz. AVN 74 (1967) 300-337.
6. Fehlertheoretische Untersuchungen und chronometrische Meßverfahren beim Einsatz von Aufsatzkreiseln in Kombination mit dem elektrooptischen Entfernungsmeßgerät AGA - Geodimeter 4 B. Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften C 112, München 1967.
7. Die Genauigkeit eines Punktes im Raum. ZfV 93 (19968) 107-111.
8. Fehlertheoretisch günstigstes Vorwärtseinschneiden. ZfV 93 (1968) 414-419.
11. Die Methode der minimalen verallgemeinerten Varianz als mehrdimensionales Analogon der Methode der minimalen Varianz. ZfV 94 (1969) 223-227.
12. Helmertsche Fußpunktkurve oder Mohrscher Kreis? AVN 76 (1969) 239-240.
15. Die Genauigkeit eines Punktes im mehrdimensionalen Euklidischen Raum. Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften C 153, München 1970.

16. Zur Wahrscheinlichkeitstheorie des Helmertschen Punktfehlers. Österr. ZfV 58 (1970) 33-41.
17. Fehlertheoretische Maxwell-Boltzmann-Verteilung. Bull. Godsique 95 (1970) 41-49.
18. Prädiktion und Punktmaße. AVN 77 (1970) 17-22.
26. Mittlere Punktfehler und Vorwärtseinschneiden ZfV 96 (1971) 41-54.
39. Genauigkeitsmaße geodätischer Netze. Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften A 73, München 1972.
58. Geodetic stochastic processes. Methoden und Verfahren der mathematischen Physik, Bd. 14 (eds. B. Brosowski/E. Martensen), BI-Verlag, Mathematical Geodesy, Mannheim 1975.
62. Geodetic applications of stochastic processes, Communication from the Geodetic Institute, Uppsala University No. 13, Uppsala 1975, Physics of the Earth and Planetary Interiors 21 (1976) 151-179.

B Andere im Text zitierte Arbeiten

BETHGE, F. (1997): Genauigkeit geometrischer Größen aus Vektordaten. DGK, C 473, München.

CASPARY, W; SCHEURING, R. (1992): Error-Bands as Measures of Geometrical Accuracy. EGIS '92, Vol. 1, Utrecht, 226-233.

CASPARY, W. (1993): Qualitätsmerkmale von Geodaten. ZfV 118, 444-450.

CASPARY, W.; JOOS, G. (1998): Quality Criteria and Control for GIS Databases. Proc. of the IAG SC 4 Symp., Eisenstadt, Austria, 436-441.

DUTTON, G. (1992): Handling Positional Uncertainty in Spatial Databases. Proc. 5th SDH, Charlston, 460-469.

ILLERT, A. (1995): Aspekte der Zusammenführung digitaler Datensätze unterschiedlicher Quellen. Nachrichten aus dem Karten- und Vermessungswesen, R. I, H. 103, 105-116, Frankfurt a. M.

JOOS, G.; BALTZER, U.; KULLMANN, K.-H. (1997): Qualitätsmanagement beim Aufbau einer topographischen Grunddatenbank am Beispiel von ATKIS in Hessen. ZfV 122, 149-159.

KRAUS, K.; HAUSTEINER, K. (1993): Visualisierung der Genauigkeit geometrischer Daten. Geo- Informations-Systeme 6, 7-12.

KRAUS, K.; KRAGER, K. (1994): Accuracy of Derived Data in a Geographical Information System. Comput., Environ. and Urban Systems, 18, 87-94.

LEHMANN, R. (1994): Zur Bestimmung des Erdschwerefeldes unter Verwendung des Maximum-Entropie-Prinzipes. DGK, C 425, München.

MEIER, S. (1990): Informationsorientierte Filterung ebener Kurven. Österr. ZfV 78, 200-211.

MEIER, S. (1991): Informationsgrößen für generalisierte Punktfelder. Geowiss. Mitt. TU Wien, H. 39, 99-106.

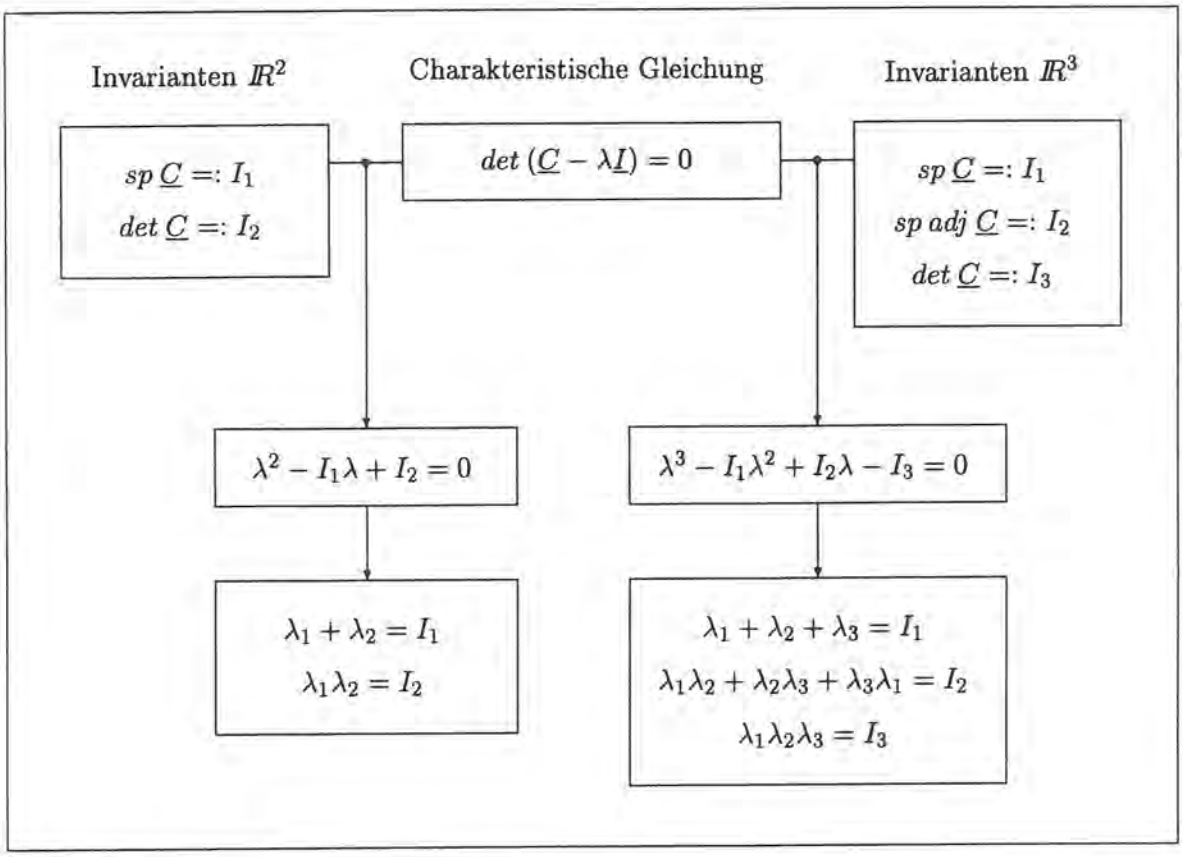
PLÜMER, L. (1996): Zur Überprüfung der Konsistenz von Geometrie und Topologie in Landkarten. Nachrichten aus dem Karten- und Vermessungswesen, Reihe I, Nr. 115, Frankfurt/M.

PLÜMER, L.; GRÖGER, G. (1997): Achieving Integrity in Geographic Information Systems - Maps and Nested Maps. Geo Informatica 1, 345-367.

SCHEURING, R. (1995): Zur Qualität von Basisdaten von Landinformationssystemen. Schriftenreihe des Studiengangs Vermessungswesen, Universität der Bundeswehr München, H. 49.

SCHILCHER, M. (1997): Qualität der Geodaten - Anspruch und Wirklichkeit des Geodatenmarktes. Fachtagung GIS '97, Inst. of Internat. Res., Wiesbaden.

STANEK, H. (1994): Datenqualität. Modellierung in GIS. Österr. ZfV u. Geoinf., 82, 14-20.



Schema 1: Beziehungen zwischen den Eigenwerten λ_i (reell, positiv) der Kovarianzmatrix $\underline{C}$ der Koordinatenfehler und den Invarianten I_i von $\underline{C}$, aufgefaßt als symmetrischer Tensor 2. Stufe (nach GRAFAREND, 1967 bis 1971). $\underline{I}$ bezeichnet die Einheitsmatrix.

σ_p^2 $\mathbb{R}^2$	Sonderfall $\underline{C} = \sigma^2\underline{I}$	Punktfehler nach	σ_p^2 $\mathbb{R}^3$	Sonderfall $\underline{C} = \sigma^2\underline{I}$
$I_1 = \lambda_1 + \lambda_2$	$2\sigma^2$	HELMERT (1868)	$I_1 = \lambda_1 + \lambda_2 + \lambda_3$	$3\sigma^2$
$\frac{I_1}{2} = \frac{\lambda_1+\lambda_2}{2}$	σ^2	MÖHLE (1936)		
		REISSMANN (1957)	$\frac{I_1}{3} = \frac{\lambda_1+\lambda_2+\lambda_3}{3}$	σ^2
$\sqrt{I_2} = \sqrt{\lambda_1\lambda_2}$	σ^2	WERKMEISTER (1920)	$\sqrt{I_2} =$	
		WILKS (1932)	$\sqrt{\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1}$	$\sqrt{3}\sigma^2$
		GRAFAREND (1970)	$\sqrt[3]{I_3} = \sqrt[3]{\lambda_1\lambda_2\lambda_3}$	σ^2

Tabelle 1: Punktfehlerquadrate σ_p^2 in Beziehung zu den Invarianten I_i und Eigenwerten λ_i in Schema 1 (nach GRAFAREND, 1967 bis 1971). $\underline{I}$ bezeichnet die Einheitsmatrix.

Objektklasse	Merkmal	Genauigkeitsmaß	Bemerkungen
Objekte mit analytisch beschreibbarer Gestalt			
punktförmige Objekte	Lagegenauigkeit	Punktfehler	mittels geodätischer Vermessung in der realen Welt realisierbar
polygonale Objekte	Linienlage Linienlänge Linienrichtung	Fehlerband Längenfehler Richtungsfehler	mittels Fehlerfortpflanzung aus Punktfehlern ableitbar
geschlossene Polygone	Flächeninhalt	Flächenfehler	
Objekte mit regelloser Gestalt			
unregelmäßig gekrümmte ebene Kurven	Linienlage Linienlänge Krümmung	Lagefehler, Fehlerband Längenfehler Krümmungsfehler	Einfluß von Diskretisierung und Erfassungsfehlern, systematische und zufällige Effekte
geschlossene ebene Kurven	Flächeninhalt	Flächenfehler	

Tabelle 2: Genauigkeitsmaße für ebene Objekte (nach BETHGE, 1997).

Punktlagefehler nach HELMERT entlang einer Polygonseite der Länge s	$\sigma_p^2(t) = \sigma_p^2(0)[(\frac{s-t}{s})^2 + (\frac{t}{s})^2]$ $t = 0, t = s : \sigma_p^2(0)$ (Max.) $t = s/2 : \sigma_p^2(0)/2$ (Min.)	Koordinaten- und Punktfehler $\sigma_x^2 = \sigma_y^2 =: \sigma^2$ $\sigma_p^2(0) = \sigma_p^2(s) =: 2\sigma^2$
Mittl. Fehler der Polygonseitenlänge	$\sigma_s^2 = 2[F(0) - G(s)]$ $F(0) = \sigma^2, G(s) = 0 : \sigma_s^2 = \sigma_p^2(0)$	G, F ... Längs-, Querkorrelationsfunktion vollständige Isotropie
Mittl. Fehler der Polygonseiten-richtung	$\sigma_r^2 = 2[F(0) - F(s)]/s^2$ $F(0) = \sigma^2, F(s) = 0 : \sigma_r^2 = \sigma_p^2(0)/s^2$	vollständige Isotropie
Mittl. Fehler des Flächeninhalts eines Polygons	$\sigma_A^2 = \frac{\sigma_p^2(0)}{4} \sum_{i=1}^{n} s_i^2$	n Seiten Gaußsche Flächenformeln vollständige Isotropie

Tabelle 3: Fehlerquadrate geometrischer Größen an polygonalen Objekten (nach BETHGE, 1997).

Hydrography and its Education

Sjamsir Mira

1. Introduction

No one will deny if we say that the sea or ocean part of the earth is going more and more important in the next future.
Talking about sea or ocean we have to talk about hydrography.

According to the United Nations Group of Experts on Hydrographic Surveying and Nautical Charting, formed by the United Nations at the Second United Nations Regional Cartographic Conference for the Americas, in Mexico September 1979, has been expressed that :

> Hydrography may be defined as the science of measuring and depicting those parameters that are necessary to describe the precise nature and configuration of the seabed, its geographical relationship to the landmass, and the characteristics and dynamics of the sea.

From the history of hydrography we know that bathymetry map as a result of hydrographic surveying mostly use for navigation, and according to the development of science and technology the use of hydrography not more only for navigation but it is used also very important for the exploration and exploitation of marine resources, and other activities.

2. Relation between Hydrography and Geodesy/Surveying, Education and Accreditation.

According to the classical definition :

- Geodesy, which mean "I divide the earth", is a science of measurement and mapping of the earth's surface.

G.L. Hosmer in his book "Geodesy", defined geodesy as follow :

- Geodesy is a science which treat of investigations of form and dimensions of the earth's surface.

If we see the definition of hydrography from U.N. Expert and compare to the definition of geodesy above, we come to the conclusion that hydrography is a part of geodesy.

If we compare curriculum of hydrographic and geodesy/surveying educations in many countries we can see so many overlapping in both curriculums. David Wells figured nicely the relation between Hydrography and another scientific fields, which mainly lectured in geodetic education (See Attachment 1). At least 60 % of the hydrographic and geodesy/surveying curriculum are mostly the same. That's why in some countries hydrographic education can be attached to the geodesy/surveying education, at least as specialisation in geodesy/surveying.
Talking about hydrography education, although there are more than 100 coastal states on the world, but less than 15 countries which already have education in hydrography and not all of them have university level.

To maintain the quality of hydrography globally, FIG and IHO formed a FIG/IHO International Advisory Board on Hydrographic Education. The Board has 8 (eight) members, which consist of 4 (four) members from FIG and 4 (four) members from IHO.
The Board meets once a year to evaluate and acreditate hydrographic education from all over the world. University Hydrographic Education can be acreditated as Academic Acreditation in Category-A.

3. Hydrographic Moduls

There are three main modules in hydrography, namely positioning, sounding and tide.
Talking about positioning one already touch the field of geodesy, because we have to talk about reference system, datum, projection, cartography, e.t.c. Modern satellite positioning system such as GPS (Global Positioning System) is widely use for positioning in hydrographic survey. As we already know if height determination using GPS, we have to take into account the undulation as the height from GPS observation refer to ellipsoid.

The objective of sounding work is to measure the depth of sea bed and operationally refer to the sea level. The problem is that sea level change continuously from time to time. That's why we have to look for a reference surface where all measuring depth to reduce to. Practically all measuring depth must be reduced to a common surface, normally Mean Sea Level (MSL); that's why we have to observe the position of sea level continuously (tidal observation). There are several methods to make tidal observations, from the simple one to the modern methods. There are two popular types of instrument normally used in practice for the tide observation, namely floating and pressure types of instrument. In practice one use MSL as geoid but we all know that MSL is not geoid. Geoid is an equipotential surface while MSL not. As MSL is not an equipotential surface, Lisitzin introduced "Sea Slope" terminology to express that two points on the MSL are located not in one equipotential surface. The distance between MSL and geoid is normally called as Sea Surface Topography (SST).

4. Geodetic Aspects on the Law of the Sea (GALOS)

In 1982 the United Nations Convention on the Law of the Sea, known as UNCLOS III, was signed in Montego Bay, Jamaica and enter into force in 1994. Actually the activity of the United Nations to produce the Law of the Sea started already in 1973; that's mean it's need 9 years to realise the UNCLOS III. In the UNCLOS III sea part of a coastal states is divided mainly into Territorial Sea (12 miles) and Economic Exclusive Zone (EEZ) (200 miles). According to the UNCOLS III, there are many countries has big sea area belong to them. In the book of "Ordering the Ocean" (See Attachment 2) there are fourteen lucky countries with the largest 200 miles zone.

Why the United Nations forced to produce the convention ?

In literature we can read that the sea area of the earth is very rich in natural resources. As technology is develop rapidly that make it possible to explore natural resources even in the offshore. As a consequence the competitive to claim sovereignty on the sea area will be very serious. That's why to order the sea/ocean is very urgent. W. Friedman in his book "The Future of the Ocean" tried to divide the ocean.

As a response to the application of UNCLOS III, in 1990 International Hydrographic Organisation (IHO) produced "A Manual on Technical Aspects of the United Nations Convention on the Law of the Sea - 1982". After he read the Manual, Petr Vanicek arranged a discussion with some geodesists who attended a conference in Miami (USA) in 1990. As the conclusion in the meeting, a geodetic conference to discuss geodetic aspects on UNCLOS III was recommended to be held in Bali 1992, called "First International Geodetic Aspects on the Law of the Sea (GALOS) Conference". Four year later (1996) the second GALOS Conference has been held in Bali again.

5. Hydrography as a science

As already mentioned above hydrographic activity shows improvement every year, not only in the routine work done by Hydrographic Office in each countries, but also in the engineering works, especially because of high technology development for mining activity in the offshore.

The existence of FIG/IHO Advisory Board for Hydrographic Education, makes sure that quality of hydrographic works done by the surveyors produced by the accredited education institutions fits the international standard. There are not so many educations involved in producing man power for hydrographic works. That's why most of the hydrographic institutions such as Hydrographic Office in each countries established their education and training by themselves to fulfil their need in hydrographic surveyors.

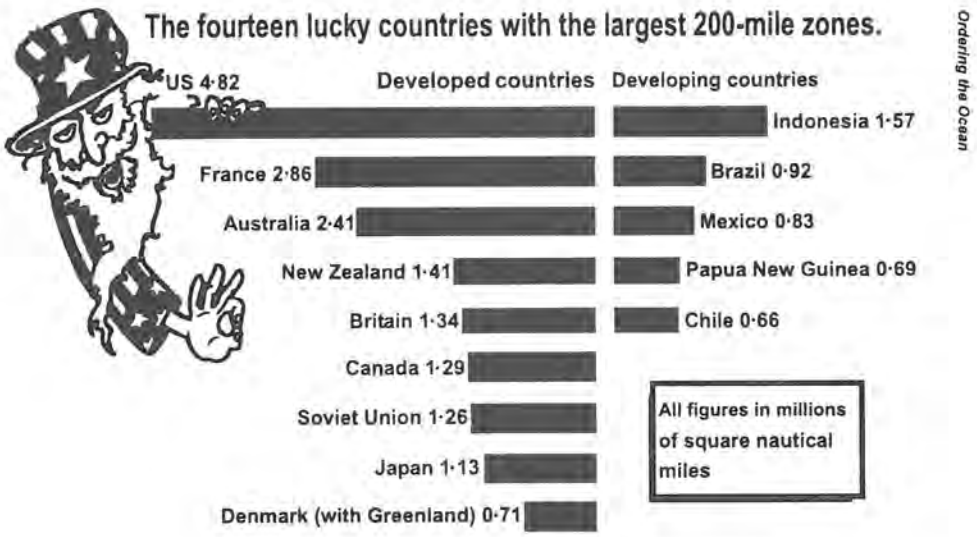
There are people who still see hydrography as only applied science which can be studied in the form of short course and training, while hydrography actually contains many scientific aspects and problems to be studied and developed. Should we let hydrography only as a normal engineering work ? Don't we want to see hydrographic scientists create new methods and technology in the field of hydrography ? To response that matter, higher education institutions should open hydrographic study program in all level of educations such as undergraduate and postgraduate programs. Actually there are already some universities offer hydrographic programs but the amount are not enough to meet a demand.

It's necessary to change the way to see and understanding hydrography by most people, not to see the side of engineering only, but it should be balanced to see hydrography as a science. We hope in the future the development of hydrography as a science will be increasing, so it can participate in many applications. In connection to this matter, maybe FIG/IHO Advisory Board for Hydrographic Education could participate in developing the idea of hydrographic higher education in University level much more than we have seen now. Universities which have already Surveying and/or Geodesy study programs, should put their attention not only to the land dimension only, but they should pay attention also to the sea. In fact, it can be the new area for them to work on.

Could we imagine universities all over the world have hydrography study programs in all levels to graduate M.Sc. and Ph.D. in the field of hydrography. To realise the idea, it's needed to increase the contacts to the universities, so the universities have better understanding about hydrography, and eventually develop hydrography education in each their universities.



RELATIONS BETWEEN HYDROGRAPHY AND ANOTHER SCIENTIFIC FIELDS
(David Wells, 1985)



The Strange Behavior of Asymptotic Series in Mathematics, Celestial Mechanics and Physical Geodesy

Helmut Moritz

In the beginning ... there was Poincaré
J.Z. Young

1 Introduction

Yes, but not quite: Heinrich Bruns (1884) (*our* Bruns!) made the race, but Poincaré (1890, 1892–1899) went much farther and deeper. He proved that "most" series used in celestial mechanics were divergent but nevertheless perfectly useful. In fact, he recognized that *mathematical* convergence or divergence may be quite irrelevant for *numerical* convergence: the series

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}\cdots$$

is convergent but numerically practically useless because the convergence is so slow. On the other hand, the series

$$\frac{e^x}{x}\left(1+\frac{1!}{x}+\frac{2!}{x^2}+\ldots\right)$$

is divergent but numerically superbly useful as we shall see in sec. 2 of this paper.
Such divergent but numerically useful series have been called *asymptotic series* by Poincaré (1892–1899, beginning of vol. 2), a terminology generally accepted by numerical mathematicians (cf. Erdélyi 1956, Press et al 1992, p. 167).
What Poincaré (1890), foreshadowed by Bruns (1884), showed was that many (or even most) series in celestial mechanics were such asymptotic "practically convergent" series.
In geodesy, we have similar series of doubtful moral behavior: the spherical–harmonic series of the geopotential at the earth's surface and Molodensky's series for the solution of the geodetic boundary value problem.
Concerning the spherical-harmonic series, every scientist aspiring to fame in physical geodesy was bound to give a wrong proof of convergence or divergence. Hopfner (1933) proved convergence; his proof was wrong. Baeschlin (1948) proved divergence; his proof was wrong. Moritz (1961) proved that the question of convergence vs. divergence was meaningless; his proof was wrong. In fact, it was basically right: an arbitrarily small change of the attracting mass (the earth) by a sand grain can change convergence into divergence. I was wrong in taking for granted that, just as complex functions in the plane have a "circle of convergence", their three–dimensional analogues, spherical harmonic functions, had a "sphere of convergence". This mistake was pointed out by a beautiful counterexample by Krarup (1969, pp. 47–49). Whereas Moritz has shown the "direct" problem, that convergence can be readily changed into divergence (the "sandgrain effect"), Krarup proved the much more difficult "inverse problem", that divergence can be changed into convergence by *an arbitrarily small change* of the geopotential.

In analogy to the correspondent theorem for the complex functions in the plane, due to C. Runge (the "Runge" from the well-known "Runge–Kutta method"), he called the three-dimensional geodetic theorem, found by him independently, with his characteristic modesty, "Runge theorem". In my book (Moritz 1980) I called it at least "Runge–Krarup theorem" because of the enormous intellectual work Krarup put in. However, the Devil, irritated by human attempts to trespass mathematically into his empire, the earth's interior, was active also here: it turned out that this theorem was known already to G. Szegö around 1925(!), cf. (Frank–Mises 1930 pp. 760–762).
Under special assumptions, convergence can of course, be proved (Moritz 1980, p. 53), Balmino (1994) and Grafarend and Engels (1994).
For details, the reader may consult (Moritz 1980, sec. 6–8) or the slightly more humorous account (Moritz 1978).
So what? The question of the convergence or divergence of spherical harmonics at the earth's surface is perfectly meaningless. Practically, we anyway operate with finite bestfitting spherical harmonic polynomials, whether to degree and order 30, 360, 1000 or 3600 (or, if you can pay for it and are very patient, 3,600.000). These are exactly bestfitting polynomials in the Runge–Szegö–Krarup sense! The author gives his blessing and hopes (probably in vain) that the Devil will be impressed enough to step off this theater to look for more profitable problems ...
Concerning the convergence of Molodensky's series it is similar: all convergence proofs known to the author (including his own in Moritz 1980, sec. 47) are probably wrong. Don't waste your time, however, to look for errors: it is also an asymptotic series and you have the pleasant alternative: either the first 3, 4 or 5 terms are sufficient, or look for another job; the following terms are anyway pure noise because the effect of measuring errors rapidly increases and wipes out the gravitational "signal".

2 A Computer Study of a Mathematical Asymptotic Series

A simple but typical example is the well-known *exponential integral* defined by

$$Ei(x) = \int_{-\infty}^{x} \frac{e^t}{t} dt \qquad (x > 0) \quad . \tag{1}$$

This is a standard mathematical function contained e.g., in the programming language MATHEMATICA; it can be called there by the name $ExpIntegralEi[x]$, cf. Fig. 1.
The standard way of computing it if MATHEMATICA is not available, is the source code in C as given by (Press et al, 1992, sec. 6.3, p. 225), function "ei(x)". Since this book is the standard work for modern mathematical computation, it is available not only in C, but also in FORTRAN, PASCAL and even BASIC. Therefore, the interested reader is simply referred to this book.
It turns out that for small x one uses an ordinary convergent power series:

$$Ei(x) = \gamma + \ln x + \frac{x}{1 \cdot 1!} + \frac{x^2}{2 \cdot 2!} + \ldots \tag{2}$$

where $\gamma = 0.5772\ldots$ is Euler's constant. For $x > 16.62\ldots$, however, the convergence of this series becomes too slow, and one uses the asymptotic series

$$Ei(x) \doteq \frac{e^x}{x}\left(1 + \frac{1!}{x} + \frac{2!}{x^2} + \ldots\right) \quad . \tag{3}$$

To compare the "true" (or rather highly accurate) function $ExpIntegralEi[x]$ in MATHEMATICA with our asymptotic series (3), one could program this asymptotic series in MATHEMATICA. Since a beautiful and fast source code for (3) is already contained in the C Program "$ei(x)$"

by Press et al., as mentioned above, we can use it too and convert it to MATHEMATICA by the auxiliary tool MATHLINK (Wolfram 1996, sec. 2.12). This is extremely simple: we add the header file "mathlink.h" in our source code "eiconv.c" (Fig. 2), and compile it (this is a bit technical) to get the exe–file "eiconv" to be installed in the MATHEMATICA program of Fig. 1 as shown there (all relevant programs may be obtained from moritz@phgg.tu-graz.ac.at).
The result is the deviation

$$err[n,x] = ExpIntegralEi[x] - expIntAsy[n,x] \tag{4}$$

of our home–made function (3) truncated after the n–th term. It is thus a function of the usual variable x and the truncation value n. The deviation is with respect to the standard MATHEMATICA function $ExpIntegralEi[x]$.

Mathematica Program ExpIntConvergence.nb

```
In[1]:= Install["eiconv"]    (* C exe-file from Mathlink *)

Out[1]= LinkObject [eiconv, 2, 2]

In[2]:= LinkPatterns[%]

Out[2]= {expIntAsy [n_Integer , x_Real ]}

In[3]:= ?expIntAsy

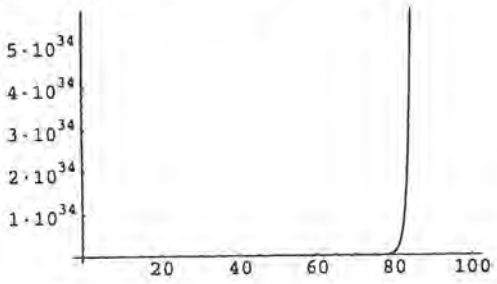
        expIntAsy [n,x] approximates the Mathematica function ExpIntegralEi [x] by the sum
          of the first n terms of an asymptotic series.

In[4]:= (* Mathematica function *)

In[5]:= ?ExpIntegralEi

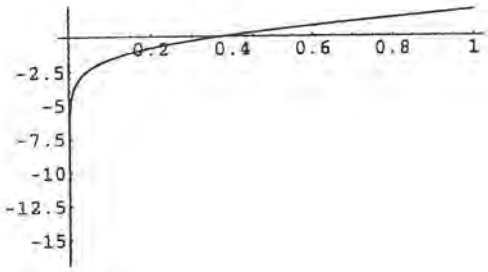
        ExpIntegralEi [z] gives the exponential integral function Ei(z).

In[6]:= Plot[ExpIntegralEi[x], {x, 0, 100}]
```




```
Out[6]= -Graphics-

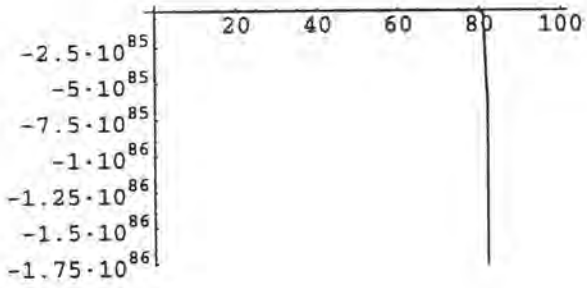
In[7]:= Plot[ExpIntegralEi[x], {x, 0, 1}]
```



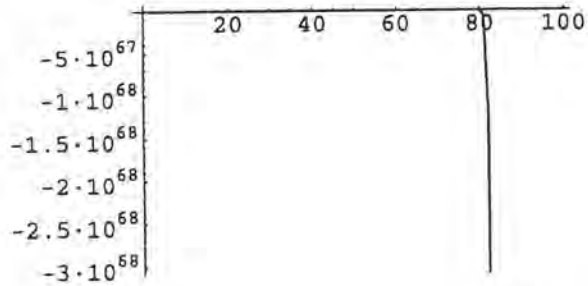

```
Out[7]= -Graphics-

In[8]:= err[n_, x_] := ExpIntegralEi[x] - expIntAsy[n, x]
```

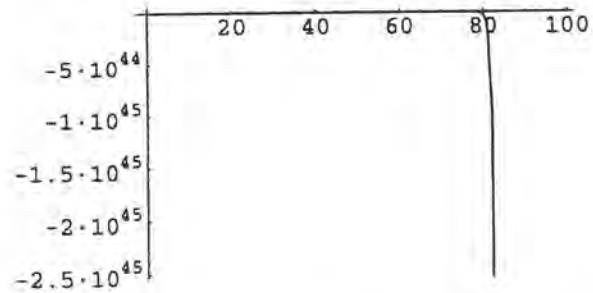
```
In[9]:= ListPlot[Table[err[n, 3.], {n, 2, 100}], PlotJoined -> True];
```




```
In[10]:= ListPlot[Table[err[n, 5.], {n, 2, 100}], PlotJoined -> True];
```




```
In[11]:= ListPlot[Table[err[n, 10.], {n, 2, 100}], PlotJoined -> True];
```




```
In[12]:= relativeErr[n_, x_] := err[n, x] / ExpIntegralEi[x]

In[13]:= {relativeErr[5, 20.], relativeErr[10, 20.], relativeErr[15, 20.]}

Out[13]= {0.0000174303 , 5.49235 ×10^-7, 9.37426 ×10^-8}

In[14]:= {relativeErr[17, 20.], relativeErr[18, 20.], relativeErr[19, 20.] }

Out[14]= {3.78099 ×10^-8, 1.4681 ×10^-8, -7.29146 ×10^-9}

In[15]:= Uninstall["eiconv"]

Out[15]= eiconv
```

```
// Source Code eiconv.c
// Study of the convergence of the asymptotic series
// for the exponential integral Ei(x) -

#include <math.h>
#include "mathlink.h"

double eias(int n, double x)  // in C : Function eias(n,x)
{                             // in MATHEMATICA : ExpIntAsy[n,x]
   int k;                     // Asymptotic Series to power n
   double prev, sum, term;

   if (x <= 0) return 0;
   else
   {
     sum=0.0;
     term=1.0;
     for(k=1;k<=n;k++)
     {
       prev=term;
       term *= k/x;
       sum += term;
     }
     return exp(x)*(1.0+sum)/x;
   }
}
int main(argc, argv)
   int argc; char* argv[];
{
   return MLMain(argc, argv);
}
```

Figure 2: Source code

To repeat, $err[n, x]$ is the deviation of the (divergent!) asymptotic series (3) to order n from the "true" function $Ei(x)$.
Of course, I expected such a behavior, but nevertheless the extremeness of the result was shocking, and I could hardly believe my eyes. Even with three terms of the asymptotic series ($n = 3$) one gets an excellent approximation and with $n = 5, 10, 15, 17, 18, 19$, we get phantastic accuracies on the order of 10^{-18}. And these we get with a few terms of a divergent series!

3 Celestial Mechanics and Geodesy

It is not surprising that Poincaré, when he first recognized these facts and applied them to the series of celestial mechanics, was overwhelmed with the joy of discovery. His work, and still less that by Bruns, was hardly understood for more than half a century. Only with the advent of fast computers was one able to render visible the phantastic pictures of modern "general nonlinear dynamics", now popularly called "chaos theory", which Poincaré had in the back of his mind and, as he said, to his regret was unable to draw.
Another fact common to Poincaré's series and spherical–harmonic series was that regular (stable) and "chaotic" (unstable) trajectories, convergent and divergent series are arbitrarily close to each other. In geodesy this is the *Runge–Szegö–Krarup (RSK) theorem* mentioned at the very beginning of this paper. (Other people could be included in this list, cf. Moritz 1980, p. 74, but questions of priority are usually rather questionable ...) In chaotic dynamics it is the *KAM*

(Kolmogorov–Arnold–Moser) theorem.
The RSK theorem, in a rather simplified form (Moritz 1980, p. 67) may be stated:

> Let K be a compact set and Γ and Ω open sets in R^3, such that their boundaries are homeomorphic to a sphere and such that $K \subset \Gamma$ and $\Gamma \subset \Omega$. If the function ϕ is harmonic in Γ and if $\epsilon > 0$ is arbitrarily small, then there exists a function ψ, harmonic in Ω, such that
>
> $$|\phi - \psi| < \epsilon \tag{5}$$
>
> uniformly on K.

The KAM theorem requires considerable knowledge from number theory. A relatively accessible presentation can be found in (Schuster 1988, p. 191). The criterion of stability or instability of trajectories is

$$\left|\frac{\omega_1}{\omega_2} - \frac{m}{s}\right| > \frac{k(\epsilon)}{s^{2.5}} \qquad (k(\epsilon \to 0) \to 0) \quad . \tag{6}$$

What does this mean? The state space of a one–dimensional oscillation (frequency ω_1) is a circle, and the state space of two regular oscillators is the topological product circle × circle, which is a torus. If the motion is perturbed, some tori will be conserved (stable motion), but some tori will break up (chaotic motion). This depends on "how irrational" the ratio ω_1/ω_2 is, how well or how poorly ω_2/ω_1 can be approximated by a rational number (m and s are integers). This is a difficult number–theoretic problem solved by eq. (6). The interested reader can work himself through the enormous literature; for the present purpose it is better to insert a picture (Fig. 3)

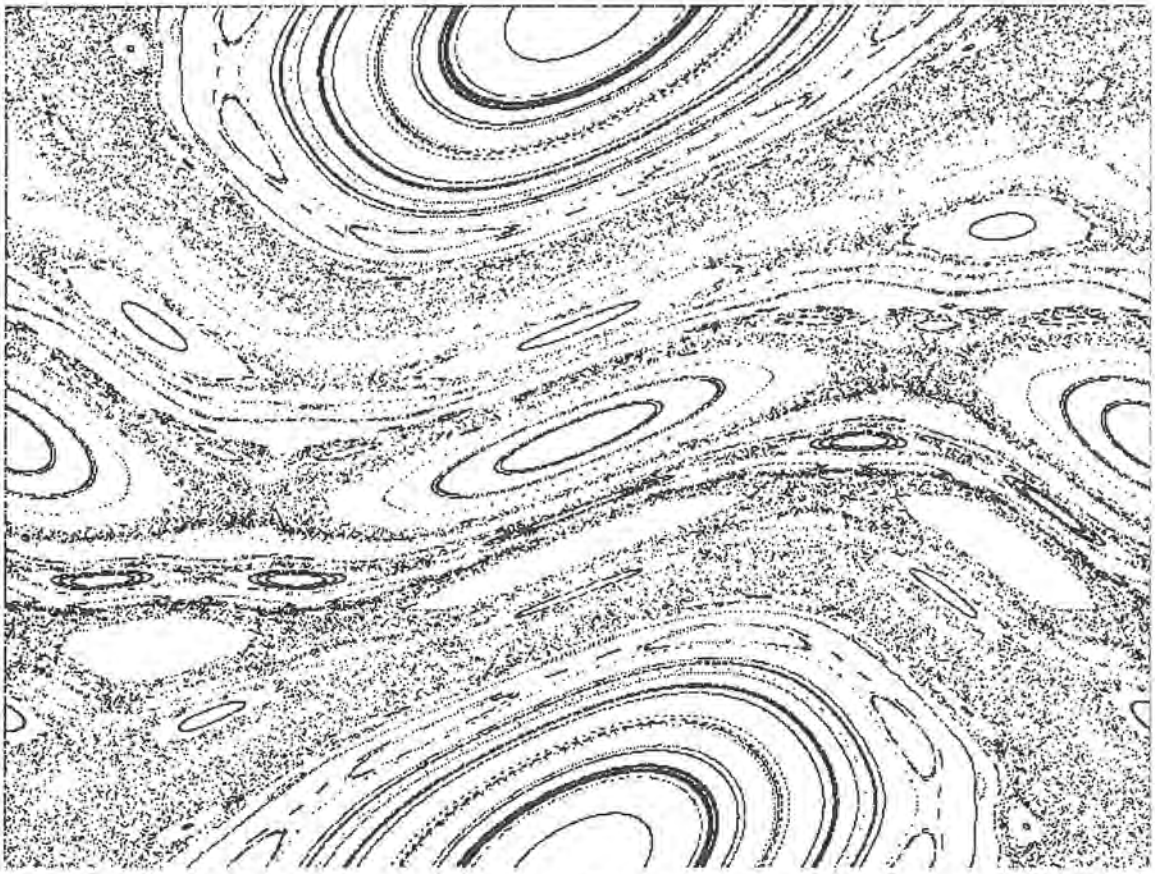


Figure 3: Poincaré

which shows the "Poincaré section" of a certain set of three–dimensional trajectories: every point corresponds to a trajectory, every "elliptic island" to a torus that has been preserved.

There are many books on chaos by computers, but to me the book (Herrmann 1994) is still unsurpassed. (Fig. 3 was computed by the author using MATHEMATICA and MATHLINK to C, but has been inspired by Herrmann's algorithms.)
From the very beginning, the problem of the present paper has fascinated me (Moritz 1969). My guess that there are relations between astronomic and geodetic series, both being asymptotic series, later proved to be correct. The work of the mathematician Heinrich Bruns both in astronomy and geodesy was striking. My early guess on Runge–type behavior of spherical harmonics (Moritz 1961) was intuitively correct and mathematically wrong, but it inspired pioneering work by Krarup (1969). When I studied the difficult but fascinating book (Sternberg 1969), I was struck by the use of "hard inverse problems" of non–linear functional analysis in the KAM problem, especially by a method of Nash which the famous mathematician Lars Hörmander later (1976) used in the first partially successful mathematical attack of existence and uniqueness of Molodensky's problem! In fact, Hörmander wsa the second eminent mathematician in this century who did significant work in geodesy. The first was Poincaré: Throughout his life (1854–1912), Henri Poincaré never ceased to work on problems of astronomy and geodesy. (In his last years, he was even the French Chief Delegate to the International Geodetic Association.)
I still I firmly believe that there is a deep mathematical relationship between nonlinear dynamic systems ("chaos theory") and geodetic problems which I, however, was never able to penetrate really deeply. We would badly need a new Poincaré. What about the young geniuses of our 60 years young Professor Erik Grafarend?

References

Balmino G (1994) Gravitational potential harmonics from the shape of a homogeneous body, Celestial Mechanics and Dynamial Astronomy, 60, 331–364.

Baeschlin C F (1948) (1949) Lehrbuch der Geodäsie, Orell–Füssli, Zürich.

Bruns H (1884) Bemerkungen zur Theorie der allgemeinen Störungen, Astron. Nachr., 109, 216–222.

Erdélyi A (1956) Asymptotic Expansions, Dover, New York.

Frank P and Mises R von (1930) Die Differentialgleichungen der Mathematik und Physik, 2nd ed., vol. 1 (reprint, Dover, New York, 1961).

Grafarend E and Engels J (1994) The convergent series expansion of the gravity field of a starshaped body, manuscripta geodaetica, 19, 18–30.

Herrmann D (1994) Algorithmen für Chaos und Fraktale, Addison–Wesley, Bonn.

Hopfner F (1933) Physikalische Geodäsie, Akadem. Verlagsgesellschaft, Leipzig.

Hörmander L (1976) The boundary problems of physical geodesy, Arch. Rat. Mech. Anal., 62, 1–52.

Krarup T (1969) A contribution to the mathematical foundation of physical geodesy, Publ. 44, Danish Geod. Inst., Copenhagen.

Moritz H (1961) Über die Konvergenz der Kugelfunktionsentwicklung für das Aussenraumpotential an der Erdoberfläche, Österr. Zeitschr. Vermessungsw., 49, 11–15.

Moritz H (1969) Convergence problems in celestial mechanics and physical geodesy, Proc. IV Symposium on Mathematical Geodesy, 81–94, Trieste.

Moritz H (1978) On the convergence of the spherical–harmonic expansion for the geopotential at the earth's surface, Boll. Geod. Sci. Affini, 37, 363–381.

Moritz H (1980) Advanced Physical Geodesy, Wichmann, Karlsruhe (2nd ed. 1989).

Poincaré H (1890) Sur le problème des trois corps et les équations de la dynamique, Acta Math., 13, 1–270.

Poincaré H (1987) Les Méthodes Nouvelles de la Mécanique Céleste, 3 vols., Albert Blanchard, Paris (reprint from the original edition from 1892–1899).

Press W H, Teukolsky S A, Vetterling W T and Flannery, B P (1992) Numerical Recipes in C, Cambridge Univ. Press.

Schuster H G (1988) Deterministic Chaos, 2nd ed., VCH, Weinheim, Germany.

Sternberg S (1969) Celestial Mechanics, Part II, W.A. Benjamin, New York.

Wolfram S (1966) The Mathematica Book, 3rd ed., Cambridge University Press.

Datum Accuracy and its Dependence on Network Geometry

Haim B. Papo

ABSTRACT

Free networks are still very popular in geodesy, surveying and mapping due mainly to their unique property of being independent of errors in external data (fixed higher order control).
The results of free net adjustment and their superior quality serve as an authentic reflection (undistorted mirror image) of the measurements' quality. The above property is particularly important today when we employ GPS measurements for densification of conventional control, where the later is usually of a relatively lower quality. Notwithstanding the optimal properties of free networks (including GPS networks as well), we examine in this paper the relationship between network geometry and the positional accuracy of its points. In particular we study the "accuracy as a function of geometry" equation as related to datum definition of a network by free net adjustment constraints. The paper presents new concepts and describes relevant parameters that are capable of defining and quantifying the datum accuracy of a free network. Following the development of conceptual and theoretical tools, the paper presents results of numerical experiments carried out with a schematic 2-D (horizontal) GPS network.
The results lead to a somewhat surprising (although intuitively acceptable) conclusion: "Positional accuracy of points in a free network depends heavily on the spatial distribution of its datum-definition points". The above dependence on geometry is in excess to the well-known and trivial property of networks where positional accuracy of a point is inversely proportional to its distance from the mass center of the datum-points' sub-net.

1. Introduction. Datum definition in free networks.

This paper deals with the initial stages in the complex process of establishing a geodetic control network. According to Grafarend (1974), at the "zero-order-design" stage the datum of the network has to be defined. In case of a densification net, as part of the hierarchical structure, datum is defined by the "higher order" control points i.e. their respective coordinates which remain fixed throughout the adjustment of the measurements. In this case there is little or no room for design variations. All the higher order points that can be accessed are used. It is wise to look for possible inconsistencies and hope, at the same time, that none would be found.

In case of an independent, "free" or "floating" network, datum is defined by the selection of a subset of its points. The selection of those datum-points is subjective and depends on a number of criteria such as, for example, the destination of the network, the specific type of measurements, the geotechnical properties of the ground, monumentation characteristics and many more. In the following sections we investigate the geometry of the datum-points subset and its possible effect on point-position accuracy in the network.

An earlier version of this paper was presented in Moscow at **International Scientific and Technical Conference** dedicated to the 220th anniversary of Moscow State University of Geodesy and Cartography, May 24-29, 1999

We should point out that in this paper the overall geometry of the network is taken for granted. There are hardly any useful analytical tools that have been developed over the years to solve satisfactorily Grafarend's (1974) "first-order-design" problem of a control network. Even if such tools were available, their practical application is in most cases problematic due to the inevitable constraints imposed by other trivial but also rather important considerations such as topography, logistics, cost, speed etc.

Let us have a network of k points in s-dimensional space (s=1,2,3). We seek an estimate of the k.s coordinates of those k points including their respective accuracies (the k.s by k.s covariance matrix of the coordinates). We note that the coordinates and their accuracies are subjective quantities as they depend heavily on the particular reference system (datum) which has been selected. The selection of a particular datum is accomplished as follows:
Out of the total of k network points, m points ($m \leq k$) are selected and declared as a datum-base subset. A datum definition weight matrix P_m is formed which elaborates the above selection by assigning to each of the m datum-points an appropriate weight. In most cases P_m is a diagonal matrix with at least d non-zero (usually positive) elements. The number d is the so-called datum defect of the free network. It depends on s (dimensionality) and on the specific type of the measurements that are being used for constructing the network. As an example: for GPS vectors "measured" in 3-D the datum defect is d=3 (origin of the reference system); for GPS vectors which have been reduced to horizontal distances (s=2) the datum defect in 2-D is d=3 (origin-2 and orientation -1 of the reference system). The approximate coordinates of the datum points are, in general, inconsistent with the correct relative positions. In spite of this inconsistency the approximate coordinates of the m datum-points together with the P_m matrix serve as a data basis for datum definition of the network [See Papo&Perelmuter, 1981].

The mathematical model of the n measurements is linearized at the approximate values of the coordinates (vector X^o) and produces the set of n observation equations:

$$L + V = A \cdot X \tag{1.1}$$

where

X is the vector (k.s) of coordinate corrections.
A is the matrix (n by k.s) of partial derivatives.
L is the vector (n) of differences between observed and approximate measurements.

The basis of null space of A [See Papo, 1987] known also as the Helmert's transformation matrix E is a k.s by d matrix with the following important property.

$$A.E = 0 \tag{1.2}$$

Datum based on the selected m subset is defined by subjecting the vector X in (1.1) to the following linear constraints [See Papo&Perelmuter, 1981; Papo, 1989].

$$E^T P_m X = 0 = E_m^T X \tag{1.3}$$

where

$$E_m = P_m \cdot E$$

We will see later that (1.3) is derived from seeking a minimum of the following quadratic form:

$$X^T P_m X = \min. \tag{1.4}$$

But first we minimize the V^TPV quadratic from in order to create the set of normal equations:

$$NX = U \tag{1.5}$$

The above set is singular due to the need for datum definition. It can be solved only if we apply on X at least d independent linear constraints. If we apply (1.3), for example, to (1.5) we will obtain a particular solution denoted as X_m. There is an infinite number of sets of d constraints which when applied to (1.5) will produce that many different solution vectors X, all of which satisfy the normal equation (1.5).

Let us consider one particular branch of such solutions, which is characterized by the contents of the P_m weight matrix. Out of the multitude of P_m matrices (and respective datums) there is one particular solution which is unique in that the P_m matrix is equal to the identity matrix. We will denote that unique solution as X_k and its datum - the k-datum. As can be seen, in this case m=k where each of the k points in the network contributes to datum definition on an even basis as P_k= I - the identity matrix of size k.s by k.s. Meissl in (1969) and also Koch in (1987) have shown that the X_k solution compared to all the rest is optimal and unique in the sense that the trace of the k.s by k.s covariance matrix is a minimum. In addition, it can be shown [See Koch, 1987] that Q_k, the unscalled (σ_o^2) covariance matrix of X_k is the pseudo-inverse of N, known also as the Moore-Penrose inverse of N ($Q_k = N^+$). According to Meissl (1969) the above optimal X_k solution is the only one which reflects faithfully the accuracy of the measurements where all the other (external) sources of error have been effectively filtered out.

According to Wolf (1977) any two X vectors, which pertain to two different datums - X_m and X_k, for example - are related through the following simple equation:

$$X_m = X_k + E\, t_{km} \tag{1.6}$$

where t_{km} is a d vector containing transformation parameters from the k - to the m - datums.
From (1.6) under the minimum condition (1.4) we derive a solution for t_{km} as follows [See Wolf, 1977 and Papo, 1987]:

$$t_{km} = -\left(E_m^T E\right)^{-1} E_m^T X_k = -H_m X_k \tag{1.7}$$

which when substituted in (1.6) results in yet another form of (1.6)

$$X_m = (I - EH_m)\, X_k = R_m\, X_k \tag{1.6'}$$

Equation (1.6') when multiplied from the left by E_m^T results in equation (1.3) - the d linear constraints which define the specific m - datum. Equations (1.7), (1.6') and the covariance matrices of t_{km} and X_m will be investigated in the following two sections to reveal the dependence of point-accuracy on datum definition in a free network.

The unscalled covariance matrices of X_m and t_{km} are evaluated as follows:

$$Q_m = R_m\, Q_k\, R_m^T \tag{1.8}$$

$$Q_{tm} = H_m\, Q_k\, H_m^T \tag{1.9}$$

where as mentioned above $Q_k = N^+$ is the pseudoinverse of N.

We note [See Meissl, 1969] that the R_m matrix is idempotent which means that:

$$Q_m = R_m\, Q_m\, R_m^T \tag{1.8'}$$

Another interesting property of the R_m-type matrices is that the result of a chain product of different R_m matrices is equal to the leftmost R_m matrix:

$$R_{mi} = R_{mi} \cdot R_{mj} \ldots . R_{mz} \tag{1.10}$$

2. The datum accuracy concept.

The relative positions of network points in an adjusted free network under minimal (d) constraints are invariant and form a perfectly rigid structure [See Meissl, 1969, Koch 1987]. That means that no matter which particular datum is selected for the free network, point-coordinates will indeed change from one selection to the next, but not the adjusted distances or spatial angles between the points. In contrast to the above invariance, point - accuracies (the covariance matrix of point coordinates) depend heavily on the specific datum that was selected to display point positions in terms of their respective coordinates.

It seems reasonable to adopt Meissl's approach in his Annex F in (1969) and to declare the k-datum (P_k= I) (based evenly on all the points in the network) as the best of all possible datums and use it as a "starting point" to arrive at any specific datum. The supremacy of the k-datum is due mainly to its minimum trace property and also to its uniqueness. In spite of its optimal and unique properties, however, the application of the k-datum (P_k= I) in practice is a complete nonsense. It is hard to find even a single case in which we would like to base the datum of a network on the approximate coordinates of all of its points.

The line of reasoning, which was adopted in this paper, is to study the effect of datum transformations from the unique but meaningless k-datum to any particular m-datum (P_m). In particular we shall pay attention to the geometry of the m-datum points and to its effect on positional accuracy of the network points. Here we have also a unique opportunity to "estimate" the nonestimable datum parameters of a network including the respective covariance matrix. As a matter of fact we will estimate the datum-transformation parameters t_{km}: from the unique but meaningless k-datum to any specific (P_m) m-datum. That means that we will analyze the Q_{tm} matrix as a function of the geometry of the m datum points.

"Geometry" is understood here as the size and distribution (within the network) of the m-subset of datum points. From elementary combinatorial analysis we can have the discrete number ${}_kC_m$ of different m-sets taken from a population of k points:

$${}_kC_m = \frac{k!}{m!(k-m)!} \tag{2.1}$$

There are cases, mainly in deformation analysis, where we are interested is investigating the consistency in the behavior (motion, rotation and strain) of a specific g-subset (a block of g points) of the network. The displacement field (or the velocity field) of the g-subset can be partitioned into rigid motion of the block and homogeneous strain (global components) and residual displacements of the g points (individual components). In such a case it would be important to reveal the connection between geometry of the datum (the m-datum) and the covariance matrix of the global parameters (rigid motion and homogeneous strain) of the g-subset. For simplicity we make the following reasonable assumptions:

(a) there have been only two observational sessions at epochs t_0 and t_1;
(b) in both sessions datum has been defined by identical m-subsets (identical approximate coordinates and identical P_m matrices);

(c) in processing the second session the approximate coordinates of the g-subset points have been set equal to the results of the adjustment of the first session.

Under the above three assumptions the estimated coordinate corrections X as a result of the adjustment of the second session would represent the displacement field of the g-subset. Partitioning of the displacement field is done as follows [See Meissl, 1969 and Papo, 1989]:

$$X_g = V_g + E_g\, t_g \tag{2.2}$$

where V_g is the vector of residual displacements;
t_g is a vector of rigid motion and homogeneous strain parameters and E_g is the extended Helmert matrix [See Papo, 1986].
The t_g parameters are estimated so as to minimize the following quadratic from:

$$V_g^T\, Q_{mg}^{-1}\, V_g = \min \tag{2.3}$$

where Q_{mg} is the covariance matrix of the g-subset coordinates (m-datum). Analysis similar to the one shown in the previous section results in:

$$t_g = \left(E_g^T\, Q_{mg}^{-1}\, E_g\right)^{-1} E_g^T\, Q_{mg}^{-1} \cdot X_g \tag{2.4}$$

$$Q_{tg} = \left(E_g^T\, Q_{mg}^{-1}\, E_g\right)^{-1}$$

An example of an E_g matrix in 3-D is shown in the following table [See Papo, 1986]:

rigid motion						homogeneous strain					
translation			rotation			scale variation			orthogonality		
1	0	0	0	z1	-y1	-x1	0	0	0	-z1	-y1
0	1	0	-z1	0	x1	0	-y1	0	-z1	0	-x1
0	0	1	y1	-x1	0	0	0	-z1	-y1	-x1	0
1	0	0	0	z2	-y2	-x2	0	0	0	-z2	-y2
...	...	...	...	...	...	...	...	...	...	...	...
0	0	1	yg	-xg	0	0	0	-zg	-yg	-xg	0

Note that xi, yi, zi (i=1,2....g) are the approximate coordinates of the points in the g-subset.

Thus we study the effects of geometries of two subsets in the network: the (datum) m-subset and the (deformation block) g-subset. The contents of the following covariance matrices, regarded as quantifiers of accuracy, will be studied:

Q_{tm} - m-datum-parameter accuracy (t_m)
Q_m - m-datum network point-position accuracy (X)
Q_{tmg} - g-block global parameters accuracy (t_{mg})

All three matrices pertain to the same m-datum and respective P_m matrix.

3. Experiments with datum and block geometries

In order to demonstrate the utility of the proposed method of analysis we designed a schematic GPS network in 2-D, consisting of 19 points (See fig. 3.1). At this time (1999) and in spite of their undisputed 3-D capacity, GPS measurements are still much more effective in constructing a horizontal (2-D) rather than a spatial (3-D) network. Mainly for this reason we have chosen to postpone, for the time being, experiments with datum geometries in 3-D.
As shown in figure 3.2 the points in the network are located at distances of exactly 20 km. from each other and form together a perfectly symmetrical 60° grid. The simulated vectors are of two typical lengths: 20.00 km. vectors such as 1-2, 5-9, etc. and 34.64 km. vectors such as 4-10, 15-17 etc. The simulated GPS vectors were adjusted in two different modes:
mode I: where a vector is represented by components: Δn (north-south) and Δe (east-west);
mode II: where each vector is regarded as a single horizontal distance Δs. ($\Delta s^2=(\Delta n)^2+(\Delta e)^2$).

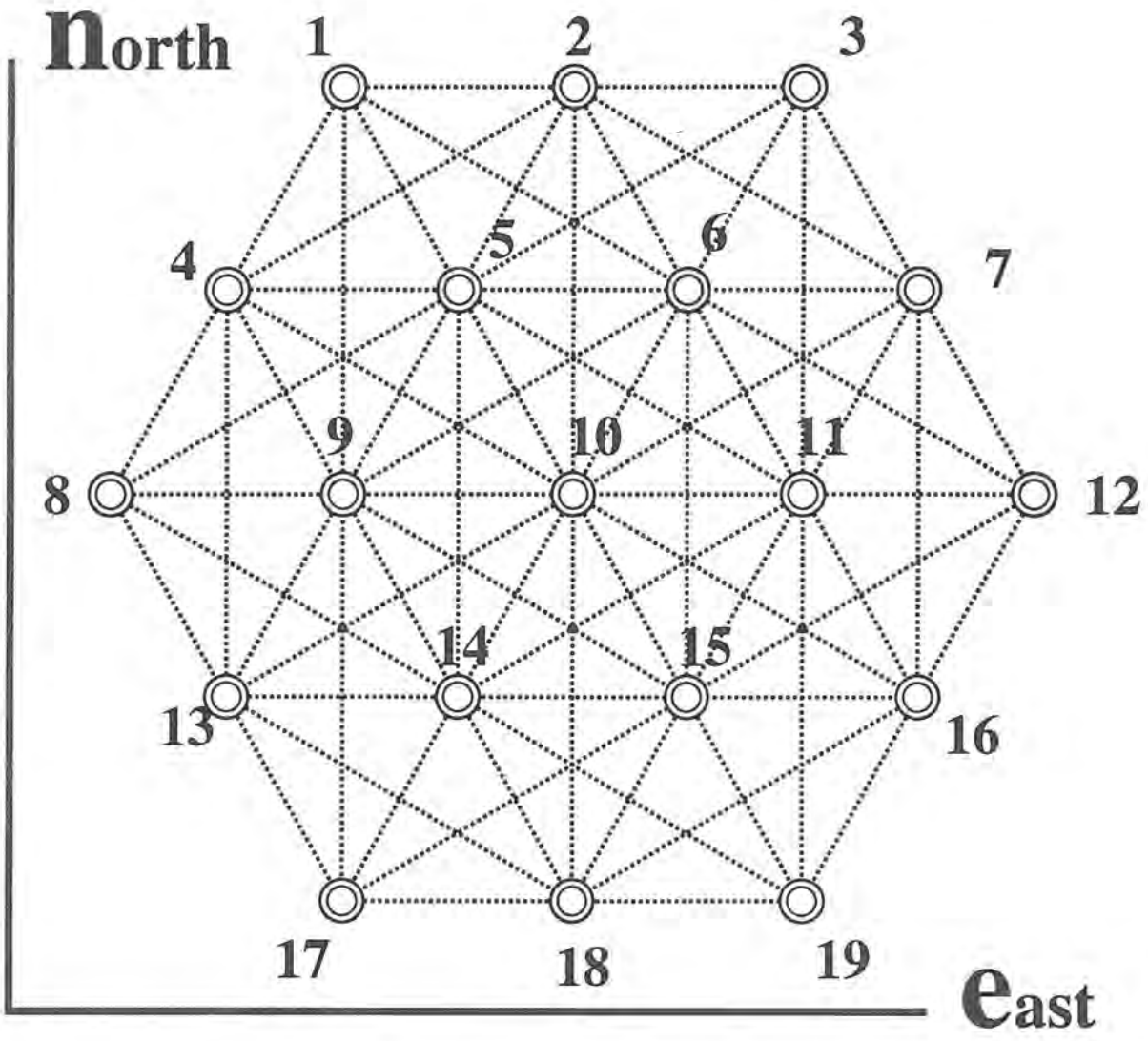


Figure 3.1: Experiments in 2-D geometry

A total of 42 (3x14) vectors of 20 km. and 30 (3x10) vectors of 34.64 km. were simulated with medium accuracy of about 1 ppm. In mode I a correlation of 70% was assumed between the two vector components. Zero correlation was assumed between any two different vectors. Weight matrices and partial derivatives were evaluated for each of the two measurement modes and finally two normal

matrices were formed - one for mode I and one for mode II, each measuring 38 by 38. The datum defect for N_I is d_I=2 due to the need to define origin in "n" and in "e". Note that orientation and scale are provided in mode I by the Δn and Δe measurements. The datum defect for N_{II} is d_{II}=3 due to the additional need to define orientation. Thus three Helmert datum transformation matrices are needed for the analysis as shown bellow. The values of ni and ei, in E_{II} and in E_g are the approximate coordinates of the respective network points. As for E_g , in this paper we limited the investigation of strain to homogeneous scale variation only. So the E_g matrix implies only four degrees of freedom as needed for partitioning the displacement field of a block in 2-D (See equation 2.2).

$$E_I^T = \begin{bmatrix} 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & \dots \end{bmatrix} \quad ; \quad E_{II}^T = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & 1 & 0 & \dots \\ -e1 & n1 & -e2 & n2 & -e3 & \dots \end{bmatrix} \begin{matrix} \text{origin for "n"} \\ \text{origin for "e"} \\ \text{orientation} \end{matrix}$$

$$E_g^T = \begin{bmatrix} 1 & 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & 1 & \dots \\ -e1 & n1 & -e2 & n2 & \dots \\ -n1 & -e1 & -n2 & -e2 & \dots \end{bmatrix} \begin{matrix} \text{translation "n"} \\ \text{translation "e"} \\ \text{rotation of axes} \\ \text{variation of scale} \end{matrix}$$

Two datum configurations (geometries) were defined as follows (See fig. 3.2):

A: a "wide" datum consisting of points 1, 8, 9, 17;
B: a "narrow" datum consisting of points 4, 8, 9, 13

Two block geometries were defined for studying the virtual displacement field as follows:

C: a "narrow" block consisting of points 7, 11, 12, 16;
D: a "wide" block consisting of points 3, 11, 12, 19.

The units of the elements of the observation equations and the weight matrices were set so as to produce variance - covariance matrices (V-CM) with the following units:

V-CM of point coordinates "n" and "e" in [cm^2] (centimeters squared);
V-CM of datum-transformation/block partitioning parameters:

t(1) origin/translation in "n" [cm^2];
t(2) origin/translation in "e" [cm^2];
t(3) orientation/rotation of axes in [μrad^2] (microradians squared);
t(4) scale variation in [ppm^2] (parts per milion squared).

All the computations were performed with 12 significant digit accuracy on a PC using the APL*PLUS software. Results of the experiments are shown in the following tables:

Table 3.1: Standard deviations of point coordinates [cm^2].

mode I: Δn, Δe							mode II: Δs					
datum	F		A		B		F		A		B	
pt#	n	e	n	e	n	e	n	e	n	e	n	e
1	1.08	1.08	.96	.96	1.19	1.19	1.07	1.08	.94	.67	1.23	1.90
2	.93	.93	1.09	1.09	1.15	1.15	1.06	.84	1.35	1.15	1.82	1.83
3	1.08	1.08	1.31	1.31	1.35	1.35	1.07	1.08	1.72	1.52	2.57	2.04
4	.93	.93	.94	.94	.75	.75	.90	1.01	.93	1.11	.74	.56
5	.75	.75	.87	.87	.91	.91	.83	.77	.95	.96	1.25	1.04
6	.75	.75	.99	.99	1.05	1.05	.83	.77	1.31	1.04	2.01	1.28
7	.93	.93	1.21	1.21	1.25	1.25	.90	1.01	1.81	1.30	2.90	1.51
8	1.08	1.08	.88	.88	.78	.78	1.09	1.07	.93	.88	.69	.80
9	.75	.75	.65	.65	.64	.64	.75	.85	.69	.79	.61	.66
10	.67	.67	.89	.89	.91	.91	.76	.76	1.09	.95	1.52	.90
11	.75	.75	1.05	1.05	1.09	1.09	.75	.85	1.49	1.09	2.40	1.12
12	1.08	1.08	1.35	1.35	1.38	1.38	1.09	1.07	2.13	1.30	3.40	1.35
13	.93	.93	.94	.94	.75	.75	.90	1.01	.93	1.11	.74	.56
14	.75	.75	.87	.87	.91	.91	.83	.77	.95	.96	1.25	1.04
15	.75	.75	.99	.99	1.05	1.05	.83	.77	1.31	1.04	2.01	1.28
16	.93	.93	1.21	1.21	1.25	1.25	.90	1.01	1.81	1.30	2.90	1.51
17	1.08	1.08	.96	.96	1.19	1.19	1.07	1.08	.94	.67	1.23	1.90
18	.93	.93	1.09	1.09	1.15	1.15	1.06	.84	1.35	1.15	1.82	1.83
19	1.08	1.08	1.31	1.31	1.35	1.35	1.07	1.08	1.72	1.52	2.57	2.04

Note: F - minimum-trace datum; A - wide-geometry datum; B - narrow-geometry datum.

Table 3.2: Error ellipses : maj./min.semi-axes (a/b) [cm.], orientation (θ) [deg].

datum	F			A			B		
pt#	a	b	θ	a	b	θ	a	b	θ
1	1.09	1.07	60.0	.96	.65	13.6	1.93	1.19	-77.9
2	1.06	.84	.0	1.39	1.09	-23.8	2.29	1.20	-45.2
3	1.09	1.07	-60.0	1.90	1.29	-35.2	3.00	1.33	-35.2
4	1.06	.84	-60.0	1.16	.87	-65.0	.74	.56	-1.7
5	.85	.75	-30.0	.99	.92	-47.8	1.36	.89	-31.7
6	.85	.75	30.0	1.32	1.03	-11.5	2.13	1.06	-22.5
7	1.06	.84	60.0	1.83	1.27	-12.0	2.99	1.31	-16.0
8	1.09	1.07	.0	.93	.88	.0	.80	.69	90.0
9	.85	.75	90.0	.79	.69	90.0	.66	.61	90.0
10	.76	.76	8.2	1.09	.95	.0	1.52	.90	.0
11	.85	.75	90.0	1.49	1.09	.0	2.40	1.12	.0
12	1.09	1.07	.0	2.13	1.30	.0	3.40	1.35	.0
13	1.06	.84	60.0	1.16	.87	65.0	.74	.56	1.7
14	.85	.75	30.0	.99	.92	47.8	1.36	.89	31.7
15	.85	.75	-30.0	1.32	1.03	11.5	2.13	1.06	22.5
16	1.06	.84	-60.0	1.83	1.27	12.0	2.99	1.31	16.0
17	1.09	1.07	-60.0	.96	.65	-13.6	1.93	1.19	77.9
18	1.06	.84	.0	1.39	1.09	23.8	2.29	1.20	45.2
19	1.09	1.07	60.0	1.90	1.29	35.2	3.00	1.33	35.2

Table 3.2 applies to mode II only. The angle θ is between the major semi-axis "a" and "n".

Table 3.3: Variance-covariance matrices of rigid-motion and scale-variation parameters (t_{mg}).

mode I								
datum \ block	**C**				**D**			
A	1.2356	.5079	.1392	.0411	.9518	.4840	.0706	.0348
	.5079	1.2356	-.0411	-.1392	.4840	.9518	-.0348	-.0706
	.1392	-.0411	.0531	.0157	.0706	-.0348	.0316	.0156
	.0411	-.1392	.0157	.0531	.0348	-.0706	.0156	.0316
B	1.3275	.5747	.1388	.0410	1.0496	.5534	.0705	.0348
	.5747	1.3275	-.0410	-.1388	.5534	1.0496	-.0348	-.0705
	.1388	-.0410	.0531	.0157	.0705	-.0348	.0316	.0156
	.0410	-.1388	.0157	.0531	.0348	-.0705	.0156	.0316

mode II								
datum \ block	**C**				**D**			
A	2.7704	.0000	.3894	.0000	1.5765	.0000	.0050	.0000
	.0000	1.2234	.0000	-.1746	.0000	.9519	.0000	-.1156
	.3894	.0000	.2330	.0000	.0050	.0000	.1066	.0000
	.0000	-.1746	.0000	.0687	.0000	-.1156	.0000	.0433
B	3.9908	.0000	-.0320	.0000	2.9288	.0000	-.4129	.0000
	.0000	1.3512	.0000	-.1796	.0000	1.1045	.0000	-.1238
	-.0320	.0000	.3872	.0000	-.4129	.0000	.2460	.0000
	.0000	-.1796	.0000	.0687	.0000	-.1238	.0000	.0433

Units in the above table are: t(1), t(2) - [cm^2] ; t(3) - [μrad^2] ; t(4) - [ppm^2]

Table 3.4: Variance-covariance matrices of datum-transformation parameters (t_m).

datum		**A**			**B**			
mode I	t(1)	.25128		.17768	.32612	.23060		[cm^2]
	t(2)	.17768		.25128	.23060	.32612		[cm^2]
mode II	t(1)	.57675	.00000	-.13049	1.91514	.00000	-.54474	[cm^2]
	t(2)	.00000	.22493	.00000	.00000	.31342	.00000	[cm^2]
	t(3)	-.13049	.00000	.04310	-.54474	.00000	.17718	[μrad^2]

4. Conclusions. The importance of geometry.

Let us first take a closer look at tables 3.1 and 3.3. In those solutions under mode I (where GPS vectors are regarded as measurements which contain orientation information) the differences between results based on datum A (wide) and - on datum B (narrow) are hardly noticeable. Solutions obtained in mode II (see table 3.1 and 3.4) display clearly the effect of datum selection. The superiority of datum A (wide) over datum B (narrow) is highly significant. It is easy to see that the size and orientation of the error ellipses are a function of distance and di-

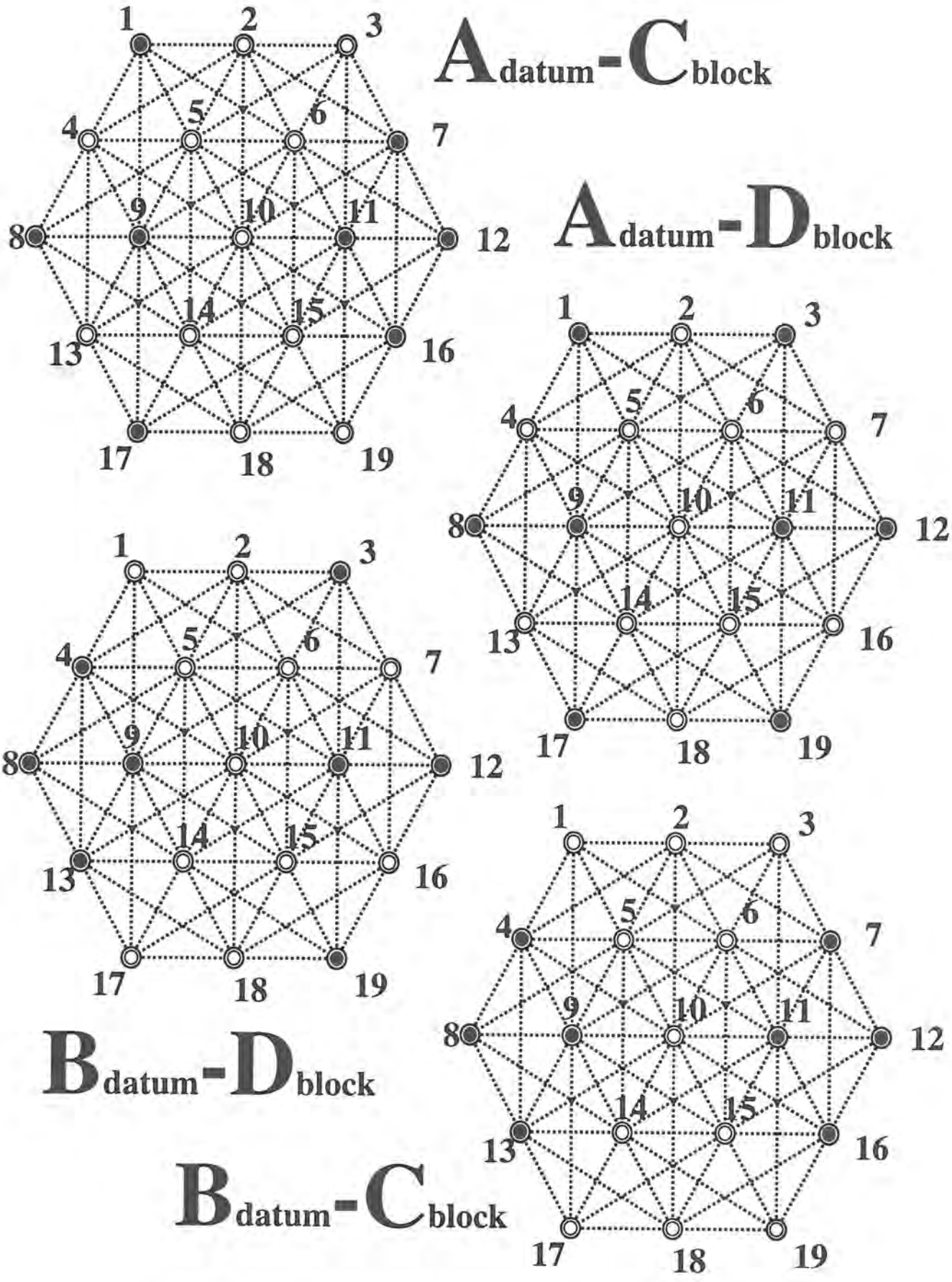


Figure 3.2: Four combinations of datum and block geometries

rection of the point relative to the "datum block" (table 3.2). The differences in "quality" of all three datum transformation parameters in mode II (see table 3.4) are indeed dramatic where variances pertaining to datum B (narrow) are 3 to 4 times larger as compared to those obtained under datum A (wide).

The effect of geometry on the quality of the parameters of block motion and strain (see table 3.3) is clearly demonstrated. The differences in mode I variances between B-C (narrow datum - narrow block) and the (wide datum-wide block) are of the order of no more than 30-40%. In mode II , however, the differences between B-C and A-D are distinctly larger. In particular it is worth noting the rotation parameter "t(3)", where due to the orientation "weakness" of datum B, the variance of the B-C t(3) parameter differs from that of A-D by almost 400%.

It is interesting to note in mode II the significant differences between the qualities of north-south t(1) and east-west t(2) origin/translation parameters. The span in the east-west direction of both the A as well as the B datum blocks is only 20 km., while the span in the north-south direction is 34.6 km. in datum B and 61 km. in datum A. The effect of the above contrast is felt consistently in the difference between variances of t(1) "n" and t(2) "e".

The results of our limited experiment as reported in this paper can be summarized as follows:

(a) The orientation content of the GPS measurements is extremely important and it made all the difference between mode I and mode II results. That is particularly important when constructing first order or "free" control networks. We should point out, however, that to ignore the geometry of the datum subset would be permissible only if the orientation content of the GPS measurements in terms of its consistency can be safely relied upon.

(b) Trilateration (mode II) as measured today by GPS is a very robust and accurate source of data. Yet, as in former "EDM times", it is notoriously week in orientation. To counteract that inherent weakness in orientation one has to make sure that the selected datum subset is characterized by a balanced and "good" geometry relative to the size and shape of the whole network.

References

Grafarend E.W. ,1974, "Optimization of Geodetic Networks", Bolletino di Geodesia a Science Affini, Vol. 33, No. 4.

Koch K.R., 1987, "Parameter Estimation and Hypothesis Testing in Linear Models", Springer Verlag, Berlin Heidelberg New York.

Meissl P., 1969, "Inner Theory of Errors of a Cluster of Points", Annex F in Systematic Investigations of Geodetic Networks in Space, (ed.) K. Rinner.

Papo H. and Perelmuter A. , 1981, "Datum Definition by Free Net Adjustment", Bulletin Geodesique, Vol.55, No.3.

Papo H., 1986, "Extended Free Net Adjustment Constraints", NOAA Tech. Rep., NOS 119 NGS 37, Rockville MD.

Papo H. , 1987, "Bases of Null Space in Analytical Photogrammetry", Photogrammetria, Vol. 41.

Papo H., 1989, "Modeling in Non-Cartesian Reference Frames", IAG General Meeting, Edinburgh, Scotland, UK.

Wolf H., 1977, "Eindeutige und mehrdeutige geodaetische Netze", Abh.d Braunschw. Wiss. Gesellschaft, Band 28, Goetingen, FRG.

Strain in the Earth — a Geodetic Perspective

W. Ian Reilly

Abstract

Recognition of the existence of horizontal displacement on faults in the Earth's crust became accepted only about a century ago, and with it came the theory of elastic rebound as the cause of earthquakes. This suggested the use of repeated geodetic measurements between widely distributed points on the Earth's surface to determine the accumulation or dissipation of elastic strain energy in the brittle crust. Such measurements can be modelled in terms of continuum mechanics, based on a three-dimensional vector field of particle velocity. The gradient of this vector field yields a set of invariants, or "estimable quantities", that characterise the rate of strain in the deforming medium. The use of the tensor calculus in formulating such continuum models concentrates attention on the underlying physical processes rather than on arbitrary coordinate systems. It also facilitates the use of higher-order spatial derivatives to describe the characteristics of heterogeneous strain — the bending of lines and the warping of surfaces — as well as providing for mathematically compact development of various interpolation schemes, including multi-dimensional polynomial expansions and least-squares collocation.

Key words: earth deformation, continuum mechanics, geodetic measurement.

Introduction

Omne tulit punctum qui miscuit utile dulci,
Lectorem delectando pariterque monendo.
(Full marks to him who combines profit with pleasure,
delighting the reader while instructing him — Horace, *Ars Poetica*)

There must be very few aspects of mathematical geodesy that have not been the subject of one of Professor Erik Grafarend's multifarious and erudite essays. His development with Burkhard Schaffrin of the concept of "estimable quantities" (Grafarend & Schaffrin 1974, 1976) continues to illuminate our understanding of what information can and cannot be extracted from a geodetic network. His work on the converse problem, the optimal design of geodetic networks, continues to be applied to deforming networks (*e.g.* Grafarend, 1986; Xu & Grafarend, 1995), and extended to the statistical analysis of the second-rank tensors that quantify strain and stress in the Earth (Xu & Grafarend, 1996a,b).

Much of the geodetic evidence of Earth deformation has come from pre-existing networks, and historic measurements, bereft of the advantage of advanced design. The present essay is concerned with an elementary continuum-mechanical interpretation of such measurements, extracting the strain invariants — the "estimable quantities" for the deforming medium. It acknowledges both the elucidative contribution of mathematical theory, and the dedication and professionalism of generations of surveyors who went into forest and desert and mountain and made better measurements than we could reasonably have demanded of them.

Origins

> "In no country, perhaps, where the English language is spoken, have earthquakes, or, to speak more correctly, the subterranean causes to which such movements are due, been so active in producing changes of geological interest as in New Zealand."
> (Lyell, 1872, vol. II: p. 82)

When Sir Charles Lyell wrote these words (with a fine chauvinistic flourish) in the Tenth Edition of his celebrated text *The Principles of Geology*, vertical displacement of the Earth's crust on faults was well known, and had been for many years: in 1802 Playfair had written

> "The greatest part of the facts relative to the fracture and dislocation of the strata, belongs to the history of veins...The frequency of these [slips], and their great extent, are well known wherever mines have been wrought." (Playfair, 1802, p 204)

In 1888 in North Canterbury, New Zealand, horizontal offsets of between 1.5 and 2.6 m were observed on the fault break that accompanied the earthquake of magnitude 7 that knocked the top off the Christchurch Cathedral spire. The Government geologist's report (McKay, 1890) was duly published and forgotten. This was but one example among many, from one small corner of the globe. Then, in 1906, the magnitude 8.3 earthquake that razed the city of San Francisco not only convinced the scientific world of the reality of horizontal displacement on faults — the San Andreas in this case — but also gave rise to Reid's theory of elastic rebound.

Geodetic evidence

It was soon realised that if earthquakes, and fault breaks, resulted from the sudden release of accumulated elastic strain in the crust, then the accumulation of strain between earthquakes might be manifest in the distortion of geodetic networks. Spurred by the Kwanto earthquake of 1923 in Japan, Terada & Miyabe (1929) derived and mapped the parameters of shear, rotation and dilatation from displacement vectors for each triangle of repeated surveys, and this type of study has since been continued. In New Zealand, H. W. Wellman, who had been a co-discoverer of the South Island's Alpine Fault, with its *ca.* 450 km dextral offset since the Cretaceous, determined the rate and orientation of shear strain from repeated triangulation in Marlborough (Wellman, 1955). This area, in the northern part of the South Island, together with the Alpine Fault itself, is now recognised as part of the obliquely convergent margin between the Pacific and the Australian lithospheric plates that passes through New Zealand.

The Marlborough region was further studied by H. M. Bibby, who developed a method of simultaneous reduction of repeated geodetic surveys, coincident in part or in whole, together with determination of deformation parameters (Bibby 1973, 1976, 1981). He found pervasive shear strain occurring over time spans of several decades in the absence of overt fault movement. This pioneering work has remained the basis for all subsequent analyses of geodetic data to determine earth deformation in New Zealand, and has had influence elsewhere.

What the geodetic evidence, then as now, leaves unresolved is the partitioning of observed shear strain between elastic (recoverable) and non-elastic (permanent) deformation. Does a low level of brittle failure indicate that strain energy is being dissipated in non-elastic creep, or simply that the observation period falls between major earthquakes, and that elastic strain is accumulating steadily? Information from more than purely geodetic measurements is needed even to discuss, much less resolve, such questions.

Continuum mechanics

A geodetic network is a discrete measuring system. The Earth's crust is riven by faults — discrete fractures. What, then, is the justification for describing deformation in terms of continuum mechanics? In essence, it is a question of scale. Even if deformation occurs by slip on a sequence of faults that separate rigid blocks, it can be treated as continuous if the fault separation is a sufficiently small fraction of the station spacing of the geodetic network. In the absence of obvious fault movement within the observing period, the continuum model is the most general and unbiased, and can always be superimposed on a discrete faulting model.

The problem of describing the deformation of the Earth's crust has much in common with fluid dynamics, and can use the methods of vector and tensor analysis made familiar in geodesy by Hotine (1969). There are two principal differences, however. The first is one of scale: the relative velocity between two lithospheric plates may be of the order of 50 mm/yr, or 1.5×10^{-9} m s^{-1}, compared, say,

with the velocity of sound at sea level of about 300 m s^{-1}: a ratio of 2×10^{11}. The second is one of the experimental environment: the experimenter in fluid dynamics is at no loss for a fixed frame of reference provided by the apparatus around, over, and through which his fluids flow; but the student of earth deformation can depend on no such reliable framework — it is as if he were floating on a broad river, with neither shore nor bottom discernible. This is no disadvantage if his object is to study the intrinsic deformation of the crust, and to rely on direct, local and differential, physical measurements: distinctions between *Lagrangian* and *Eulerian* coordinates can be cast aside, and tedious arguments about datums become irrelevant.

A primary advantage of using the methods of the tensor calculus in continuum mechanics resides in the immediate identification of vector and tensor quantities with objective physical fields, such as the particle velocity field of the deforming medium, independently of arbitrary coordinate systems. A second advantage is the ease with which covariant differentiation of such spatially variable velocity fields opens the way to the calculation of higher-order invariants — the "estimable quantities" of continuum mechanics — and the clarity with which these can be seen to be independent of particular coordinate systems.

Deformation in three dimensions

If the velocity of a material particle P is represented by the vector $\mathbf{u}^{\mathbf{r}}$, then the deformation in the vicinity of P is given by the gradient of the velocity vector, and represented by the covariant derivative of $\mathbf{u}^{\mathbf{r}}$, *viz.* $\mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}$. The tensor $\mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}$ has in general nine independent coefficients; it can be decomposed into symmetric and antisymmetric parts

$$\mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}} = \sigma^{\mathbf{r}}{}_{\mathbf{s}} + \tau^{\mathbf{r}}{}_{\mathbf{s}} \qquad (1)$$

The symmetric tensor $\sigma^{\mathbf{r}}{}_{\mathbf{s}}$, the strain rate tensor of six independent terms, describes the intrinsic deformation of the medium. An idealised test apparatus would comprise a regular tetrahedron embedded in the medium, where any one of its six sides could shorten or lengthen independently of the other five. The use of such an apparatus is unlikely, although a geodetic approximation could be attained by a suitable array of benchmarks on the floor and flanks of a deep valley, interconnected by distance measurements.

The antisymmetric tensor $\tau^{\mathbf{r}}{}_{\mathbf{s}}$, the rotation rate tensor of three independent terms, describes the mean rotation of the small volume about P with respect to some external reference frame. It can be represented by its equivalent rotation-rate vector $\mathbf{T}^{\mathbf{r}}$, defined by

$$\mathbf{T}^{\mathbf{r}} = -\tfrac{1}{2}\,\varepsilon^{\mathbf{rst}}\,\mathbf{a}_{\mathbf{su}}\,\tau^{\mathbf{u}}{}_{\mathbf{t}} = -\tfrac{1}{2}\,\varepsilon^{\mathbf{rst}}\,\mathbf{a}_{\mathbf{su}}\,\mathbf{u}^{\mathbf{u}}{}_{\mathbf{t}} \qquad (2)$$

where $\varepsilon^{\mathbf{rst}}$ is the alternating tensor, and $\mathbf{a}_{\mathbf{su}}$ the metric tensor in three dimensions.

Inverting (2) and substituting in (1), we have

$$\mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}} = \sigma^{\mathbf{r}}{}_{\mathbf{s}} + \mathbf{a}^{\mathbf{rp}}\,\varepsilon_{\mathbf{pqs}}\,\mathbf{T}^{\mathbf{q}} \qquad (3)$$

Dilatation and shear

The symmetric strain rate tensor $\sigma^{\mathbf{r}}{}_{\mathbf{s}}$ can be subjected to a principal axis decomposition. Let the principal axes be denoted by the set of orthogonal unit vectors ($\mathbf{i}^{\mathbf{r}}, \mathbf{j}^{\mathbf{r}}, \mathbf{k}^{\mathbf{r}}$), ordered such that the eigenvalues resulting from the decomposition of $\sigma^{\mathbf{r}}{}_{\mathbf{s}}$ are ranked as

$$\sigma^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{i}_{\mathbf{r}}\,\mathbf{i}^{\mathbf{s}} \geq \sigma^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{j}_{\mathbf{r}}\,\mathbf{j}^{\mathbf{s}} \geq \sigma^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{k}_{\mathbf{r}}\,\mathbf{k}^{\mathbf{s}} \qquad (4)$$

These eigenvalues represent the rates of extensional strain in the principal directions; their sum defines the rate of volumetric dilatation Θ by

$$\begin{aligned} 3\,\Theta &= (\,\sigma^{r}{}_{s}\,\mathbf{i}_{r}\,\mathbf{i}^{s} + \sigma^{r}{}_{s}\,\mathbf{j}_{r}\,\mathbf{j}^{s} + \sigma^{r}{}_{s}\,\mathbf{k}_{r}\,\mathbf{k}^{s}) \\ &= \sigma^{r}{}_{s}\,\delta^{s}{}_{r} = \sigma^{r}{}_{r} = \mathbf{u}^{r}{}_{s}\,\delta^{s}{}_{r} = \mathbf{u}^{r}{}_{r} \end{aligned} \tag{5}$$

where $\delta^{s}{}_{r}$ is the Kronecker (or substitution) tensor in three dimensions, and $\sigma^{r}{}_{r}$ is the trace of the symmetric tensor $\sigma^{r}{}_{s}$.

Subtraction of the effect of isotropic expansion or contraction from the strain rate tensor $\sigma^{r}{}_{s}$ leads to the following tensor of five independent terms that represents the rate of shearing, or pure change of shape

$$\begin{aligned} (\sigma^{r}{}_{s} - \Theta\,\delta^{r}{}_{s}) = {} & (\,\sigma^{u}{}_{t}\,\mathbf{i}_{u}\,\mathbf{i}^{t} - \Theta)\,\mathbf{i}^{r}\,\mathbf{i}_{s} + (\,\sigma^{u}{}_{t}\,\mathbf{j}_{u}\,\mathbf{j}^{t} - \Theta)\,\mathbf{j}^{r}\,\mathbf{j}_{s} \\ & + (\,\sigma^{u}{}_{t}\,\mathbf{k}_{u}\,\mathbf{k}^{t} - \Theta)\,\mathbf{k}^{r}\,\mathbf{k}_{s} \end{aligned} \tag{6}$$

where $\sigma^{r}{}_{s}$ has been expanded in terms of the principal directions.

One way of representing this tensor is to introduce two scalar magnitudes, γ_1 and γ_2, in addition to the three independent terms that define the three principal axes, *viz.*

$$\begin{aligned} \gamma_1 &= \sigma^{u}{}_{t}\,\mathbf{i}_{u}\,\mathbf{i}^{t} - \Theta \\ \gamma_2 &= -(\,\sigma^{u}{}_{t}\,\mathbf{k}_{u}\,\mathbf{k}^{t} - \Theta\,) \end{aligned} \tag{7}$$

where γ_1 and γ_2 correspond to the greatest and least rates of extensional strain in the principal directions, respectively, and are defined so that both $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$.

In terms of these parameters, the eigenvalues are now

$$\begin{aligned} \sigma^{r}{}_{s}\,\mathbf{i}_{r}\,\mathbf{i}^{s} &= \Theta + \gamma_1 \\ \sigma^{r}{}_{s}\,\mathbf{j}_{r}\,\mathbf{j}^{s} &= \Theta + (\,\gamma_2 - \gamma_1\,) \\ \sigma^{r}{}_{s}\,\mathbf{k}_{r}\,\mathbf{k}^{s} &= \Theta - \gamma_2 \end{aligned} \tag{8}$$

and the deformation rate tensor can be expressed as

$$\mathbf{u}^{r}{}_{s} = \Theta\,\delta^{r}{}_{s} + \gamma_1\,(\,\mathbf{i}^{r}\,\mathbf{i}_{s} - \mathbf{j}^{r}\,\mathbf{j}_{s}\,) + \gamma_2\,(\,\mathbf{j}^{r}\,\mathbf{j}_{s} - \mathbf{k}^{r}\,\mathbf{k}_{s}\,) + \mathbf{a}^{rp}\,\varepsilon_{pqs}\,\mathbf{T}^{q} \tag{9}$$

The change of shape of the medium in the vicinity of a point P can be measured by the rate of change in the angle between two distinct lines of material particles. Let $\mathbf{l}^{r}$, $\mathbf{m}^{r}$, $\mathbf{n}^{r}$, be an arbitrary right-handed set of orthogonal unit vectors, of which $\mathbf{l}^{r}$, $\mathbf{m}^{r}$, represent two such lines of particles. The rate of change of the right angle between $\mathbf{l}^{r}$ and $\mathbf{m}^{r}$ is given by the difference in the rates of rotation of the particles in these directions about the axis $\mathbf{n}^{r}$, *viz.*

$$\mathbf{u}^{r}{}_{s}\,\varepsilon_{pqr}\,\mathbf{n}^{q}\,(\,\mathbf{l}^{s}\,\mathbf{l}^{p} - \mathbf{m}^{s}\,\mathbf{m}^{p}\,) = -\mathbf{u}^{r}{}_{s}\,(\,\mathbf{m}^{s}\,\mathbf{l}_{r} + \mathbf{l}^{s}\,\mathbf{m}_{r}\,) \tag{10}$$

In substituting for $\mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}$ from (9), the terms in Θ and $\mathbf{T}^{\mathbf{q}}$ will vanish, and the resultant expression will contain the scalar terms γ_1 and γ_2, and the three Euler angles that relate the ($\mathbf{l}^{\mathbf{r}}$, $\mathbf{m}^{\mathbf{r}}$, $\mathbf{n}^{\mathbf{r}}$) triad to the principal directions ($\mathbf{i}^{\mathbf{r}}$, $\mathbf{j}^{\mathbf{r}}$, $\mathbf{k}^{\mathbf{r}}$).

Because of the restriction of extensive geodetic measurements to the Earth's surface, the vertical gradient of the velocity vector is in general unobservable, *i.e.* if $\mathbf{n}^{\mathbf{r}}$ is a unit vector in the vertical direction, the values of $\mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{n}^{\mathbf{s}}$ are usually unattainable. The "estimable quantities" are therefore reduced to six, and most of the discussion of the results of geodetic measurement of earth deformation is in terms of two-dimensional deformation, either in plane or spherical approximation.

Deformation in two dimensions

Four of the above six estimable quantities appear in the two-dimensional form of the expression for the velocity gradient of equation (9), *viz.*

$$\begin{aligned} \mathbf{u}^{\alpha}{}_{\beta} &= \sigma^{\alpha}{}_{\beta} + \tau^{\alpha}{}_{\beta} \\ &= \Delta\,\delta^{\alpha}{}_{\beta} + \gamma\,(\mathbf{j}^{\alpha}\mathbf{j}_{\beta} - \mathbf{k}^{\alpha}\mathbf{k}_{\beta}) - \varepsilon^{\alpha\eta}\,\mathbf{a}_{\eta\beta}\,\Omega \end{aligned} \tag{11}$$

where the Greek subscripts & superscripts now denote two-dimensional surface vectors and tensors, and we have introduced Δ as the rate of areal dilatation, Ω as the scalar rate of mean rotation in the two-dimensional surface, and γ as the magnitude of the rate of shear strain (the *tensor shear*, in contrast to the *engineering shear*, $2\,\gamma$). To these three invariants we can add one principal direction (either $\mathbf{j}^{\alpha}$ or $\mathbf{k}^{\alpha}$) to fully specify the rate of deformation. In the case of the intrinsic strain rate $\sigma^{\alpha}{}_{\beta}$ in two dimensions, the idealised test apparatus would comprise an equilateral triangle attached to the surface, an arrangement that is closely approximated by triangulation and trilateration networks.

Again denoting by $\mathbf{n}^{\mathbf{r}}$ a unit vector in the vertical direction, the rate of areal dilatation Δ, a linear invariant of $\mathbf{u}^{\alpha}{}_{\beta}$, can be related to the volumetric dilatation Θ by the definition

$$\Delta = \tfrac{1}{2}\,\mathbf{u}^{\alpha}{}_{\beta}\,\delta^{\alpha}{}_{\beta} = \tfrac{1}{2}\,\mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}\,(\delta^{\mathbf{s}}{}_{\mathbf{r}} - \mathbf{n}^{\mathbf{s}}\,\mathbf{n}_{\mathbf{r}}) = \tfrac{1}{2}\,(3\,\Theta - \mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{n}^{\mathbf{s}}\,\mathbf{n}_{\mathbf{r}}) \tag{12}$$

Thus the rate of areal dilatation differs from the rate of volumetric dilatation by the magnitude of the rate of vertical extensional strain. If it can be assumed that the volumetric strain Θ is zero, *i.e.* that the medium is incompressible, then the rate of areal dilatation Δ can be taken as a measure of the vertical extensional strain rate.

The scalar rotation rate Ω, a second linear invariant of $\mathbf{u}^{\alpha}{}_{\beta}$, is related to the three-dimensional vector $\mathbf{T}^{\mathbf{q}}$ through the definition

$$\Omega = -\tfrac{1}{2}\,\mathbf{u}^{\alpha}{}_{\beta}\,\mathbf{a}_{\alpha\eta}\,\varepsilon^{\eta\beta} = -\tfrac{1}{2}\,\varepsilon^{\mathbf{rst}}\,\mathbf{a}_{\mathbf{su}}\,\mathbf{u}^{\mathbf{u}}{}_{\mathbf{t}}\,\mathbf{n}_{\mathbf{r}} = \mathbf{T}^{\mathbf{r}}\,\mathbf{n}_{\mathbf{r}} \tag{13}$$

The magnitude γ of the shear strain rate can be derived as the quadratic invariant

$$\gamma^2 = \tfrac{1}{4}\,\mathbf{u}^{\alpha}{}_{\beta}\,\mathbf{u}^{\gamma}{}_{\delta}\,(\mathbf{a}_{\alpha\gamma}\,\mathbf{a}^{\beta\delta} - \varepsilon_{\alpha\gamma}\,\varepsilon^{\beta\delta}) \tag{14}$$

The unit vectors $\mathbf{j}^{\alpha}$ and $\mathbf{k}^{\alpha}$ give the principal directions of the symmetric tensor $\sigma^{\alpha}{}_{\beta}$ corresponding to the directions of maximum and minimum extensional strain, respectively. The eigenvalues are thus

$$\sigma^{\alpha}{}_{\beta}\,\mathbf{j}_{\alpha}\,\mathbf{j}^{\beta} = \mathbf{u}^{\alpha}{}_{\beta}\,\mathbf{j}_{\alpha}\,\mathbf{j}^{\beta} = \Delta + \gamma$$

$$\sigma^{\alpha}{}_{\beta}\, k_{\alpha}\, k^{\beta} = u^{\alpha}{}_{\beta}\, k_{\alpha}\, k^{\beta} = \Delta - \gamma \tag{15}$$

By analogy with (10), the rate of shear with respect to an arbitrary pair of orthogonal unit vectors $\mathbf{l}^{\mathbf{r}}$, $\mathbf{m}^{\mathbf{r}}$, is

$$u^{\alpha}{}_{\beta}\, \varepsilon_{\alpha\delta}\, (l^{\beta} l^{\delta} - m^{\beta} m^{\delta}) = -u^{\alpha}{}_{\beta}\, (m^{\beta} l_{\alpha} + l^{\beta} m_{\alpha}) = 2\gamma \sin 2\varphi \tag{16}$$

where φ is the angle between $\mathbf{l}^{\alpha}$ and the principal direction $\mathbf{i}^{\alpha}$ in the direction of maximum relative extension.

The advantage in calculating the shear strain rate is twofold:

- it can be derived from observations of changes in shape only, where no accurate length scale is available (as for repeated triangulations);
- it is the quantity that most accurately reflects the continuous accumulation of elastic strain energy, and thus presages brittle failure in elastic media.

Shear strain rate alone can be depicted as a line symbol of magnitude γ, with the direction (though not the sense) of either the maximum relative extension $\mathbf{j}^{\alpha}$, the maximum relative contraction $\mathbf{k}^{\alpha}$, or even of one or other of the directions of maximum shear which bisect the right angle between the directions $\mathbf{j}^{\alpha}$ and $\mathbf{k}^{\alpha}$. Alternatively, the magnitude γ of the shear strain rate can be plotted and contoured as a scalar variable without reference to the directions of the principal axes.

Heterogeneous strain in two dimensions: bending

If the rate of strain is constant over some region, the strain is said to be *homogeneous*, and the gradient of the deformation rate tensor is then zero

$$u^{\alpha}{}_{\beta\gamma} = 0 \tag{17}$$

On the other hand, if the rate of strain varies within the region, the strain is *heterogeneous*.

If the observed strain is heterogeneous, this will be made obvious in plotting different values of the rate of dilatation, or different values of the magnitude of the shear strain rate and its associated direction across the region. The question arises, however, as to whether one or more functions of $u^{\alpha}{}_{\beta\gamma}$ might be used to display the character of the heterogeneous strain.

From an inspection of equation (11), it is apparent that the gradients of the two linear invariants, Δ and Ω, will yield vectors that point towards regions of greater or lesser rates of dilatation or rotation. However, as even the gradient of the intrinsic strain rate tensor $\sigma^{\alpha}{}_{\beta\gamma}$ has six independent parameters, there is a large number of derived functions available.

One characteristic of homogeneous strain is that any line of material particles that was originally straight remains straight after straining. Under heterogeneous strain, such a line of particles would in general become curved. This suggests that the *rate of bending*, or of change of curvature, is a quantity that not only reflects an observed characteristic of many geological structures, but also could be used to describe one aspect of heterogeneous strain.

Introducing an arbitrary pair of orthogonal unit vectors, $\mathbf{l}^{\alpha}$ and $\mathbf{m}^{\alpha}$, we may express the velocity of a line of particles in the direction of $\mathbf{l}^{\alpha}$ resolved into components along the line and normal to it

$$u^{\alpha}{}_{\beta}\, l^{\beta} = e\, l^{\alpha} + r\, m^{\alpha} \tag{18}$$

where $\mathbf{e}$ is the rate of extensional strain in the direction $\mathbf{l}^{\alpha}$, and $\mathbf{r}$ is the rate of rotation of the line of particles in the same direction. Hence

$$\mathbf{r}(\mathbf{l}^{\alpha}) = \mathbf{u}^{\alpha}{}_{\beta}\,\mathbf{l}^{\beta}\,\mathbf{m}_{\alpha} \tag{19}$$

We define the rate of bending $\rho(\mathbf{l}^{\alpha})$ of the line of particles in the direction $\mathbf{l}^{\alpha}$ as the gradient of the rate of rotation in that same direction, *viz.*

$$\rho(\mathbf{l}^{\alpha}) = \mathbf{r}_{\gamma}(\mathbf{l}^{\alpha})\,\mathbf{l}^{\gamma} = \mathbf{u}^{\alpha}{}_{\beta\gamma}\,\mathbf{l}^{\beta}\,\mathbf{m}_{\alpha}\,\mathbf{l}^{\gamma} \tag{20}$$

If we have the values of $\mathbf{u}^{\alpha}{}_{\beta\gamma}$, then we can derive the directions of extreme or zero bending at a point. Since (20) is a cubic function of direction, there will be either one or three axes of maximum bending, and one or three axes of zero bending, which is a little more complicated than the intuitive concept of the bending of a linear structure, such as a beam.

There are in all three distinct contractions of $\mathbf{u}^{\alpha}{}_{\beta\gamma}$ with the orthogonal unit vectors $\mathbf{l}^{\alpha}$ and $\mathbf{m}^{\alpha}$ of the form given in (20), as well as analogous forms in three dimensions (Reilly, 1986), but the bending rate in two dimensions is probably the most accessible.

Tilting and warping of surfaces

The vertical component of the velocity of a material point P is denoted by

$$\mathbf{h}_{\mathbf{s}} = \mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{n}_{\mathbf{r}} \tag{21}$$

where $\mathbf{n}_{\mathbf{r}}$ is a unit vector in the direction of the vertical. The *tilt rate vector* in three dimensions is defined as the gradient of the vertical velocity

$$\mathbf{h}_{\mathbf{s}} = \mathbf{u}^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{n}_{\mathbf{r}} = \sigma^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{n}_{\mathbf{r}} + \mathbf{n}^{\mathbf{p}}\,\varepsilon_{\mathbf{pqs}}\,\mathbf{T}^{\mathbf{q}} \tag{22}$$

The tilt rate in the direction of an arbitrary horizontal unit vector $\mathbf{l}^{\mathbf{r}}$ is

$$\mathbf{h}_{\mathbf{s}}\,\mathbf{l}^{\mathbf{s}} = \sigma^{\mathbf{r}}{}_{\mathbf{s}}\,\mathbf{n}_{\mathbf{r}}\,\mathbf{l}^{\mathbf{s}} + \mathbf{n}^{\mathbf{p}}\,\mathbf{l}^{\mathbf{s}}\,\varepsilon_{\mathbf{pqs}}\,\mathbf{T}^{\mathbf{q}} \tag{23}$$

showing that the tilt rate is a combination of the intrinsic shear strain rate in the vertical plane containing $\mathbf{l}^{\mathbf{r}}$, and the extrinsic mean rotation rate about a horizontal axis normal to $\mathbf{l}^{\mathbf{r}}$. These components cannot be separated on the basis of horizontal tilt measurements alone, as where the tilt rate is found by such geodetic measurements as repeated spirit levelling, or repeated GPS observations.

It is usually more convenient to define the tilt rate as a two-dimensional vector $\mathbf{h}_{\alpha}$ in the horizontal plane: this accounts for the remaining two of the six "estimable quantities" generally attainable by geodetic measurements on the Earth's surface. An accessible measure of the *intrinsic* strain rate is then given by its gradient $\mathbf{h}_{\alpha\beta}$ (*cf.* Hein & Kistermann 1981). If $\mathbf{l}^{\alpha}$ is a unit surface vector coincident with the space vector $\mathbf{l}^{\mathbf{r}}$ of (23) above, the *rate of change of surface curvature* in the direction of $\mathbf{l}^{\alpha}$ is

$$\mathbf{h}_{\alpha\beta}\,\mathbf{l}^{\alpha}\,\mathbf{l}^{\beta} = \mathbf{u}^{\mathbf{r}}{}_{\mathbf{st}}\,\mathbf{n}_{\mathbf{r}}\,\mathbf{l}^{\mathbf{s}}\,\mathbf{l}^{\mathbf{t}} \tag{24}$$

By comparison with equation (20), the expression on the right-hand-side of (24) is seen to be equivalent to the *rate of bending* in the vertical plane of the line of particles in the direction $\mathbf{l}^{\mathbf{r}}$.

The tensor $\mathbf{h}_{\alpha\beta}$ is symmetric; its decomposition is analogous to that for two-dimensional strain in (11), *viz.*

$$\mathbf{h}_{\alpha\beta} = \mathbf{H}\,\mathbf{a}_{\alpha\beta} + \mathbf{D}\,(\mathbf{j}_\alpha\,\mathbf{j}_\beta - \mathbf{k}_\alpha\,\mathbf{k}_\beta) \tag{25}$$

where the *rate of change of mean curvature* of the surface is

$$\mathbf{H} = \tfrac{1}{2}\,\mathbf{h}_{\alpha\beta}\,\mathbf{a}^{\alpha\beta} \tag{26}$$

and the *rate of change of torsion* $\mathbf{D}$ of the surface is found from

$$\mathbf{D}^2 = \tfrac{1}{4}\,\mathbf{h}_{\alpha\beta}\,\mathbf{h}_{\gamma\delta}\,(\mathbf{a}^{\alpha\gamma}\,\mathbf{a}^{\beta\delta} - \varepsilon^{\alpha\gamma}\,\varepsilon^{\beta\delta}) \tag{27}$$

The maximum and minimum values of the rate of change of curvature (or of bending in the vertical plane) are given by the eigenvalues of the symmetric tensor $\mathbf{h}_{\alpha\beta}$, *viz.*

$$\begin{aligned} \mathbf{h}_{\alpha\beta}\,\mathbf{j}^\alpha\,\mathbf{j}^\beta &= \mathbf{H} + \mathbf{D} \\ \mathbf{h}_{\alpha\beta}\,\mathbf{k}^\alpha\,\mathbf{k}^\beta &= \mathbf{H} - \mathbf{D} \end{aligned} \tag{28}$$

Determination of strain from geodetic observations

Following the general principle introduced by Bibby (1973, 1976, 1981), the displacements of geodetic bench-marks can be modelled so as to permit a unified solution of geodetic measurements made at different epochs. The results comprise

- a set of coordinates for each bench-mark at some reference epoch;
- a set of parameters defining the velocity field, either continuous or discrete in time and space.

Amongst the continuum models used or proposed for interpolating velocities and rates of deformation, Grafarend (1986) has noted that the geodetic network is an actualisation of a finite element model, and coupled this with a local spline interpolation. Spline interpolation is also the basis of an application by Haines & Holt (1993) of a finite element model to the spherical surface of the Earth. Two further models — polynomial expansion and least-squares collocation — will be briefly mentioned here.

Polynomial approximation

The velocity vector at a point P can be expressed as a Taylor's series expansion about a suitable origin P_0

$$\mathbf{u}^\alpha = \mathbf{b}^\alpha + \mathbf{b}^\alpha{}_\beta\,\mathbf{y}^\beta + \mathbf{b}^\alpha{}_{\beta\gamma}\,\mathbf{y}^\beta\,\mathbf{y}^\gamma/2! + \mathbf{b}^\alpha{}_{\beta\gamma\delta}\,\mathbf{y}^\beta\,\mathbf{y}^\gamma\,\mathbf{y}^\delta/3! + \ldots \tag{29}$$

where

$\mathbf{y}^\beta$ is a position vector, with gradient $\mathbf{y}^\beta{}_\gamma = \delta^\beta{}_\gamma$ (in a Euclidean space), and

$\mathbf{b}^\alpha$, $\mathbf{b}^\alpha{}_\beta$, $\mathbf{b}^\alpha{}_{\beta\gamma}$, $\mathbf{b}^\alpha{}_{\beta\gamma\delta}$, ..., are constant coefficients to be determined.

The deformation rate tensor is found by covariant differentiation of (29) as

$$\mathbf{u}^\alpha{}_\eta = \mathbf{b}^\alpha{}_\eta + \mathbf{b}^\alpha{}_{\beta\eta}\,\mathbf{y}^\beta + \mathbf{b}^\alpha{}_{\beta\gamma\eta}\,\mathbf{y}^\beta\,\mathbf{y}^\gamma/2! + \ldots \tag{30}$$

With adequate error determination, the series expansion can be truncated to exclude insignificant terms. Low-order expansions are of value in smoothing the results from noisy data. Application to extensive regions (such as the order of 10^5 km^2 in Reilly, 1990) can be criticised as forcing a pattern on complex data, a universal hazard of polynomial approximation methods.

Least-squares collocation

The interpolation of a vector field of displacements or velocities of material points would seem to be an ideal subject for least-squares collocation. Deakin *et al.* (1994) have applied the method to interpolating the displacements of the three-dimensional coordinates of points of a geodetic network in Victoria, Australia. In this they used a triplet of covariance functions, one for each coordinate direction. In a study of the prediction of horizontal strain in Japan, El-Fiky && Kato (1999) assumed the covariance between point displacements to be "homogeneous and isotropic", but used a separate covariance function for each of the E-W and N-S components of the observed displacement vectors.

In neither of these examples is it clear that the difference between the covariance function for different components is of any real significance in the interpolation process. Moreover, the assignation of different covariance functions to different components of the displacement field amounts to defining a covariance function for the *vector field* that is anisotropic with respect to the azimuth of the parallel components of the two vectors, with axes of anisotropy coinciding with arbitrarily chosen coordinate directions. In short, there seems to be no good reason to go beyond a simple function that is isotropic both with respect to the relative orientation of pairs of points, and also with respect to the orientation of any arbitrary pair of parallel vector components, and where the correlation between orthogonal vector components is zero.

Let a material particle P have a position denoted by the vector $\mathbf{x}^i$, and to be moving with a velocity $\mathbf{u}^i(P) = \mathbf{d}\,\mathbf{x}^i(P) / \mathbf{d}\,\mathbf{t}$. Let a similar particle Q have a position $\mathbf{x}^i(Q)$ and velocity $\mathbf{u}^i(Q)$. Assuming a Euclidean space, we denote the vector PQ by

$$\mathbf{p}^i = \mathbf{r}\,\mathbf{m}^i = \mathbf{x}^i(Q) - \mathbf{x}^i(P) \tag{31}$$

where $\mathbf{r}$ is the length of the vector $\mathbf{p}^i$,

$\mathbf{m}^i$ is the unit vector in the direction of $\mathbf{p}^i$, such that $\mathbf{m}^i\,\mathbf{m}_i = 1$.

As an example of a covariance tensor $\mathbf{C}^{jk}$ between the vectors $\mathbf{u}^j(P)$ and $\mathbf{u}^k(Q)$ based on a simple Gaussian function, that is simultaneously homogeneous, and isotropic in both the senses discussed above, we can write

$$\mathbf{C}^{jk}\{\mathbf{u}^j(P), \mathbf{u}^k(Q)\} = \mathbf{C_0}\,\mathbf{a}^{jk} \exp(-\mathbf{r}^2 / 2\mathbf{d}^2) \tag{32}$$

where $\mathbf{C_0}$ is a constant of dimension $(velocity)^2$,

$\mathbf{d}$ is a constant of dimension *length*,

$\mathbf{a}^{jk}$ is the metric tensor in three dimensions.

If $\mathbf{f}_j$, $\mathbf{g}_j$ are two arbitrary unit vectors, then the covariance between two velocity components is the scalar quantity

$$\begin{aligned}\mathbf{C}^{jk}\{\mathbf{u}^j(P)\,\mathbf{f}_j, \mathbf{u}^k(Q)\,\mathbf{g}_k\}\,\mathbf{f}_j\,\mathbf{g}_k &= \mathbf{C_0}\,\mathbf{a}^{jk}\,\mathbf{f}_j\,\mathbf{g}_k \exp(-\mathbf{r}^2 / 2\mathbf{d}^2) \\ &= \mathbf{C_0} \cos\theta \exp(-\mathbf{r}^2 / 2\mathbf{d}^2)\end{aligned} \tag{33}$$

where θ is the angle between $\mathbf{f}_j$ and $\mathbf{g}_j$.

The argument can be extended to calculate the covariance between the velocity vector $\mathbf{u}^{k}(Q)$ at Q and the deformation rate tensor $\mathbf{u}^{j}_{l}(P)$ at P by taking the covariant derivative of (32) with respect to $\mathbf{x}^{j}(P)$

$$\mathbf{C}^{jk}_{l}\{\mathbf{u}^{j}_{l}(P), \mathbf{u}^{k}(Q)\} = \mathbf{C}_0\, \mathbf{a}^{jk}\, \mathbf{m}_l\, (\mathbf{r}/\mathbf{d}^2) \exp(-\mathbf{r}^2/2\mathbf{d}^2) \qquad (34)$$

If $\mathbf{f}_j$, $\mathbf{g}_j$, $\mathbf{h}_j$ are three arbitrary unit vectors, then the scalar covariance between arbitrary components of the velocity and of the deformation rate tensor

$$\begin{aligned}\mathbf{C}^{jk}_{l}\{\mathbf{u}^{j}_{l}(P)\,\mathbf{f}_j, \mathbf{u}^{k}(Q)\,\mathbf{g}_k\,\mathbf{h}^l\}\,\mathbf{f}_j\,\mathbf{g}_k\,\mathbf{h}^l \\ = \mathbf{C}_0\, \mathbf{a}^{jk}\, \mathbf{f}_j\, \mathbf{g}_k\, \mathbf{h}^l\, \mathbf{m}_l\, (\mathbf{r}/\mathbf{d}^2) \exp(-\mathbf{r}^2/2\mathbf{d}^2) \\ = \mathbf{C}_0 \cos\theta \cos\psi\, (\mathbf{r}/\mathbf{d}^2) \exp(-\mathbf{r}^2/2\mathbf{d}^2) \qquad (35)\end{aligned}$$

where ψ is the angle between $\mathbf{h}^l$ and the direction $\mathbf{m}_l$ of the line PQ. This provides a basis for interpolating the deformation rate tensor $\mathbf{u}^{j}_{l}$ at P from observed velocities at a series of N points (Q_1, $Q_2 \ldots Q_N$).

Given that the line PQ between two points on the Earth's surface will generally be nearly horizontal, then for vertical derivatives of $\mathbf{u}^{j}$, the angle ψ will be close to a right angle, and the scalar covariance in this case will tend to zero. This is just another way of stating that the vertical gradient of the velocity field is not determinable from observations of the velocity vector distributed over the surface, and that only six of the nine components of the deformation rate tensor can in general be found from such observations.

Conclusions

Repeated geodetic observations yield estimates of the particle velocity field of the deforming Earth, sampled at the network of observing points. Evaluation of such observations in terms of the spatial gradient of a continuous velocity field leads to the determination of such invariants as the rates of dilatation, rotation, and shear. These are the "estimable quantities" that best characterise the state of strain in the Earth, and for which purpose are more suited than the velocity field itself. The methods of the tensor calculus are particularly apt for the clear and unambiguous derivation of such invariants, as for many other manipulations of vector fields, and thus are very much in the spirit of the rigorous approach to geodetic problems that has been demonstrated by Erik Grafarend and his co-workers over many years.

I am grateful to Dr Hugh Bibby for his comments on the draft of this manuscript.

References

Bibby, H. M., 1973: The reduction of geodetic survey data for the detection of earth deformation. *Geophysics Division Report No. 84*, Department of Scientific & Industrial Research, Wellington.

Bibby, H. M., 1976: Crustal strain across the Marlborough Faults, New Zealand. *New Zealand Journal of Geology & Geophysics 19:* 407-425.

Bibby, H. M., 1981: Geodetically determined strain across the southern end of the Tonga-Kermadec-Hikurangi subduction zone. *Geophysical Journal of the Royal Astronomical Society 66:* 513-533.

Deakin, R. E., Collier, P. A., Leahy, F. J., 1994: Transformation of coordinates using least squares collocation. *The Australian Surveyor, March 1994:* 6-20.

El-Fiky, G. S., Kato, T., 1999: Continuous distribution of horizontal strain in the Tohoku district, Japan, predicted by least-squares collocation. *Journal of Geodynamics 27:* 213-236.

Grafarend, E., Schaffrin, B., 1974: Unbiased free net adjustment. *Survey Review 22*: 200-218.

Grafarend, E., Schaffrin, B., 1976: Equivalence of estimable quantities and invariants in geodetic networks. *Zeitschrift für Vermessungswesen 101:* 485-491.

Grafarend, E. W., 1986: Three-dimensional deformation analysis; global vector spherical harmonic and local finite element representation. *Tectonophysics 130:* 337-359.

Haines, A. J., Holt, W. E., 1993: A procedure for obtaining the complete horizontal motions within zones of distributed deformation from the inversion of strain rate data. *Journal of Geophysical Research 98:* 12,057-12,082.

Hein, G. W., Kistermann, R., 1981: Mathematical foundation of non-tectonic effects in geodetic recent crustal movement models. *Tectonophysics 71:* 315-334.

Hotine, Martin, 1969: *Mathematical Geodesy*. U. S. Department of Commerce, Washington, D.C.

Lyell, Charles, 1872: *Principles of Geology*. 11th Ed., John Murray, London.

McKay, A., 1890: On the earthquakes of September 1888, in the Amuri and Marlborough districts of the South Island. *New Zealand Geological Survey report of geological explorations 20:* 1-16.

Playfair, John, 1802: *Illustrations of the Huttonian Theory of the Earth*. Dover, New York (reprint of 1964).

Reilly, W. I., 1986: Heterogeneous strain in earth deformation. *Geophysics Division Report No. 210*, Department of Scientific & Industrial Research, Wellington.

Reilly, W. I., 1990: Horizontal crustal deformation on the Hikurangi Margin. *New Zealand Journal of Geology & Geophysics 33:* 393-400.

Terada, T., Miyabe, N., 1929: Deformation of the Earth Crust in Kwansai Districts and its relation to the Orographic Feature. *Bulletin Earthquake Research Institute 7:* 223-239.

Wellman, H. W., 1955: New Zealand Quaternary tectonics. *Geologische Rundschau 43:* 248-257.

Xu, P., Grafarend, E., 1995: A multi-objective second-order optimal design for deforming networks. *Geophysical Journal International 120:* 577-589.

Xu, P., Grafarend, E., 1996a: Probability distribution of eigenspectra and eigendirections of a two-dimensional, symmetric rank two random tensor. *Journal of Geodesy 70:* 419-430.

Xu, P., Grafarend, E., 1996b: Statistics and geometry of the eigenspectra of thee-dimensional second-rank symmetric random tensors. *Geophysical Journal International 127:* 744-756.

Continuously Operating GPS Reference Station Networks: New Algorithms and Applications of Carrier Phase-Based, Medium-Range, Static and Kinematic Positioning

Chris Rizos, Shaowei Han, Horng-Yue Chen and Goh Pong Chai

ABSTRACT & INTRODUCTORY REMARKS

Continuously operating GPS networks have been used for many years to address two types of positioning applications. The first, perhaps best known application, is in relation to geodetic objectives such as the determination of crustal motion on a variety of spatial scales - from the measurement of the broad kinematics of tectonic plates to deformation monitoring of local areas undergoing subsidence (e.g. due to fluid extraction and underground mining), surface inflation (e.g. due to magma intrusion under volcano domes), or the complex faulting in seismically active zones. Currently, permanent GPS stations around the world which have been established to address such geodetic applications number well over a thousand. Hundreds of these stations are now formally part of the global network organised under the auspices of the International Association of Geodesy (IAG) as the well-known International GPS Service (IGS). Data from this network is collected on a daily basis, centrally archived at data centres and is available via the Internet to users. Many of these users, through the application of special carrier phase processing techniques, are able to achieve relative station coordinate accuracies at the "few parts per billion" level (sub-centimetre accuracy for receiver separations perhaps a thousand kilometres or more).

The second class of applications that have traditionally been addressed by continuously operating GPS receiver networks are those that require real-time, differential GPS (DGPS) services to determine coordinates to accuracies of the order of a few metres. Local Area DGPS systems rely on a single reference station generating correction messages, which are transmitted to users over a wireless link. The DGPS corrections to pseudo-range data can also be determined at groups of permanent GPS stations, to service wide continental or oceanic regions. This mode of positioning is therefore generally referred to as Wide Area DGPS (WADGPS). Such WADGPS services are offered by several multinational companies, who own and operate networks of reference stations, and who use communications satellites to deliver the DGPS corrections to their customers.

The population of continuously operating GPS receivers is growing rapidly. Some are being established to address a single application, e.g. crustal deformation measurement, for atmospheric monitoring, to support precise real-time navigation, etc. Less frequently, GPS networks are designed for multi-functional use. In some countries permanent GPS stations are intended to support local surveying users. However, the GPS reference receivers, because of their high cost, cannot be established in a dense enough configuration to satisfy all user requirements. For example, GPS surveying typically requires receiver separations of less than 30km (and even <20km for high efficiency OTF-kinematic and rapid static positioning techniques). In this paper three scenarios for user positioning which are optimised for permanent GPS reference receiver separations of the order of 50-100km are introduced. (These implementations are referred to here as addressing "medium-range" positioning applications in Surveying and Geodesy.) The techniques all use a linear combination observation model based on the double-differenced phase observations between user receiver (either moving or

stationary) and multiple GPS reference receivers. In such a model the orbit bias and ionospheric delay can be largely eliminated, and the tropospheric delay, multipath and observation noise can be significantly reduced.

One geodetic implementation is a form of "densification" of a network of dual-frequency GPS receivers in which a permanent array of low-cost, single-frequency GPS receivers is deployed to increase the receiver network spatial resolution. Such an implementation would benefit, for example, deformation monitoring applications. Another implementation allows for the use of cost effective operational procedures based on short station occupation times (as in the "rapid static" GPS Surveying technique). The configuration requires two types of GPS receivers: (a) the permanent, continuously operating, dual-frequency, reference receiver network, and (b) the *mobile*, single-frequency receivers which are used to visit a large number of benchmarks for short periods of time. Data from some stations of Japan's Geographical Survey Institute and Taiwan's Institute of Earth Sciences, Academia Sinica, permanent GPS networks have been used to test these algorithms, and the results are presented here.

In addition to such *static* positioning applications in Geodesy, centimetre-level accuracy kinematic GPS positioning with the aid of a sparse continuously operating GPS reference receiver network is also possible. This paper describes how an integrated ambiguity resolution method (with improvements to the real-time stochastic model, new criteria to verify the selected ambiguity set, and a fault detection and adaptive procedure), when used with the proposed multiple GPS reference station observation model, permits the integer ambiguities to be resolved using as little as a single epoch of data, even for receiver separations of several tens of kilometres. Such an enhanced carrier phase-based technique can be used for real-time positioning applications when implemented within a multiple GPS reference station network such as is being currently established in Singapore, in a collaborative R&D project for The University of New South Wales (UNSW) and the Nanyang Technological University (NTU). What makes this network comparatively unique is that the data from the four permanent GPS reference receivers are continuously delivered by dedicated high speed datalinks to a single control station, in a manner analogous to the pseudo-range-based WADGPS systems.

PROGRESS IN CARRIER PHASE-BASED GPS POSITIONING

Since the early 1980s several innovations to GPS carrier phase-tracking hardware, data processing software and field procedures have made precise GPS positioning a comparatively routine activity. The standard mode of precise differential positioning is for one (or more) GPS reference receivers to be located at stations whose coordinates are known, while other GPS user receivers are located at points of interest. All GPS receivers simultaneously track the L-band carrier waves transmitted by the GPS satellites and, following sophisticated data processing, the coordinates of the GPS user receivers are determined relative to those of the reference receiver(s). Carrier phase-based GPS positioning is now extensively used for many geodetic, surveying and precise navigation applications on land, at sea and in the air.

GPS Geodesy: Techniques & Applications

We may distinguish the techniques and applications of *GPS Geodesy* from those of standard *GPS Surveying*. GPS Geodesy typically can be characterised by the following:

(a) ***Hardware*** - receivers and high quality antennas able to make carrier phase and pseudo-range measurements on both L-band frequencies.
(b) ***Software*** - sophisticated computer programs to process the collected GPS data.
(c) ***Field Procedures*** - simultaneously tracking instruments, recording data over a lengthy observation session, with static (non-moving) antennas.
(d) ***Specialist Applications*** - typically those requiring relative positioning accuracy of 0.1-0.01ppm, and stable reference datum connections.

(1ppm is equivalent to a centimetre accuracy for a 10km reference-user receiver separation.) GPS Geodesy techniques have been progressively refined in order to address the *constraints* of carrier phase-based GPS positioning. The two most significant constraints which GPS Geodesy has been able to dramatically overcome are: (a) ppm-level accuracy, and (b) reference-user receiver spacing. Nowadays, as a result of increasing sophistication in measurement modelling, better satellite tracking coverage, a global network of permanent GPS tracking stations belonging to the International GPS Service (IGS), a well-defined global reference frame such as the International Terrestrial Reference System (ITRS), and the availability of dual-frequency GPS instrumentation, significantly higher *relative accuracies of the order of 0.01ppm are commonplace.* Furthermore, such accuracy is achieved for receiver separations ranging from tens of kilometres to several thousands of kilometres, when large numbers of receivers are used simultaneously (some at reference datum sites with ITRS coordinates).

The quintessential *GPS Geodesy technique* is that which has been refined since the early 1980s, and typically used to establish national (and global) geodetic control networks, to measure the movement of tectonic plates, or to monitor the long-term stability of station benchmarks. At the heart of the technique is sophisticated software which can process the data collected by a network of dual-frequency GPS receivers deployed at fixed benchmarks. The quintessential *GPS Geodesy application* is the measurement, or monitoring, of 3-D station velocities to a high accuracy. Station motion being due to, for example, global geodynamics and regional tectonics, ground subsidence as a result of water or oil extraction, or underground mining activities, volcano and hill slope instability, and even movement of engineering structures such as dams, bridges, buildings, open cut mine walls, etc. An experiment can be established in order to measure the rate of deformation, by using GPS to measure the change in length (as well as height difference and orientation) of baselines connecting receivers in a carefully monumented station network. Although GPS techniques which address such applications are generally implemented on a *field-campaign* basis, implying the periodic (often annual) re-survey of a network of station benchmarks, there is increasing interest in the use of *permanent, continuously operating GPS stations*. These are now deployed globally by the IGS, although regional GPS networks have also been established in the USA, Japan, Europe, and many other countries, to address a variety of crustal motion monitoring applications.

The range of GPS Geodesy applications that can be addressed using GPS Geodesy techniques is growing rapidly and includes:

- establishment of national geodetic datums,
- maintenance of the ITRS, including the determination of earth orientation parameters,
- determination of the magnitude and pattern of regional and global crustal motion, both in the horizontal and vertical sense,
- local (ground or structural) deformation monitoring,
- precise determination of the coordinates of benchmarks, or the trajectory of receivers, in a well-defined reference frame, in support of a number of specialist applications,
- precise determination of satellite orbits, and
- atmospheric studies, including water vapour measurement and the monitoring of ionospheric activity.

There is, in addition, intensive R&D into *new* GPS Geodesy techniques which challenge a number of further constraints to precise carrier phase-based positioning. These include techniques that are applicable when the GPS antennas are in motion (as in so-called "kinematic geodesy"), the length of observation sessions are significantly shortened, the positioning results need to be determined in "real-" or "near-real-time", the density of points to be surveyed is very high, and so on. Some of these techniques require multiple GPS reference stations.

High Productivity GPS Surveying

GPS Surveying requires a minimum of two GPS receivers, and until recently the cost of such technology has been quite high. What has made GPS Surveying technology expensive, apart from the high capital cost of the instrumentation, was the inflexible field procedures which required that the antennas be stationary (over a groundmark) for periods of up to several hours, as well as the complex post-mission data processing.

Over the last half decade, however, considerable R&D has been invested by instrument manufacturers to make the GPS Surveying technology more attractive. That is, if the antenna could be moving (that is, the so-called "kinematic positioning" mode), then new applications for the GPS technology could be addressed. If the length of time required to collect phase data for a reliable solution could be shortened, and if the results were available immediately (that is, the "real-time" positioning mode), then GPS carrier phase-based positioning *productivity* would significantly improve and the technology would be attractive for many more precise positioning applications. GPS Surveying is now a mature technology, capable of delivering relative accuracies of the order of a few parts per million (ppm) for reference-user receiver separations up to 10-20 kilometres using commercial-off-the-shelf (COTS) GPS products.

Present COTS "real-time kinematic" (RTK) GPS systems are: (a) able to be used in the kinematic positioning mode, (b) require comparatively short observation times, and (c) are capable of real-time operation (when provision is made for a communications link between the two GPS receivers). At the heart of all such high accuracy GPS systems is the Ambiguity Resolution (AR) algorithm. The challenge has been (and continues to be): *how to carry out AR quickly, reliably and with a minimum of constraints?* Addressing the constraints of AR in RTK systems is crucial if GPS is to be used for time-critical applications such as machine control, GPS-guided excavations, precision farming, automated container port operations, and so on.

If GPS signals were continuously tracked and loss-of-signal-lock never occurred, the integer ambiguities determined at the beginning of a survey would be valid for the whole period that GPS was being used. However, the GPS satellite signals can be shaded (e.g. due to buildings in "urban canyon" environments, or when the receiver passes under a bridge or through a tunnel), in which case the ambiguity values are "lost" and must be redetermined. The length of this "time-to-AR" may range from several tens of seconds up to a few minutes with present GPS COTS systems, *but only when the reference-to-mobile-user receiver distance is less than about 20km.* During this "re-initialisation" period centimetre accuracy positioning is not possible, and hence there is "dead" time until sufficient data has been collected for AR. If interruptions to the GPS signals occur repeatedly, then ambiguity re-initialisation is at the very least an irritation, and at worse a significant weakness of GPS COTS carrier phase-based systems. In addition, the longer the period of tracking required to ensure reliable "on-the-fly" AR (OTF-AR), the greater the risk that cycle slips will occur during this crucial (re-)initialisation period. (These shortcomings are also present in any system based on data post-processing as well, however implementing an RTK system is more challenging.)

The algorithm improvements that can address the baseline length constraint, and shorten the "time-to-AR" to just one epoch of data, for kinematic applications, rely on new multiple GPS reference station implementations and associated improvements to data processing algorithms.

ADDRESSING THE CONSTRAINTS FOR MEDIUM-RANGE STATIC & KINEMATIC GPS POSITIONING

GPS Geodesy has provided considerable impetus for: (a) improvements in instrumentation, (b) the development of the IGS infrastructure, and (c) the ITRS reference frame. GPS Surveying is (almost) capable of instantaneous centimetre-level accuracy positioning of moving antennas *if certain conditions are met*. The following constraints can be addressed via multiple GPS reference station implementations and associated improvements in data processing algorithms:

(1) Insistence on site occupations of several days for geodetic applications such as control network densification, and surveys undertaken for pre- and post-seismic network distortion measurement.

(2) Insistence on the use of high cost, dual-frequency instrumentation for geodetic applications such as deformation monitoring.

(3) Insistence on short reference-to-mobile-user receiver separations for kinematic applications using OTF-AR techniques, for both real-time and post-mission implementations.

In the scenarios described in this paper, the implication is that medium-range positioning (defined here as involving baseline lengths of the order of 50-100km) is carried out in such a manner that *baseline length dependent biases are mitigated*. For geodetic applications ((1) and (2) above) this means that sub-ppm accuracy can still be delivered even when the instrumentation (single-frequency receivers) and/or field procedures (short observation periods) deviate from those traditionally insisted upon. In the case of precise kinematic applications ((3) above), this implies that OTF-AR is just as easy and reliable as over short baselines.

The most important baseline length dependent biases are satellite orbit, ionospheric and tropospheric biases. Multiple GPS reference stations surrounding the area of survey serve to generate empirical correction terms for the user receiver in a manner analogous to Wide Area DGPS systems. A linear combination model has been proposed by Han (1997), Han & Rizos (1997), which can account for orbit bias and ionospheric delay, as well as mitigate tropospheric delay, multipath and measurement noise across the network. The basis of this approach is that the data from the reference station network can be used to develop corrections to the double-differenced carrier phase data formed between a mobile receiver and a *single reference receiver*, hence making it possible to implement it within COTS single-reference-receiver static and kinematic data processing software.

Geodetic Procedures Based on Multiple Reference Receivers

Four data sets were used in these experiments, three provided by the permanent GPS network of Japan's Geographical Survey Institute (GSI), and one from Taiwan's Institute of Earth Sciences - Academia Sinica (IESAS). These are located in Figure 1 (in Hokkaido, Tokai and Kyushu, in Japan, and on the east coast of Taiwan). For each geographical location, seven days of data (dual-frequency data with sample rate 30 seconds, 24 hour files), from four different seasons in 1997 (day-of-year: 001-007, 091-097, 181-187 and 271-277), from a total of about 30 stations, were used in the analysis. All station coordinates were determined using scientific GPS software in the ITRS, with respect to three IGS stations. These coordinates would be used for comparisons with alternative processing strategies proposed by the authors.

The aim was to test the performance of the linear combination observation model based on double-differenced observations between a single-frequency (user) receiver and multiple GPS reference receivers. A variety of reference receiver network configurations, regions and seasons were chosen. The user-reference receiver baseline lengths ranged from about 24km to over 90km. Two experi-

ments were conducted on all the data sets: (a) a test of rapid static, single-frequency surveys, and (b) a test of continuous GPS network containing both dual- and single-frequency receivers. The results are summarised below.

All baselines are first determined using dual frequency ionosphere-free combinations over seven days are considered and the results are considered true values for the according 7-day session. All GPS data sets are then split into files of 240 epochs (2hr length), resulting in 12 files per day per station (Rizos, et al, 1998). The 84 results for each baseline were derived and subsequently 84 differences can be derived. The mean value of the 84 differences and the standard deviation of the 2-hour session results are plotted in the left hand side of Figure 2 versus the baseline for four regions in the Autumn campaign. The solid circles (3 baselines), white squares (6 baselines), solid triangles (4 baselines) and white diamonds (1 baseline) represent the different networks: Hokkaido, Tokai, Kyushu, IESAS. The mean offsets of all 14 baselines are 2.8mm, 4.5mm, 14.1mm, 3.2mm for the latitude, longitude, height and baseline length components, respectively. It can be seen that the biases and standard deviation of the baseline results are almost baseline independent using the corrections from multiple reference stations. The right hand side of Figure 2 summaries all results over the four seasons. It notes that the variation of the 2 hour results deriving during the Summer period is larger than for other seasons. Further investigations on this issue are being undertaken.

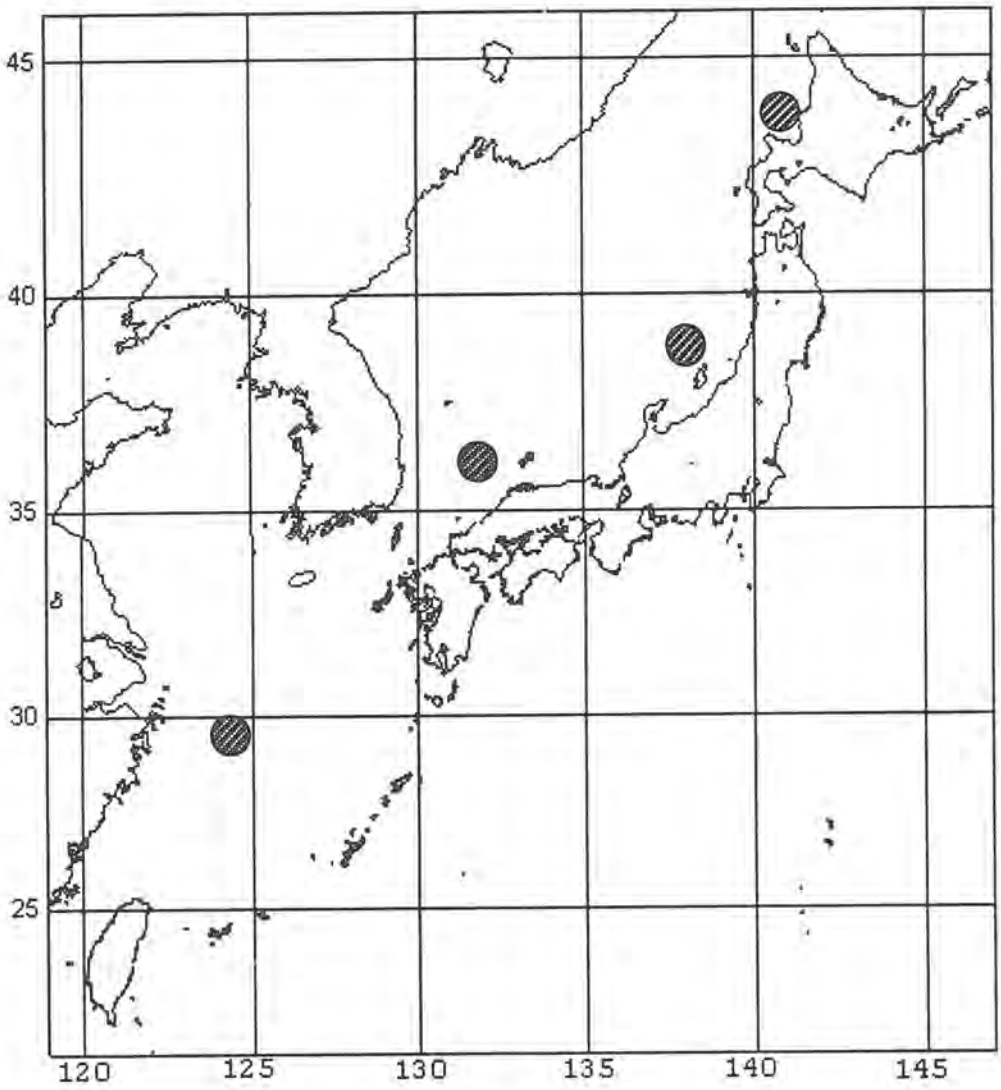


Figure 1. Four test areas of permanent GPS networks in Japan and Taiwan.

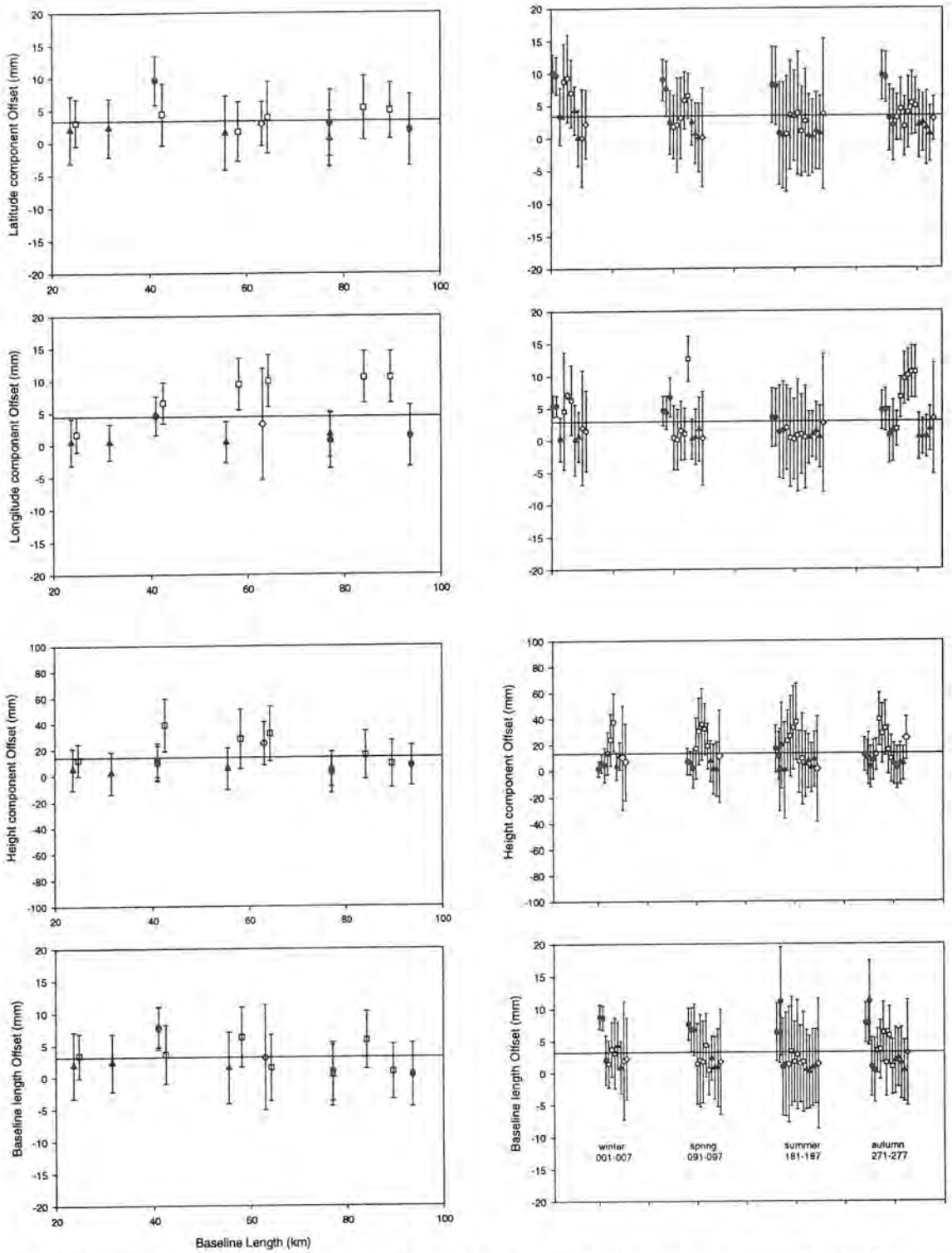


Figure 2. Mean baseline offsets (relative to the 7 day ionosphere-free results) and standard deviations of the 2 hour baselines. The results are shown for various user-reference receiver distances, in the different networks, in the lefthand column plots (the x-axis is distance in km) in the Autumn campaign, and in the righthand column plots over the four seasons. The symbols, ●, □, ▲, ◆ represent the different networks: Hokkaido, Tokai, Kyushu, IESAS.

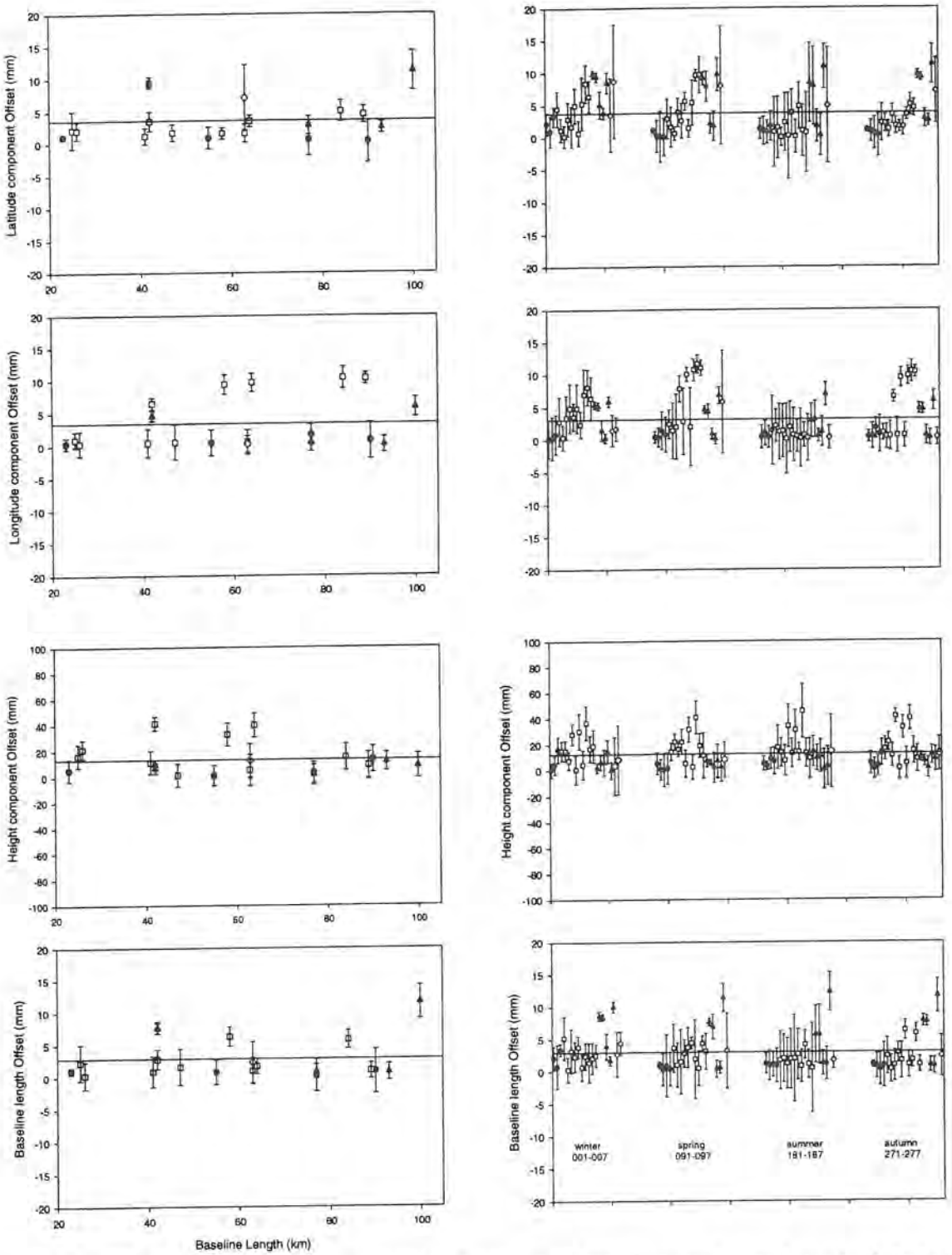


Figure 3. Mean baseline offsets (relative to the 7 day ionosphere-free results) and standard deviations of the 24 hour baselines. The results are shown for various user-reference receiver distances, in the different networks, in the lefthand column plots (the x-axis is distance in km) in the Autumn campaign, and in the righthand column plots over the four seasons. The symbols, ●, □, ▲, ◆ represent the different networks: Hokkaido, Tokai, Kyushu, IESAS.

In summary, it appears that 1cm accuracy horizontal component and 3cm height determination is possible with single-frequency, rapid static techniques (2hr sessions) under certain conditions. Although these are the subject of ongoing investigation, it is obvious that, for example, larger errors (especially in the height component) are evident when the user receiver elevation is different from the surrounding network receivers (see Tokai network). It therefore appears feasible therefore that sub-ppm accuracy (ppm: 1cm accuracy over 10km receiver separation) is achievable using low-cost receivers and comparatively short observation sessions, for user-reference receiver separations of several tens of kilometres.

The second scenario simulates a permanent network configuration that may be used for ground or structural deformation monitoring, based on a sparse network of 3 dual-frequency receivers surrounding a denser, inner network of low-cost, single-frequency receivers. The same four data sets were used for the tests with the exception that 24hr files were processed in place of the rapid static (2hr) results referred to above. Figure 3 contains the same information as Figure 2, and summarises the results of mean baseline offsets relative to the 7 day ionosphere-free results and standard deviations of the 24 hour baselines. The most noticeable difference is not in the overall accuracy (as indicated by the mean values), but in the standard deviations, which are lower. The conclusion that can be drawn is that when single-frequency data is processed in an optimal way, such that the reference receiver network is used to generate corrections to the double-differenced data between a user receiver and one of the reference receivers, the accuracy of the derived coordinate results are very high and may be adequate for addressing certain geodetic applications such as the monitoring of local ground deformation phenomena.

GPS Surveying Based on Multiple Reference Receivers

A test was carried out on 14 December 1996 (Han & Rizos, 1997). A permanent GPS station at The University of New South Wales (UNSW) was selected as one of the reference stations. The other two reference stations were located at Stanwell Park, to the south of Sydney, and at Springwood, to the west of Sydney. The mobile receiver was mounted on a car and the experiment started at the side of a highway, 31.44km, 34.11km and 46.5km distant from the UNSW receiver, Springwood receiver and Stanwell Park receiver, respectively. During the test the car-mounted receiver travelled along the highway, and then back to nearly the same point as the start point, collecting a total of 1903 epochs of data. Two algorithmic innovations were tested together: (a) the multiple GPS reference station methodology (Han, 1997), and (b) an integrated OTF-AR algorithm. The UNSW OTF-AR algorithm can be used with a single epoch of dual-frequency, carrier phase and pseudo-range data (Han, 1996). In addition to the standard AR search technique, it consists of three further refinements: (1) new criteria to validate the integer ambiguity set, (2) a real-time stochastic model, and (3) an adaptive procedure. The results in Table 1 have been separated to illustrate the improvements from applying these three steps. Firstly, the integrated method with step (1) is used, and the results displayed in row 2 of Table 1. Then, the integrated method with steps (1) and (2) is used, and the results displayed in row 3 of Table 1. Finally, the three-step UNSW OTF-AR algorithm (with multiple reference station corrections) is applied, and the results displayed in row 4 of Table 1.

This instantaneous OTF-AR methodology illustrates one of the benefits of using multiple GPS reference stations. Such OTF-AR performance, for baselines longer than 30km, would not be possible using a single GPS reference station. Conversely, multiple epoch OTF-AR could be a routine operation for baselines up to 100km in length were there a GPS reference network to surround the survey area. Finally, it is possible to implement such algorithms in real-time if the reference station network was *integrated* via high speed data links.

The implications of such medium-range positioning performance is worthy of comment. By overcoming the short-range AR constraint in this way the logistical costs associated with operating a close-by GPS reference station are reduced. Furthermore, if the operation of the GPS reference network were the responsibility of an agency (public or private), the user would be relieved of a considerable burden, and the costs of GPS surveys would be reduced further. The establishment of an ap-

propriate multi-reference station infrastructure across a major city could therefore be a significant boon for all carrier phase-based GPS positioning applications, including challenging RTK implementations.

Table 1. Single epoch ambiguity resolution for medium-range, multi-reference station, GPS kinematic positioning, using the integrated UNSW OTF-AR methodology.

	Total Epochs	Fix Ambiguities		
		Correct	Wrong	Reject
Integrated OTF-AR with (1)	1903	1840	0	63
Integrated OTF-AR with (1, 2)	1903	1849	0	54
Integrated OTF-AR with (1, 2, 3)	1903	1903	0	0

CONCLUDING REMARKS

Considerable R&D still needs to be undertaken in order to determine, for example, how closely spaced the network of reference stations should be in order to derive bias corrections with sufficient accuracy to resolve integer ambiguities, as well as investigate new configurations of precise static and kinematic GPS positioning systems. The critical resource for this research is a multiple GPS reference station infrastructure such as that being established to service researchers and users in the Republic of Singapore (see Figure 4). The project involves the installation of four permanent GPS reference receivers around the island, the leasing of dedicated telephone lines linking the receivers to a central computer server located at the Nanyang Technological University (NTU), and the development of software to manage day-to-day operations and the data flow within the network. Such an integrated facility will be crucial for testing possible real-time implementations of the "multiple GPS reference station approach" for GPS Surveying and GPS Geodesy.

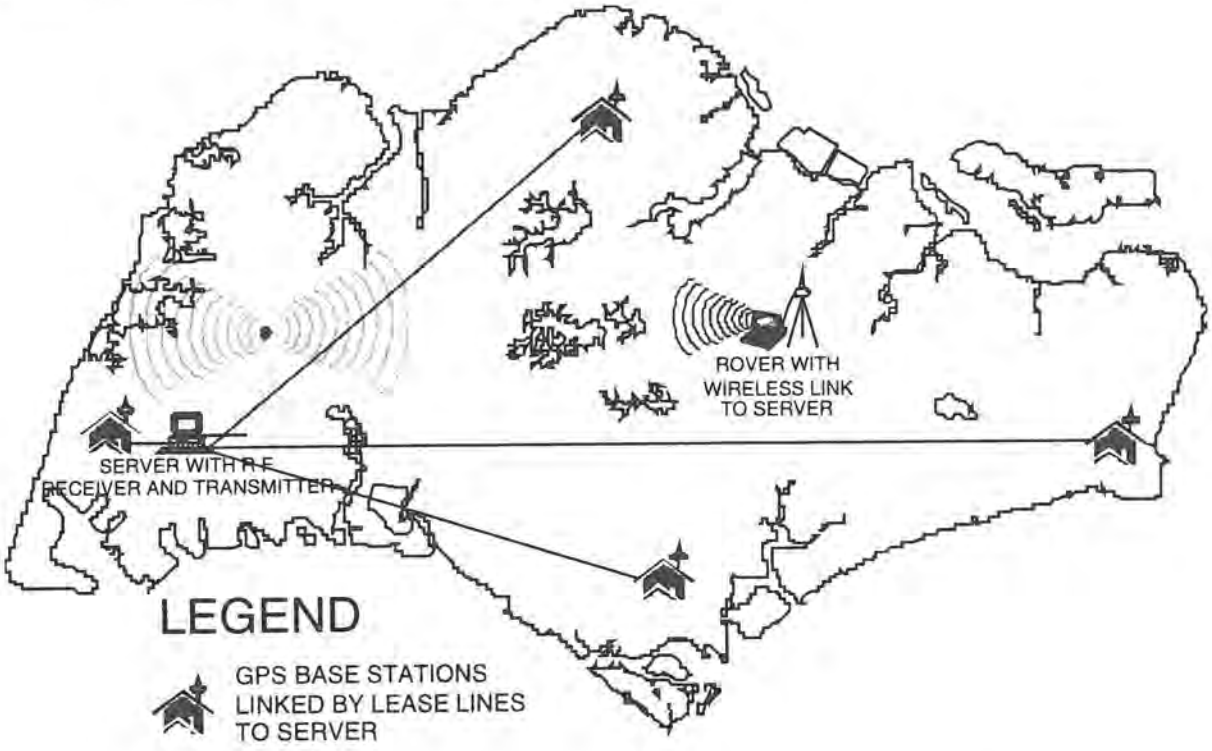


Figure 4. The integrated multiple GPS reference station infrastructure in Singapore (station locations only approximate).

Over the last decade, both the cost-effectiveness of GPS Surveying techniques and the accuracy and reliability of GPS Geodesy techniques has improved considerably. However, such performance has implied high cost, dual-frequency instrumentation and the use of rather rigid (and constrained) operational procedures. There are several new algorithms and implementation strategies that take advantage of multiple GPS reference networks which can overcome some of the constraints of medium-range positioning, using lower cost receiver hardware and/or less rigid field procedures, without "trading off" too much in performance (defined in terms of accuracy, time-to-survey, time-to-AR, reliability of AR, etc.). In this paper the authors have briefly described:

(1) The manner in which a "rapid static" surveying technique may be implemented using low-cost mobile GPS receivers to address "near-real-time" geodetic applications such as rapid surveys of large numbers of benchmarks in response to pre- or post-seismic activity.
(2) A scenario in which a dense network of single-frequency GPS receivers, in combination with a sparse network of permanent dual-frequency receivers, can be used cost-effectively for deformation monitoring applications.
(3) The scenario of a user owning only a single GPS receiver, and applying the "multiple GPS reference station approach", within an enhanced OTF-AR system that addresses the short-range AR constraint of present COTS systems.

Feasibility tests have been carried out at a number of network locations. However, over the next few years, an "open air laboratory" such as the NTU-UNSW GPS network in Singapore will permit engineering challenges (such as real-time implementations) to be addressed, as well as aid in identifying potential user applications of, and services for, such integrated networks.

ACKNOWLEDGEMENTS

The third author is supported by a study grant from the Republic of Taiwan's Institute of Earth Sciences, Academia Sinica, to undertake research towards a PhD at the School of Geomatic Engineering, The University of New South Wales.

REFERENCES

Han, S. (1996), Quality control issues relating to ambiguity resolution for real-time GPS kinematic positioning, Proc. *9th Int. Tech. Meeting of The Satellite Division of The U.S. Institute of Navigation*, Kansas City, Missouri, 17-20 September, 1419-1430.

Han, S. (1997), *Carrier Phase-Based Long-Range GPS Kinematic Positioning*. PhD dissertation, Unisurv rept. S-49, School of Geomatic Eng., UNSW, Sydney, 185pp.

Han, S. & C. Rizos (1997), An instantaneous ambiguity resolution technique for medium-range GPS kinematic positioning, Proc. *10th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation*, Kansas City, Missouri, 16-19 September, 1789-1800.

Rizos, C., S. Han & H.-Y. Chen (1998), Carrier phase-based, medium-range, GPS rapid static positioning in support of geodetic applications: algorithms and experimental results, to be published in proc. of *Spatial Information Science & Technology (SIST'98)*, Wuhan Technical University of Surveying & Mapping, Wuhan, P.R. China, 13-16 December.

The analysis of the Neumann and oblique derivative problem. Weak theory

Youri Rozanov and Fernando Sansò

Abstract:
In this review paper the simple Neumann and oblique derivative problem is formulated for an exterior domain and mapped by the Kelvin-Raleigh transform, to an internal domain. The weak formulation of the two problems is subsequently studied and standard theorems of existence, uniqueness and well-posedness are proved.
The conditions of validity for such theorems have a clear interpretation for the geometric point of view.
An extensive Appendix, mostly without proofs, provides the relevant material on theory of distributions and Sobolev spaces.

1 Motivation and formulation of the problem

In geodesy some of the most fundamental problems of the gravity field determination from boundary observations are translated into exterior boundary value problems (BVP) for the Lapalce or Poisson equation (cfr. [Sansò, 1995], [Sansò, 1997]).
After suitable linearization and reductions-simplifications of various kinds we finally come out with a problem that can typically be formulated as follows:
Given a simply connected bounded open domain B with boundary S and the exterior open domain Ω, given some known distribution f in Ω and a boundary datum g, to find a function (potential) u in Ω such that

$$\Delta u = f \ in \ \ \Omega \tag{1.1}$$

$$\underline{e} \cdot \nabla u + bu = g \ on \ \ S \tag{1.2}$$

$$u = 0\left(\frac{1}{r}\right) \quad for \ \ r \to \infty \tag{1.3}$$

Please note that we have been purposly ambiguous in denoting f as a distribution as it can be legitimately interpreted in both senses, physically as a mass distribution outside S (this happens when classical geodetic reductions are applied so that S lies partly inside the masses), or mathematically, in L. Schwarz sense. It has to be remarked that the physical situation in geodesy is such that f has a bounded support in Ω although we shall go beyond this hypothesis for the sake of completeness in the analysis. As a byproduct of this remark we shall not insist on the condition (1.3), which derives from the requirement that u is a regular harmonic function outside a sphere of sufficiently large radius, but we shall define in the next paragraph conditions of regularity at infinity suitable for the functional spaces we are going to work with.
As for the boundary condition (1.2) we remark that, excluding the case of mixed BVP's, $\underline{e}$ can be taken as a unit vector field on S. Typically in geodesy $\underline{e}$ (cfr. Fig. 1.1) is the direction of the normal gravity $\underline{\gamma}$, or its opposite. In turn $\underline{\gamma}$ is always directed fairly close to the radial direction $\underline{e}_r$, while the outer normal $\underline{\nu}$ to S may be very distinct to $\underline{e}$, although this happens only on a small portion of the surface. In any event, we shall make the quite reasonable assumption that

$$\cos\alpha = \underline{e} \cdot \ \underline{\nu} \geq \alpha > 0 \ (\text{on } S) \ ; \tag{1.4}$$

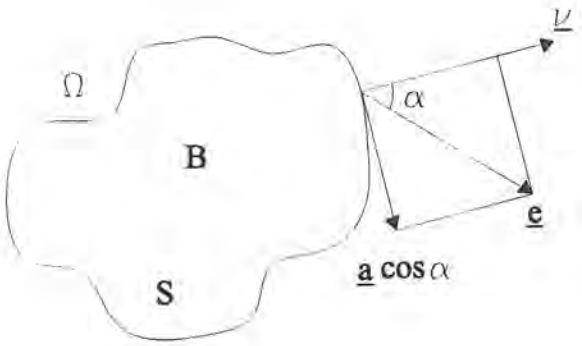


Figure 1.1: The geometry of BVP's analyzed.

this qualifies mathematically (1.1), (1.2), (1.3) as a regular oblique derivative problem. In this evenience, by dividing (1.2) by $\cos\alpha$ and rearranging the symbols in an obvious way, we can write

$$\underline{\nu} \cdot \nabla u + \underline{a} \cdot \nabla u + bu = g \tag{1.5}$$

or

$$\frac{\partial u}{\partial \nu} + \underline{a} \cdot \nabla_t u + bu = g \tag{1.6}$$

with $\underline{a}$ tangent to S,

$$\underline{a} \cdot \underline{\nu} = 0, (\text{on } S). \tag{1.7}$$

and $\nabla_t = \nabla - \underline{\nu}\partial_\nu$, the tangent component of the gradient. Let us remark that S has to display some regularity and in this paper to fix the ideas we shall accept that S is a $C^{2,\lambda}$ surface, i.e. it has λ Holder continuous second derivatives in local coordinates.
In (1.6) $\underline{a}$ is small in the average (apart from mountainous areas) and $\underline{a} \cdot \nabla_t$ can be considered as a perturbation with respect to the main operator ∂_ν In one important instance, when gravity anomalies data are reduced to the ellipsoid for the determination of the geoid, we have identically $\underline{a} \equiv 0$.
Finally, the term in b might or might not be present, depending on the problem considered: this is indeed not irrelevant to the mathematical analysis, since the uniqueness or non-uniqueness of the solution does depend on the sign of b. For instance, the simple Molodensky problem with boundary operator $B = \left(\frac{\partial}{\partial r} + \frac{2}{r}\right)$, is well known to have a null space of dimension 3. On the other hand, the operator $b \cdot u$ is much milder than $\partial_\nu \cdot u$ so that (1.6) can typically be turned into a Fredholm type equation, once the corresponding problem with $b = 0$ has been analyzed. Furthermore, $b = 0$ does correspond to the (linearized) fixed boundary gravimetric problem (cfr. [Sansò, 1997]), which is becoming a realistic problem with nowadays GPS observations, and which geodesy shares with another geoscience: geomagnetism.
Concluding, we will be considering the two problems

a) The exterior Neumann problem

$$\begin{cases} \Delta u = f & \text{in } \Omega \\ \dfrac{\partial u}{\partial \nu} & \text{on } S \\ u \to 0 & \text{at } \infty \end{cases} \tag{1.8}$$

b) The exterior regular oblique derivative problem

$$\begin{cases} \Delta u = f & \text{in } \Omega \\ \dfrac{\partial u}{\partial \nu} + \underline{a} \cdot \nabla_t u = g & \text{on } S \\ u \to 0 & \text{at } \infty \, . \end{cases} \tag{1.9}$$

In this paper we shall present the theory of existence, uniqueness and stability of the solution of the two problems in the framework of the weak concept, so called because the differential operators have to be interpreted in distribution sense. In a forthcoming paper we shall present the strong theory too, or the so-called analysis of the regularization problem, including its extension to the corresponding stochastic problems, where u has to be interpreted as a generalized random field (cfr. [Sansò, 1997], [Sansò, 1995], [Sansò, Venuti, 1998]). Although there is a very large mathematical litterature for problems like (1.8), (1.9) and although the results we shall derive are not surprising, yet in the methods used for their proof and in particular in the use of a generalized Ladyzenkaya identity for the regularization, there is some new material analysis. Furthermore, the precise formulation of the condition for the existence and uniqueness of the solution of the pure oblique derivative problem is sufficiently simple to be interpreted in a geodetic sense. On the other hand, a number of remarks that could be considerably shortened for mathematicians, are nevertheless included into the paper to make it self-consistent for readers from a geodetic environment.

2 The Kelvin-Raleigh transform

This transform, also called the inverse radii transform, is useful here because it maps (1.8), (1.9) into BVP problems for an interior domain Ω, which is in this case bounded.
The transform is simultaneously a transformation of coordinates and of the unknown function: more precisely, assume that B is such as to cover the unit ball

$$B \supset B_1 \equiv \{r \leq 1\} \, ,$$

so that if in this paragraph we call s the "exterior" radial variable, we have

$$s_P > 1 \, , \qquad \forall P \in \Omega \, ;$$

then we define a new variable r and a new unknown function (potential) v, as

$$s = \frac{1}{r}(s > 1), \quad r = \frac{1}{s} \quad (r < 1) \tag{2.1}$$

$$v = (r, \sigma) = \frac{1}{r} u(\frac{1}{r}, \sigma) = s \; u(s, \sigma) \, . \tag{2.2}$$

As it is obvious, by (2.1) Ω is mapped into a set $\Omega' \subset B_1$.

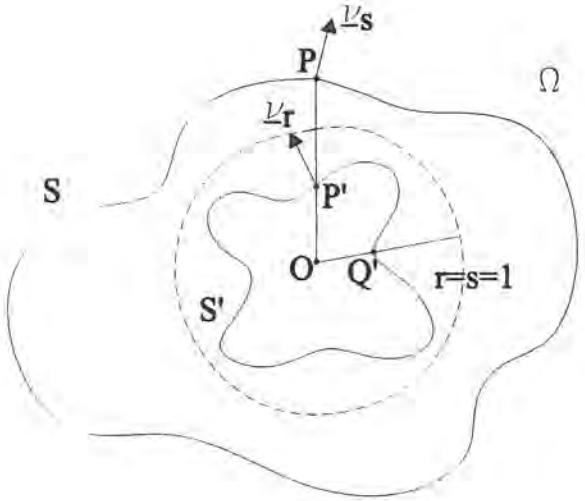


Figure 2.1: The geometry of the Kelvin-Raleigh transform.

so we have

$$\begin{aligned} &\Omega \text{ exterior to } S \\ &\Omega' \text{ interior to } S' \end{aligned}$$

Although not necessary, one can think that S and S' have (at least locally) equations

$$\begin{aligned} S &\rightarrow s = S(\sigma) \\ S' &\rightarrow r = R(\sigma) = \frac{1}{S(\sigma)} . \end{aligned} \tag{2.3}$$

A simple computation then shows that the two *exterior* normals $\underline{\nu}_s, \underline{\nu}_r$ are symmetrically placed with respect to the radial unit vector $\underline{e}_r = \underline{e}_s$. A straightforward computation shows that

$$\Delta_r v \equiv s^5 \Delta_s u \equiv s^5 f \ , \tag{2.4}$$

$$\frac{\partial v}{\partial \nu} + \frac{v}{\sqrt{R^2 + R_\vartheta^2 + \frac{1}{\sin^2\vartheta} R_\lambda^2}} = -S^3 \frac{\partial u}{\partial \nu_s} = -S^3 g \ . \tag{2.5}$$

Accordingly we can map problem (1.8) into a problem of the form

$$\begin{cases} \Delta v = F & \text{in } \Omega' \\ \dfrac{\partial u}{\partial \nu} + bv = G & \text{on } S' \end{cases} \tag{2.6}$$

where

$$F(r,\sigma) = \frac{1}{r^5} f\left(\frac{1}{r}, \sigma\right) \ ; \ G(\sigma) \equiv -S^3(\sigma) g(\sigma) \tag{2.7}$$

and

$$\bar{b} \geq b(\sigma) = \frac{1}{\sqrt{R^2(\sigma) + R_\vartheta^2 + \frac{R_\lambda^2}{\sin_\vartheta^2}}} \geq b_0 = 1 \,. \tag{2.8}$$

As for the oblique derivative problem (1.9) one sees that it is transformed into

$$\begin{cases} \Delta v = F & \text{in } \Omega' \\ \dfrac{\partial v}{\partial \nu} + \underline{a} \cdot \nabla_t v + bv = G & \text{on } S' \end{cases} \tag{2.9}$$

where this time

$$\underline{a} = \frac{\underline{e}}{\underline{e} \cdot \underline{\nu}_r} - \underline{\nu}_r \,. \tag{2.10}$$

It is an important remark that if $\underline{e}$ is directed close to the radial direction, then $\underline{a}_r$ given by (2.10) and

$$\underline{a}_s = \frac{\underline{e}}{\underline{e} \cdot \underline{\nu}_s} - \underline{\nu}_s$$

must be close one another in size. The situation is illustrated in Fig. 2.2 when $\underline{e} \equiv \underline{e}_r$

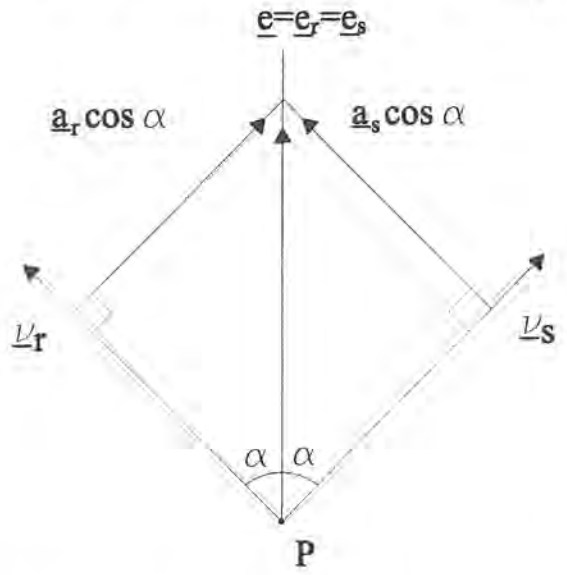


Figure 2.2: Geometry of the oblique derivative problem when $\underline{e} = \underline{e}_r = \underline{e}_s$.

Accordingly, when $\underline{a} \cdot \nabla_t$ can be considered as a perturbation in (1.9), the same is true in (2.9) and vice versa.

Remark 2.1: now that we have established the correspondence between external and internal problems at the level of notation we shall call again B instead of Ω' the internal domain and S instead of S' its boundary.

Remark 2.2: with the inverse radii transform there is a certain modification of the functional properties of the known terms. However, since by hypothesis $S(\sigma) \in C^{2+\lambda}$ we do not have, in the context of this paper, significant differences between g and G. On the contrary, since $F = s^5 f$ we see that $F \in L^2(\Omega')$, considering that $d\Omega' = \dfrac{d\Omega}{s^6}$, implies

$$\int_{\Omega'} F^2 d\Omega' = \int_{\Omega} s^4 f^2 d\Omega < +\infty \,,$$

which imposes a well-defined asymptotic constraint on f.

3 Standard weak solutions of the Neumann and oblique derivative problems

We start from the Neumann problem

$$\begin{cases} \Delta v = F & \text{in } B \\ \dfrac{\partial v}{\partial \nu} + bv = G & \text{on } S \\ (\bar{b} \geq b|_S \geq b_0 > 0) \ . \end{cases} \tag{3.1}$$

Since we want to look for a solution less regular than H^2 in B we have to find a way to write (3.1) in a form that is equivalent to it for classical solutions, $v \in H^2$, but which involves only first order derivatives, $v \in H^1$. This form is obtained by an application of a Green's identity, namely $\forall \varphi \in \mathfrak{R}_B \mathcal{D}$

$$\begin{aligned} \int_B \nabla v \cdot \nabla \varphi dB = \int_S \frac{\partial v}{\partial \nu} \varphi dS - \int_B \Delta v \varphi dB = \\ = \ - \int_S bv\varphi dS + \int_S G\varphi dS - \int_B F\varphi dB \end{aligned} \tag{3.2}$$

Let us recall that $\forall \varphi \in \mathfrak{R}_B \mathcal{D}$ means that $\forall \varphi \in \mathcal{D}$ (C^∞ functions with compact support) we take its restriction to B (open). Obviously if we take $v \in H^2$, then $F \in L^2(B)$ and $G \in H^{3/2}(S) \subset H^{1/2}(S)$ (cfr. Appendix) so that each single term of (3.2) is finite and meaningful $\forall \varphi \in \mathfrak{R}_B \mathcal{D}$; it is a basic issue that each term in (3.2) can be extended by continuity to the case that $v, \varphi \in H^1(B)$, $F \in [H^1(B)]', G \in H^{-1/2}(S)$.

Lemma 3.1: let $v, \varphi \in H^1(B)$, $F \in [H^1(B)]', G \in H^{-1/2}(S)$ then, interpreting the integrals by continuity as limits of sequences of regular functions,

$$\left| \int_B \nabla v \cdot \nabla \varphi dB \right| \leq \|v\|_{H^1} \cdot \|\varphi\|_{H^1} \tag{3.3}$$

$$\left| \int_S bv\varphi dS \right| \leq \text{const} \cdot \|v\|_{H^1(B)} \cdot \|\varphi\|_{H^1(B)} \tag{3.4}$$

$$\left| \int_S G\varphi dS \right| \leq \text{const} \cdot \|G\|_{H^{-1/2}(S)} \cdot \|\varphi\|_{H^{1/2}(S)} \tag{3.5}$$

$$\left| \int_B F\varphi dS \right| \leq \text{const} \cdot \|F\|_{(H^1)'} \cdot \|\varphi\|_{H^1(B)} \tag{3.6}$$

□ (3.3) derives from Schwarz inequality and observing that $\{\int_B \nabla v^2 dB\}^{1/2} \leq |v|_{H^1}$; (3.4) is again given by Schwarz inequality, recalling that $b \leq \bar{b}$, and trace theorems (see Appendix)

$$\begin{aligned} \left| \int_s bv\varphi dS \right| \leq \text{const} \, \|v\|_{L^2(S)} \, \|\varphi\|_{L^2(S)} \leq \\ \leq \text{const} \, \|v\|_{H^{1/2}(S)} \, \|\varphi\|_{H^{1/2}(S)} \leq \text{const} \, \|v\|_{H^1(B)} \, \|\varphi\|_{H^1(B)} \end{aligned}$$

(3.5) is a direct application of Lemma A.7 and of Sobolev trace Theorem. (3.6) is a direct application of Lemma A.8. □

From Lemma 3.1 and Riesz representation theorem, of the dual of a Hilbert space with the same space, we see that (3.1) can be transformed into a simple equation from $H^1(B)$ into $H^1(B)$; in

fact we can claim that

$$A(v,\varphi) = \int_B \nabla v \cdot \nabla\varphi dB + \int_S bv\varphi dS \equiv < Av, \varphi >_{H^1(B)}, \tag{3.7}$$

$$\int_S G\varphi dS = < \Gamma G, \varphi >_{H'} \tag{3.8}$$

$$\int_B F\varphi dB = < CF, \varphi >_{H'} \tag{3.9}$$

with, respectively

$$\|Av\|_{H^1(B)} \leq \overline{A}\,\|v\|_{H^1(B)} \tag{3.10}$$

$$\|\Gamma G\|_{H^1(B)} \leq \overline{\Gamma}\,\|G\|_{H^{-1/2}(S)} \tag{3.11}$$

$$\|CF\|_{H^1(B)} \leq \overline{C}\,\|F\|_{[H^1(B)]'} \ . \tag{3.12}$$

Accordingly (3.2), extended to the whole $H^1(B)$, becomes

$$\langle Av, \varphi \rangle_{H^1(B)} = \langle \Gamma G - CF, \varphi \rangle_{H^1(B)} \quad \forall \varphi \in H^1(B)$$

or

$$Av = \Gamma G - CF \ , \ v \in H^1(B) \ . \tag{3.13}$$

Remark 3.1: from the form of (3.7) one immediately realizes that A is a selfadjoint operator in $H^1(B)$, in fact by symmetry

$$\langle Av, \varphi \rangle_{H'} = A(v,\varphi) = A(\varphi, v) = \langle A\varphi, v \rangle_{H^1} = \langle v, A\varphi \rangle_{H^1} \ . \tag{3.14}$$

Since for sure $\Gamma G - CF \in H^1(B)$ the study of equation (3.13) is reconducted to the question whether A is an isomorphism (i.e. continuous, invertible and with continuous inverse) of $H^1(B)$ into itself.
To answer to that one can use a very basic lemma that we formulate here as Lemma 3.2 and prove in the Appendix as Lemma A.10.
Lemma 3.2: $\forall v \in H^1(B)$ the following inequality holds

$$\int_B v^2(P)dB \leq \text{const} \left\{ \int_B \nabla v^2 dB + \int_S v^2(P) dS_P \right\} \ . \tag{3.15}$$

Remark 3.2: from Lemma 3.2 we basically see that $A(v,v)$ is the square of an equivalent norm in $H^1(B)$. In fact from (3.10) we already know that

$$A(v,v) = \langle Av, v \rangle_{H^1(B)} \leq \overline{A}\,\|v\|^2_{H^1(B)} \tag{3.16}$$

while from (3.15) we clearly get

$$\begin{aligned} \|v\|^2_{H^1(B)} &= \int_B \nabla v^2 dB + \int_B v^2 dB \leq \widetilde{A} \left\{ \int_B \nabla v^2 dB + \int_S v^2 dS \right\} \leq \\ &\leq \widetilde{A}' \left\{ \int_B \nabla v^2 + \int_S bv^2 dS \right\} = \widetilde{A}' A(v,v) \ . \end{aligned} \tag{3.17}$$

so we have

$$\frac{1}{\widetilde{A}'} \|v\|^2_{H^1(B)} \leq A(v,v) \leq \overline{A}\,\|v\|^2_{H^1(B)} \ . \tag{3.18}$$

This of course proves at once that A is an isomorphism of $H^1(B)$ onto itself. In fact by (3.17) A is an invertible operator ($Av = 0 \rightarrow \|v\| = 0 \rightarrow v = 0$) and its range is closed in H^2. Moreover its range is dense in H^1, because by selfadjointness if $u \in H^1$ is such that

$$\forall v \in H^1, 0 = \langle Av, U \rangle = \langle v, Au \rangle$$

we must have as well

$$Au = 0 \quad \Rightarrow \quad u = 0.$$

Then the range of A is the whole H^1 and we have just proved the following theorem.
Theorem 3.1: $\forall F \in [H^1(B)]'$, $\forall G \in H^{-1/2}(S)$ *the Neumann problem (3.1), translated into the weak form (3.2), has one and only one solution in* $v \in H^1(B)$. *Furthermore*

$$\|v\|_{H^2} \leq C \left\{ \|F\|_{[H^1(B)]'} + \|G\|_{H^{-1/2}(S)} \right\} . \tag{3.19}$$

Remark 3.3: one might wonder why our Neumann problem has a unique solution while usually in analysis it is claimed to have a null space constituted by constant functions.
Indeed this would be the case if we had $b = 0$ in (3.1); however in our case $b \geq b_0 > 0$ exactly because we have inherited our BVP from an external formulation.
Since the external homogeneous Neumann problem is known to have only the zero solution, the same happens to its internal image (3.1).
We can come now to the oblique derivative problem, formulated as

$$\begin{cases} \Delta v = F & \text{in } B \\ \dfrac{\partial v}{\partial \nu} + \underline{a} \cdot \nabla_t v + bv = G & \text{on } S . \end{cases} \tag{3.20}$$

Following the same reasoning as for (3.1) we immediately come to the weak formulation

$$\forall \varphi \in H^1, \int_B \nabla v \cdot \nabla \varphi dB = - \int_S \underline{a} \cdot \nabla_t v \varphi dS - \int_S bv\varphi dS + \\ + \int_S G\varphi dS - \int_B F\varphi dB . \tag{3.21}$$

By using the same symbolism as before and putting

$$\alpha(v, \varphi) = \langle \alpha v, \varphi \rangle_{H_1} = \int (\underline{a} \cdot \nabla_t v)\varphi dS \tag{3.22}$$

we come straightforwardly to the formulation

$$\langle Av, \varphi \rangle_{H^1} + \langle \alpha v, \varphi \rangle_{H^1} = \langle \Gamma G - CF, \varphi \rangle_{H^1} \tag{3.23}$$

or

$$Av + \alpha v = \Gamma G - CF \ , \ v \in H^1 . \tag{3.24}$$

Since we have proved in (3.18) that

$$A \geq \frac{1}{\widetilde{A'}} I$$

in $H^1(B)$, then (3.24) will have one and only one solution on condition that α is a bounded operator in $H^1(B)$ and that for instance

$$\|\alpha\| < \frac{1}{\widetilde{A'}} .$$

Fortunately, following the famous theorem of Lax and Milgram (cfr. e.g. [Miranda, 1970]), we can find a milder condition for the existence and uniqueness of the solution of (3.23), which is summarized in the two requirements that

$$|A(v,\varphi) + \alpha(v,\varphi)| \leq \text{const}\, \|v\|_{H^1} \|\varphi\|_{H^1} \tag{3.25}$$

$$A(v,v) + \alpha(v,v) \geq \text{const}\, \|v\|^2_{H^1} . \tag{3.26}$$

In fact from (3.26) we immediately see that

$$(A+\alpha)v = 0 \quad \Rightarrow \quad v = 0 \tag{3.27}$$

i.e. the operator $(A+\alpha)^{-1}$ exists. Moreover $(A+\alpha)$ must have a dense range in H^1 since

$$u \in H^1, \langle (A+\alpha)v, u\rangle = 0 \quad \forall v, \rightarrow \langle (A+\alpha)u, u\rangle = 0$$
$$\Rightarrow \quad u = 0$$

Finally, combining (3.26) and (3.25) one gets

$$\|v\|^2_{H^1} \leq \text{const}\, \langle (A+\alpha)v, v\rangle \leq \text{const}\, \|(A+\alpha)v\|_{H^1} \cdot \|v\|_{H^1}$$

entailing

$$\|v\|_{H^1(B)} \leq \text{const}\, \|(A+\alpha)v\|_{H^1(B)} , \tag{3.28}$$

which means that $(A+\alpha)^{-1}$ is continuous, i.e. the range of $A+\alpha$ is closed and then it is the whole $H^1(B)$.
To prove (3.25) we need only to verify that

$$|\alpha(v,\varphi)| \leq \text{const}\, \|\nabla_t v\|_{H^{-1/2}(S)} \|\underline{a}\varphi\|_{H^{1/2}(S)} . \tag{3.29}$$

Now assume that $\underline{a} \in C^\lambda(S), \lambda > 1/2$, then $\underline{a}$ is a multiplier in $H^{1/2}(S)$, i.e.

$$\|\underline{a}\varphi\|_{H^{1/2}(S)} \leq \text{const}\, \|\varphi\|_{H^{1/2}(S)} \leq \text{const}\, \|\varphi\|_{H^1(B)} . \tag{3.30}$$

On the other hand (cfr. the Appendix)

$$\|\nabla_t v\|_{H^{-1/2}(S)} \leq \text{const}\, \|v\|_{H^{1/2}(S)} . \tag{3.31}$$

So (3.29) and then (3.26) is proved under the only condition $\underline{a} \in C^\lambda(\lambda > 1/2)$. As for (3.26) we first of all have

$$\begin{aligned} \alpha(v,v) &= \int_S (\underline{a}\cdot\nabla_t v) v dS = \frac{1}{2}\int_S \underline{a}\cdot\nabla_t(v^2) dS \\ &= -\frac{1}{2}\int_S (\nabla_t\cdot\underline{a}) v^2 dS . \end{aligned} \tag{3.32}$$

Therefore (recalling also (3.15), (3.17))

$$A(v,v) + \alpha(v,v) \equiv \int_B \nabla v^2 dB + \int_S \left[b - \frac{1}{2}(\nabla_t\cdot\underline{a})\right] v^2 dS \geq \text{const}\, \|u\|^2_{H^1(B)} ,$$

if

$$b - \frac{1}{2}\nabla_t \cdot \underline{a} \equiv \beta_0 > 0 \,. \tag{3.33}$$

Therefore we have just proved the following theorem.
Theorem 3.2: $\forall F \in [H^1(B)]', \quad \forall G \in H^{1/2}(S)$ *and for every* $\underline{a}$ *such that*

$$\underline{a} \in C^\lambda(S)(\lambda > 1/2), \qquad b - \frac{1}{2}\nabla_t \cdot \underline{a} \geq \beta_0 > 0 \tag{3.34}$$

we have one and only one solution v *of (3.20) in* $H^1(B)$. *Moreover*

$$\|v\|_{H^1(B)} \leq const\left\{\|F\|_{[H^1(B)]'} + \|G\|_{H^{1/2}(S)}\right\} \,. \tag{3.35}$$

In this way we have accomplished the main analysis of the problem initially defined, (1.8), (1.9), in the light of the standard theory of weak solutions of BVP's for the Laplace operator.
From the mathematical point of view, to complete this analysis, one has to verify whether by adding regularity conditions to the data one gets a corresponding regularity improvement of the solution: for instance is it true that if we assume $F \in L^2(B)$ and $G \in H^{1/2}(S)$ (i.e. data one order of derivation more regular) we have also for the solution $v \in H^2(B)$? This question will be answered in a paper to follow this one.
Remark 3.4: s a last comment, let us observe that the condition (3.34) has a simple rough interpretation from the geometric point of view, in fact since

$$\underline{a} = \frac{\underline{\nu_e}}{\cos I} - \underline{\nu} \,,$$

where $\underline{\nu_e}$ is the normal to the ellipsoid through the point, $\underline{\nu}$ is the normal to S and I the inclination of S with respect to the normal vertical, for regions where $I \sim 0, \cos I \sim 1$

$$\nabla_t \cdot \underline{a} \cong 2(c_e - c_s) \tag{3.36}$$

where c_e is the mean curvature of the ellipsoid and c_s is the mean curvature of S. Going through the reciprocal radii transformation, mean curvatures become mean curvature radii ρ_e, ρ_s, while $b \leq \frac{1}{R}$ becomes just the radial distance of the point P on the surface from the origin; so (3.34) with (3.36) transformed becomes just

$$r_P \geq (\rho_e - \rho_s)_P$$

which is certainly a reasonable assumption. A closer look to this relation should be given when the surface S becomes rougher and the inclination I plays a major role.

A Appendix

In this appendix we shall try to summarize, mostly without proofs, the theory of Sobolev spaces and a few facts about functional analysis which have been used throughout the paper.
Definition A.1: $\mathcal{D}$ is the space of functions $\varphi \in C^\infty(R^3)$[1] endowed with the notion of limit

$$\left\{\varphi_n \underset{\mathcal{D}}{\to} \varphi\right\} \Leftrightarrow \{\text{Supp } \varphi_n, \text{Supp } \varphi \subset K \text{compact fixed set}, \quad \varphi_n^{(K)} \to \varphi^{(K)} \text{uniformly on } K\} \tag{A.1}$$

We note that indeed in (A.1) (A.1) can change from sequence to sequence but has to be fixed with respect to n. We recall also that([A indicating the closure of A)

$$\text{Supp } \varphi \equiv [\{x; \varphi(x) \neq 0\}] .$$

Definition A.2: $\mathcal{D}(B), B$ open, is the subspace of $\mathcal{D}$

$$\mathcal{D}(B) \equiv \{\varphi \in \mathcal{D} \ ; \ \text{Supp } \varphi \subset B\} \tag{A.2}$$

Lemma A.1: $\mathcal{D}(B)$ is a closed subspace of $\mathcal{D}$.
Definition A.3: $\mathcal{D}'$ is the topological vector space of distributions, T, in R^3, i.e. of linear continuous functionals on $\mathcal{D}$

$$T, \quad \langle T, \varphi\rangle \in R \ , \ |\langle T, \varphi\rangle| < +\infty \qquad \forall \varphi \in \mathcal{D} \tag{A.3}$$
$$\langle T, \lambda\varphi + \mu\psi\rangle = \lambda \langle T, \varphi\rangle + \mu \langle T, \psi\rangle \tag{A.4}$$
$$\{\varphi_n \underset{\mathcal{D}}{\to} \varphi\} \Rightarrow \langle T, \varphi_n\rangle \to \langle T, \varphi\rangle \ . \tag{A.5}$$

Lemma A.2: $\mathcal{D}'$ is a complete topological vector space, with the weak dual topology

$$\{T_n \to T\} \Leftrightarrow \langle T_n, \varphi\rangle \to \langle T, \varphi\rangle \ , \quad \forall \varphi \in \mathcal{D} \tag{A.6}$$

Remark A.1: let f be a measurable function $f \in L^2_{loc}$ (i.e. $\int_{(r \leq R)} f^2 dB < +\infty \ \ \forall R$); then

$$\langle T_f, \varphi\rangle \equiv \int_{R^3} f\varphi dx \tag{A.7}$$

is a distribution T_f which we identify with the function f

$$T_f = f$$

In particular $T = 0$ can be made to coincide with any measurable function equal to zero almost everywhere, since

$$\langle T, \varphi\rangle = \int f\varphi dx = 0 \quad \forall \varphi \in \mathcal{D}$$

implies

$$f = 0 \quad a.e.$$

Definition A.4: $\mathcal{D}'(B)$ is the topological dual of $\mathcal{D}(B)$.

[1]The notions given here are all valid in R^n but we limit ourselves to the case of R^3, relevant to geodetic problems.

Definition A.5: let Ω be the largest open set such that

$$\langle T, \varphi \rangle = 0 \quad \forall \varphi, \text{Supp } \varphi \subset \Omega \ ;$$

then the support of T is the closed set

$$\text{Supp } T = \Omega^C \tag{A.8}$$

Remark A.2: $\mathcal{D}'(B)$ is isomorphic to the closed subspace of $\mathcal{D}'$ of all T such that Supp $T \subset B$.
Definition A.6: first we observe that any differential monomial

$$D^s = D_1^{s_1} D_2^{s_2} D_3^{s_3} \ , \quad |s| = s_1 + s_2 + s_3 \ ,$$

is a continuous linear operator such that

$$D^s \varphi = \psi \in \mathcal{D} \ , \quad \forall \varphi \in \mathcal{D} \ ,$$

because Supp $\psi \subseteq$ Supp φ; then we define

$$D^s T = U \Leftrightarrow \langle U, \varphi \rangle \equiv (-1)^{|s|} \langle T, D^s \varphi \rangle \ , \quad \forall \varphi \in \mathcal{D} \tag{A.9}$$

Remark A.3: any measurable function considered as a distribution has distributional derivatives of any order.
In particular two functions f, g coinciding almost everywhere have the same derivatives, since $u = f - g = 0$ a.e. so that

$$\langle D_i u, \varphi \rangle = - \langle u, D_i \varphi \rangle = - \int u D_i \varphi dx = 0 \ .$$

Note that, with Definition A.6 for any function f with continuous derivatives $D_i f$, the distributional derivativs are the same functions.
Definition A.7: the Sobolev space H^k (k integer ≥ 0) is defined as the linear subspace of $\mathcal{D}'$ of functions f such that

$$\left(D^j = D_1^{j_1} D_2^{j_2} D_3^{j_3} \right), \int \sum_{|j|=0}^{k} (D^j f)^2 dx < +\infty \ , \tag{A.10}$$

for instance for $f \in H^1$

$$\int \left[f^2 + \sum_{j=1}^{3} (D_j f)^2 \right] dx < +\infty \ .$$

Let us observe that (A.10) is the square of a norm derived from a scalar product, then H^k is at least a pre-Hilbert space.
Remark A.4: by using the Fourier transform $\widehat{f}$ of f and using the Parseval's identity, we see that (A.10) is equivalent to

$$\int \left(\sum_{|j|=0}^{k} \xi^{2j} \right) \left| \widehat{f} \right|^2 d\xi < +\infty \ . \tag{A.11}$$

where $\xi^{2j} = \xi_1^{2j_1} \xi_2^{2j_2} \xi_3^{2j_3}$.
Since clearly the polynomial

$$\sum_{|j|=0}^{k} \xi^{2j} = 1 + \xi_1^2 + \xi_2^2 + \xi_3^2 + \ldots \geq 1$$

is strictly positive, we see from (A.11) that $\widehat{f} \in L^2$, i.e. $f \in L^2$ and more precisely that

$$f \in H^k \quad \Leftrightarrow \quad \widehat{f} = \left(\sum_{|j|=0}^{k} \xi^{2j} \right)^{-1/2} \widehat{g} \; ; \; \widehat{g} \in L^2 \tag{A.12}$$

Lemma A.3: as a consequence of (A.12) we see that H^k is complete, i.e. it is a Hilbert space.
Remark A.5: since we have, $\forall \xi$

$$a \left(1 + |\xi|^2\right)^k \leq \sum_{|j|=2}^{k} \xi^{2j} \leq b \left(1 + |\xi|^2\right)^k, (a, b > 0),$$

we see that condition (A.11) is equivalent to

$$\int (1 + |\xi|^2)^k \left| \widehat{f} \right|^2 d\xi < +\infty \tag{A.13}$$

which defines an equivalent norm in H^k.
Definition A.8: for any real $s \geq 0$ we define H^s as the space of $f \in L^2$ such that

$$\int (1 + |\xi|^2)^s \left| \widehat{f} \right|^2 d\xi \; < +\infty \tag{A.14}$$

In this way we can define as well fractionary Sobolev spaces like $H^{1/2}, H^{3/2}$ etc.
We note explicitely that (A.14) makes sense also $\forall s$ real, $s < 0$; so we can introduce as well Sobolev spaces with negative order.
Definition A.9: let $\Re_B$ be the operator of restriction to B of a function f defined in R^3; then

$$H^s(B) \equiv \{ \Re_B f \; ; \; f \in H^s \} \; . \tag{A.15}$$

Lemma A.4: $H^s(B)$ is a Hilbert space and, when $s = k$ integer,

$$\|f\|^2_{H^k(B)} = \int_B \left(\sum_{|j|=0}^{k} D^j f \right)^2 dB \tag{A.16}$$

Lemma A.5: for any real s, s', we have the embedding chain

$$\mathcal{D} \subset H^s \subset H^{s'} \qquad (s' \leq s) \; ; \tag{A.17}$$

meaning that

$$f \in \mathcal{D} \Rightarrow f \in H^s \; , \; f \in H^s \Rightarrow f \in H^{s'} \qquad (s' \leq s) \; ;$$

the embedding operator J

$$J : H^s \to H^{s'} \qquad Jf \equiv f$$

is dense and, when $s' < s$, compact.
This means that the image of $\mathcal{D}$ in any H^s is dense and that given a sequence $\{f_n\}$ bounded in H^s ($\|f_u\|_{H^s} < \text{const}$), it has at least an accumulation point $\overline{f}$ in $H^{s'}$.
The same holds true for $\mathcal{D}(B), H^s(B), H^{s'}(B)$.
We underline that the above statements are valid for positive as well as for negative s.
Remark A.6: with the help of local coordinates systems one can extended the concept of Sobolev spaces to surfaces. To make it simple let's assume that S is a surface with finite, continuous curvature so that we can introduce local systems of coordinates

$$\underline{\sigma} \equiv (\vartheta, \lambda), \; \sigma \in Q \equiv \{ \vartheta_1 \leq \vartheta \leq \vartheta_2, \; \lambda_1 \leq \lambda \leq \lambda_2 \}$$

with orthogonal coordinate lines along the principal curvature directions.
Then $\forall f$ sufficiently smooth

$$\nabla_t f = \underline{e}_\vartheta \frac{1}{\rho_\vartheta}\frac{\partial f}{\partial \vartheta} + \underline{e}_\lambda \frac{1}{\rho_\lambda}\frac{\partial f}{\partial \lambda} \tag{A.18}$$

($\rho_\vartheta, \rho_\lambda$ curvature radii).
Assume now that $f \equiv 0$ outside the patch $A_{\underline{\sigma}} \equiv \underline{x}(\underline{\sigma}), (\underline{\sigma} \in Q)$ where the local system $\underline{\sigma}$ is defined above, then, considering that $dS = \rho_\vartheta \rho_\lambda d\vartheta d\lambda$, we can set

$$\int_S (\nabla_t f)^2 dS = \int_{A_{\underline{\sigma}}} (\nabla_t f)^2 dS = \int_\vartheta \left[\frac{\rho_\lambda}{\rho_\vartheta}\left(\frac{\partial f}{\partial \vartheta}\right)^2 + \frac{\rho_\vartheta}{\rho_\lambda}\left(\frac{\partial f}{\partial \lambda}\right)^2 \right] d\vartheta d\lambda$$

showing that, if $a \leq \dfrac{\rho_\lambda}{\rho_\vartheta} \leq b$, the two conditions

$$\int_S |\nabla_t f|^2 \, dS < +\infty \, , \qquad \int_Q |\nabla_\sigma f|^2 \, dQ < +\infty \tag{A.19}$$

are equivalent.
Since we can split S into a finite number of regular overlapping patches, we see that the case $f \in H^1(S)$ can be defined through coordinates transformations, stretching S on $\Re^2$.
Definition A.10: let S be a surface with a parametric representation, $\underline{x}(t_1, t_2)$, continuous up to k-th derivatives, then we define H^k as the closure of $\{\Re_S \varphi \, , \, \varphi \in \mathcal{D}\}$ in the norm

$$\int_S \sum_{s_1+s_2=0}^{k} \left(\frac{\partial^{s_1}}{\partial t_1}\frac{\partial^{s_2}}{\partial t_2} f \right)^2 dS = \|f\|^2_{H^k(S)} \; ; \tag{A.20}$$

in particular

$$\int_S \left(f^2 + |\nabla_t f|^2 \right) dS = \|f\|^2_{H^1(S)} \; ;$$

moreover, by stretching S on $\Re^2$ one can define as well $H^s(S)$ for any real $s \geq 0$.
Lemma A.6: the operator of multiplication of $f \in H^{1/2}(S)$ by a function $\alpha \in C^\lambda$ ($\lambda > 1/2$) is bounded in $H^{1/2}(S)$

$$\|\alpha f\|_{H^{1/2}(S)} \leq \text{const} \cdot \|f\|_{H^{1/2}(S)} \; . \tag{A.21}$$

This is easy to understand by using an equivalent definition of the $H^{1/2}(S)$ norm

$$\|f\|^2_{H^{1/2}(S)} = \int_S f^2 dS + \int_S dS_y \int_S dS_x \frac{|f(x) - f(y)|^2}{|x-y|^4} \, ,$$

given by Gagliardo (cfr. [Lions, Magenes, 1968]).
Lemma A.7: let $f \in H^s(B)$, $s > 1/2$; then, if we call $\Re_S f$ the trace of f on S, we have

$$\Re_S f \in H^{s-1/2}(S) \; ; \; \|\Re_S f\|_{H^{s-1/2}(S)} \leq C \, \|f\|_{H^s(B)} \tag{A.22}$$

Remark A.7: let us note explicitely that $H^s(B)$, when $s > 1/2$, is a space of functions which on S can have a trace $\neq 0$. Therefore one cannot say that $\mathcal{D}(B)$ is dense in $H^s(B)$. We call

$$(s > 1/2), H^s_0(B) \equiv [\mathcal{D}(B)]_{H^s(B)} \; ; \tag{A.23}$$

$H_0^s(B)$ is a proper closed subspace of $H^s(B)$. On the contrary, when $s < 1/2$, it is not possible to define a continuous operator of trace $\Re_S$, therefore we have

$$(s < 1/2) \quad H^s(B) \equiv [\mathcal{D}(B)] \,. \tag{A.24}$$

(A.24) holds for negative values of s, too. On the other hand, as claimed in Lemma A.5, when $B \equiv \Re^3, \mathcal{D}$ is dense in all the H^s.

Remark A.8: let us observe that $\forall \varphi, \psi \in \mathcal{D}$ the following inequalities hold

$$\left| \int \varphi \psi dx \right|^2 \equiv \left| \int \widehat{\varphi}^* \widehat{\psi} d\xi \right|^2 \leq \int |\widehat{\varphi}|^2 (1 + |\xi|^2)^{-s} d\xi \cdot \tag{A.25}$$
$$\cdot \int \left| \widehat{\psi} \right|^2 (1 + |\xi|^2)^s d\xi \equiv \|\varphi\|^2_{H^{-s}} \|\psi\|^2_{H^s} \,,$$

where the integrals refer to two whole R^3.
Since in $R^3\mathcal{D}$ is dense in both H^{-s} and H^s we see that if we take

$$\varphi_n \in \mathcal{D} \qquad \varphi_n \underset{H^{-s}}{\to} f$$

$$\psi_n \in \mathcal{D} \qquad \psi_n \underset{H^s}{\to} g$$

we can extend the symbol $\int \varphi \psi dx$ to

$$\int f g dx \equiv \lim_{n,m \to \infty} \int \varphi_n \psi_m dx \,; \tag{A.26}$$

furthermore we have

$$\forall f \in H^{-s}, \ \forall g \in H^s; \left| \int f g dx \right| \leq \|f\|_{H^{-s}} \|g\|_{H^s} \,. \tag{A.27}$$

If we repeat the same reasoning for $s > 1/2$ and $\varphi_n \in \mathcal{D}(B), \psi_n \in \mathcal{D}(B)$, since $\mathcal{D}(B)$ is dense in $H_0^s(B)$ but not in $H(B)$ we see that

$$\forall f \in H^{-s}(B), \quad \forall g \in H_0^s(B) \,; \ \left| \int_B f g dB \right| \leq \|f\|_{H^{-s}(B)} \|g\|_{H^s(B)} \tag{A.28}$$

On the other hand when we take a closed surface S, since by coordinate transformation we map it onto R^2, we have, like in (A.27),

$$\forall f \in H^{-s}(S), \ \forall g \in H^s(S) \,; \left| \int_S f g dS \right| \leq \|f\|_{H^s(S)} \leq \|g\|_{H^{-s}(S)} \,. \tag{A.29}$$

Therefore we claim that, indicating by X' the dual of a space X,

$$H^{-s} \equiv (H^s)' \,, \quad H^{-s}(B) \equiv (H_0^s(B))' \,, \quad H^{-s}(S) = (H^s(S))' \tag{A.30}$$

and by these identifications we mean that if $F \in (H^s)'$ then there is $f \in H^{-s}$ such that

$$F(g) \equiv \int f g dx, \quad \forall g \in H^s$$

and so forth.

The above Remark is a particular case of a more general result.

Lemma A.8: let X, Y be two Hilbert spaces with $X \subset Y$, the embedding being dense and continuous; then if we identify $Y' \equiv Y$ via the Riesz theorem we can write

$$X \subset Y \equiv Y' \subset X' \tag{A.31}$$

with $Y' \equiv Y$ continuously and densely embedded in X', so that $\forall x' \in X'$ we can write

$$x'(x) \equiv (x', x)_Y \tag{A.32}$$

understanding the scalar product as a limit of $(y_n, x)_Y$ with $y_n \in Y$ and $y_n \underset{X'}{\to} x'$.

Lemma A.9: if D^k is any differential operator of order $|k|$, then D^k is a continuous linear operator of H^s into $H^{s-|k|}$, i.e.

$$\left\| D^k f \right\|_{H^{s'}} \leq \text{const} \, \|f\|_{H^{s'+|k|}} \; ; \tag{A.33}$$

moreover if A is any continuous linear operator $A : H^s \to H^{s'}$ $(s' > s)$, then A is a compact operator of H^s into itself.

Lemma A.10: we want to sketch the proof of the following fundamental inequality (Rellich): $\forall f \in H^1(B)$ we have

$$\int_B f^2 dB \leq \text{const} \left\{ \int_S f^2 dB + \int_B |\nabla f|^2 \, dB \right\} . \tag{A.34}$$

□ It is enough to prove (A.34) $\forall \Re_B f, f \in \mathcal{D}$. Let us consider the Green's function $G(\underline{x}, \underline{y})$ of the Laplacian in the domain B and put

$$w(\underline{x}) \equiv G(\underline{x}, 0) \, .$$

The function $w(\underline{x})$ is harmonic in $B \backslash \{0\}, w = 0$ on S while $w = 0\left(\frac{1}{r}\right)$ when $r \to 0$; moreover the surfaces $S_{\overline{w}} \equiv \{\underline{x}; w(\underline{x}) = \overline{w}\}$ are interior to one another while $\overline{w} \to \infty$, so that if $B_{\overline{w}} \equiv$ interior $\{S_{\overline{w}}\}$, one has

$$B_0 = B, \; B_{w_1} \subset B_{w_2} \text{ iff } \; w_1 \supset w_2 \, , \tag{A.35}$$

in addition the vector $-\nabla w$ is such that

$$|\nabla w| \neq 0 \text{ in } B \backslash \{0\} \, , \; |\nabla w| = 0\left(\frac{1}{r^2}\right) \text{ for } r \to 0 \tag{A.36}$$

so that

$$|\nabla w| \geq G_0 > 0 \; ; \tag{A.37}$$

furthermore the force lines of $-\nabla w$ never cross, while they have a focus in 0 and if we introduce a curvilinear coordinate ℓ_Q as in Fig. A.1, $\forall Q \neq 0$ we have a couple $\{P, \ell\}$ such that $P \in S, 0 \leq \ell \leq L$ which identifies univocally Q.

We also observe that $-\nabla w$ is directed as the exterior normal to S_w and that

$$|\nabla w| = \frac{\partial u}{\partial \ell} \; ; \tag{A.38}$$

furthermore, from Gauss theorem, we know that

$$\left(\frac{\partial w}{\partial \ell}\right)_Q dS_Q \equiv \left(\frac{\partial w}{\partial \ell}\right)_P dS_P \, , \tag{A.39}$$

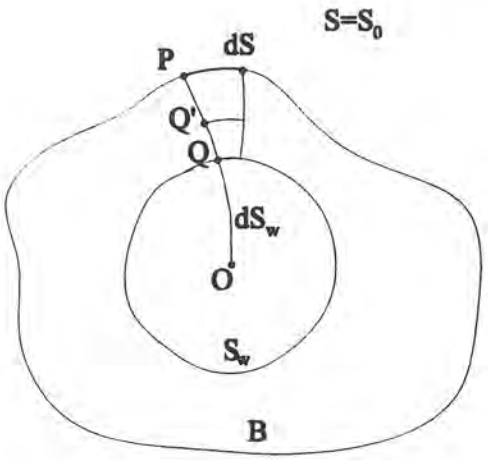


Figure A.1: Green's coordinates in B.

implying that

$$\forall Q; P \in B\backslash\{0\}, \quad dS_Q \leq C dS_P \; ; \tag{A.40}$$

likewise $\forall Q'$ with $\ell_{Q'} \leq \ell_Q$ we have

$$dS_{Q'} \leq J dS_Q \; . \tag{A.41}$$

Then, for $f \in \mathcal{D}$, we write

$$f(Q) = f(P) + \int_P^Q \left(\frac{\partial f}{\partial \ell}\right) d\ell \quad \Rightarrow \quad |f(Q)|^2 \leq 2\left[f^2(P) + L\int_P^Q \left(\frac{\partial f}{\partial \ell}\right)^2 d\ell\right]$$

so that, using (A.40) and (A.41),

$$|f(Q)|^2 \, dS_Q \leq \text{const}\left[f^2(P) dS_P + \int_P^Q \left(\frac{\partial f}{\partial \ell}\right)^2 dS_{Q'} \cdot d\ell\right] \; .$$

Observing that $dS_Q d\ell = dB_Q$ and extending the integration from Q up to the origin O along the force line L_P, we get

$$\int_{L_P} |f(Q)|^2 \, dB_Q \leq \text{const} \cdot L\left[f^2(P) dS_P + \int_{L_P} \left(\frac{\partial f}{\partial \ell}\right)^2 dB_{Q'}\right]$$

which finally integrated over all $P \in S_0$ proves (A.34). □

References

[Sansò, 1995] Sansò, 1995. "The Long road from measurement to boundary value problems in physical geodesy". XXI IUGG General Assembly, Boulder, Colorado, July 2-14 1995. *Manuscripta Geodaetica*, vol. 20, 1995, n. 5.

[Sansò, 1997] F. Sansò, 1997. "The hierarchy of Geodetic BVP's." In: Lecture Notes in Earth Sciences, n. 65, 1997.

[Rozanov, Sansò, 1997] Y. Rozanov, F. Sansò, 1997. "BPV for harmonic random fields". In: Lecture Notes on Earth Sciences, n. 65, 1997

[Holota, 1997] P. Holota, 1997. "Variational methods for geodetic BVP's." In: Lecture Notes on Earth Sciences, n. 65, 1997.

[Lions, Magenes, 1968] J. L. Lions, E. Magenes, 1968. "Problemes aux limites non homogénes". Paris, 1968.

[Gilbarg, Trudinger, 1982] T. Gilbarg, N. S. Trudinger, 1982. "Elliptic partial differential equations of second order". Grund. der Mathematischen Vissenschaften, vol. 224, 1982.

[Miranda, 1970] C. Miranda, 1970. "Partial differential equations of elliptic type." Springer, 1970.

[Magenes, Stampacchia, 1954] E. Magenes, G. Stampacchia, 1954. "Problemi al contorno per le equazioni differenziali di tipo ellittico". Annali della Scuola Normale Superiore di Pisa, n. 12, 1954.

[Treves, 1967] F. Treves, 1967. "Topological vector spaces, distributions and kernels". Academic Press, 1967.

[Sansò, Venuti, 1998] F. Sansò, G. Venuti, 1998. "White noise stochastic BVP's and Cimmino's theory". Presented to Hotine-Marussi Symposium, in print.

From the Generalized Bruns Transformation to Variations of the Solution of the Geodetic Boundary Value Problem

Reiner Rummel and Martin van Gelderen

1 Introduction

Erik Grafarend has inspired geodesy in so many ways, that selecting one area automatically implies to do injustice to all others. Nevertheless, one can probably say that the main focus of his research is mathematical, physical and statistical geodesy. But this is also to say his focus was and is geodesy as a whole. There is a deep love for geometry in Grafarend's work. This is also reflected in his contributions to physical geodesy. Periods where he expressed his thoughts in plane coordinates were followed by others in which he used spherical or ellipsoidal coordinates. For us 'Earthlings' it is impossible to follow his pace. All what we can try is to select one single aspect of his enormous work and reflect on it.

This shall be tried in the sequel, where we take as point of departure his publication *The Bruns transformation and a dual setup of geodetic observational equations.* It was published in 1980 in Washington, D.C., during his stay at the National Geodetic Survey of the US. It is well known that observables in physical geodesy depend on both the gravity field and the position where the measurement is taken. When establishing linear boundary conditions the position part can be eliminated by appropriate combination of observations; e.g. of potential and gravity anomaly. It is the well-known Bruns transformation. In the above work Grafarend could show that the Bruns transformation can be generalized to three or more dimensions and to various observables of physical geodesy. It also shows that the principle of elimination of unknowns, known from classical adjustment theory, can be translated to field quantities, too, and in particular to boundary functions as they are met when solving the geodetic boundary value problem (GBVP). This also means that free boundary value problems can be transformed into a fixed form to which standard solutions apply. For us this was the starting point for the solution of the GBVP in a more generalized fashion, that includes uniquely determined as well as overdetermined cases. We refer to [Rummel and Teunissen(1986)], [Rummel *et al.*(1989)], [Rummel and Van Gelderen(1992)] and [Rummel and Van Gelderen(1999)].

In this article, dedicated to Erik Grafarend at the occasion of his 60th birthday, we deal with GBVPs in several coordinate systems, their solution by separation and their determination. The determination step is the procedure of fixing the unknown parameters of the mathematical solution of the Laplace equation. It is achieved on the basis of gravimetric observations (potential differences, gravity anomalies, deflections of the vertical, torsion balance measurements etc.) carried out at the Earth's surface or reduced to some reference surface. The generalized Bruns transformation is thereby employed to arrive at boundary conditions free of geometric unknowns, e.g. coordinate or height corrections. Now, in 'GPS-age', one may argue that the geometry part of GBVPs is anyway taken care of by 3D positioning. In reality we still have to deal mostly with old data collected over decades. Our classical GBVPs will therefore still maintain their relevance for quite some time.

2 Solution of the Laplace equation by separation

[Moon and Spencer(1961)] have shown the solution of the three-dimensional Laplace equation for eleven orthogonal coordinate systems. For them a relatively simple solution by separation is possible. The coordinate system is usually chosen so as to fit best possible to the surface on which the boundary data are given. In geodesy spherical, rectangular, circular-cylinder, oblate spheroidal and ellipsoidal coordinates are of relevance. We shall deal only with the spherical, rectangular and circular-cylinder cases here and derive some special cases from them that may be of interest. The procedure consists of the following four steps (ibid.):

1. Formulation of Laplace equation in the chosen coordinate system.
2. Separation into a set of three second-order ordinary differential equations.
3. Solution of the ordinary differential equations.
4. Determination of the mathematical solution, as obtained from the superposition of the solutions under 3, by a set of approximate boundary conditions. Formulation of a closed solution whenever possible.

The first three steps have been solved by [Moon and Spencer(1961)] once for all. Thus this part can be kept rather short. We start with an outline of the general approach.
In the coordinate system $\{x^a\} = \{x^1, x^2, x^3\}$ Laplace equation reads

$$\nabla^2 V = \nabla_{ab} V g^{ab} = 0, \tag{1}$$

where ∇_{ab} is the second covariant derivative operator and g^{ab} the metric tensor of the coordinate frame. For orthogonal curvilinear coordinates $g^{ab} = 0 \quad (a \neq b)$ and Laplace equation can be written as [Moon and Spencer(1961), eq. 1.09]

$$\nabla^2 V = \frac{1}{\sqrt{g}} \sum_{a=1}^{3} \nabla_a \left(\frac{\sqrt{g}}{g_{aa}} \nabla_a V \right) = 0.$$

If the coordinate system fulfills certain conditions (ibid.) the solution of Laplace's equation can be found by the *separation of variables*. For V we substitute thereby

$$V(x^1, x^2, x^3) = f(x^1) g(x^2) h(x^3)$$

into (1) and three independent, second-order differential equations are obtained. Their general solution is written as

$$f_{\ell m}(x^1), \quad g_{\ell m}(x^2) \quad \text{and} \quad h_{\ell m}(x^3),$$

respectively, where ℓ and m are two constants of integration. Their possible values will be determined by the type of solution required. The general solution of Laplace equation is obtained from a linear combination of all possible solutions. Assuming for simplicity that ℓ, m take only integer values we can write

$$V(x^1, x^2, x^3) = \sum_{\ell m} a_{\ell m} f_{\ell m}(x^1) g_{\ell m}(x^2) h_{\ell m}(x^3).$$

The value of the constants $a_{\ell m}$ has to be determined from boundary conditions on some surface S. Generally this is not an easy task but for some special cases the solution can be found easily. If the geometry of S coincides with one of the coordinate surfaces, e.g., $x^3 =$ constant, and if

the functions $f_{\ell m}(x^1)g_{\ell m}(x^2)$ form a complete basis in some function space on S then the given boundary function $b(x^1, x^2)$ on S can be written as

$$b(x^1, x^2) = \sum_{\ell m} b_{\ell m} f_{\ell m}(x^1) g_{\ell m}(x^2).$$

With the Dirichlet boundary condition

$$V(x^1, x^2, x^3 = \text{constant}) = b(x^1, x^2)$$

the solution of the boundary value problem is

$$a_{\ell m} = \frac{b_{\ell m}}{h_{\ell m}(x^3 = \text{constant})} \Rightarrow$$

$$V(x^1, x^2, x^3) = \sum_{\ell m} \frac{b_{\ell m}}{h_{\ell m}(x^3 = \text{constant})} f_{\ell m}(x^1) g_{\ell m}(x^2) h_{\ell,m}(x^3).$$

For boundary conditions of other type an analogous solution can be found.

2.1 Solution in spherical coordinates

In spherical coordinates $\{\theta, \lambda, r\} = \{\text{co-latitude, longitude, radial distance}\}$, Laplace equation applied to a potential field V takes the well-known form

$$\nabla^2 V = \frac{\partial^2 V}{\partial r^2} + \frac{2}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial V}{\partial \theta} + \frac{1}{\sin^2\theta}\frac{\partial^2 V}{\partial \lambda^2} = 0. \tag{2}$$

After insertion of first

$$V(\theta, \lambda, r) = Y(\theta, \lambda) h(r) \tag{3a}$$

and then

$$Y(\theta, \lambda) = f(\theta) g(\lambda) \tag{3b}$$

it can be separated, with (3a), into

$$\frac{\partial^2 Y}{\partial \theta^2} + \cot\theta \frac{\partial Y}{\partial \theta} + \frac{1}{\sin^2\theta}\frac{\partial V}{\partial \lambda} + n(n+1)Y = 0 \tag{4}$$

and

$$\frac{\partial^2 h}{\partial r^2} + \frac{2}{r}\frac{\partial h}{\partial r} - \frac{n(n+1)}{r^2} h = 0 \tag{5}$$

($n \in \mathbb{N}$). Eq. (4) is the so-called surface Laplace or Beltrami equation with $Y(\theta, \lambda)$ the surface spherical harmonics. After insertion of (3b) one arrives at ($m \in \mathbb{Z}, |m| \leq n$)

$$\frac{\partial^2 f}{\partial \theta^2} + \cot\theta \frac{\partial f}{\partial \theta} + \left[n(n+1) - \frac{m^2}{\sin^2\theta} \right] f = 0, \tag{6}$$

and

$$\frac{\partial^2 g}{\partial \lambda^2} + m^2 g = 0. \tag{7}$$

The solutions of the ordinary second order differential equations (5),(6) and (7) are:

$$h(r) = a r^n + b r^{-(n+1)} \tag{8a}$$

$$f(\theta) = a P_{nm}(\cos\theta) + b Q_{nm}(\cos\theta) \tag{8b}$$

with $P_{nm}(\cos\theta)$ and $Q_{nm}(\cos\theta)$ the associated Legendre functions of the first and second kind, respectively, and

$$g(\lambda) = a\sin m\lambda + b\cos m\lambda \tag{9a}$$
$$= c\exp(im\lambda) + d\exp(-im\lambda). \tag{9b}$$

The two solutions (9a,9b) are equivalent. We select (9b) because it leads to a more compact form.With the surface spherical harmonics

$$Y_{nm}(\theta,\lambda) = P_{nm}(cos\theta)\exp(im\lambda) \tag{10}$$

and equivalently

$$Z_{nm}(\theta,\lambda) = Q_{nm}(cos\theta)\exp(im\lambda) \tag{11}$$

the complete set of solutions for all admissible integer degrees n and orders m becomes:

$$\begin{aligned} V(\theta,\lambda,r) = &\sum_{n=0}^{\infty} r^{-(n+1)} \sum_{m=-n}^{n} a_{nm}Y_{nm}(\theta,\lambda) \\ &+ \sum_{n=0}^{\infty} r^{n} \sum_{m=-n}^{n} b_{nm}Y_{nm}(\theta,\lambda) \\ &+ \sum_{n=0}^{\infty} r^{-(n+1)} \sum_{m=-n}^{n} c_{nm}Z_{nm}(\theta,\lambda) \\ &+ \sum_{n=0}^{\infty} r^{n} \sum_{m=-n}^{n} d_{nm}Z_{nm}(\theta,\lambda). \end{aligned} \tag{12}$$

Special case – V independent of λ (axial-symmetric) For V independent of λ eq. (6) changes into

$$\frac{\partial^2 f}{\partial\theta^2} + \cot\theta\frac{\partial f}{\partial\theta} + n(n+1)f = 0, \tag{13}$$

the characteristic equation for the Legendre polynomials, which has the solution

$$f(\theta) = aP_n(cos\theta) + bQ_n(cos\theta). \tag{14}$$

For this case the complete set of solutions reads:

$$\begin{aligned} V(\theta,\lambda) = &\sum_{n=0}^{\infty} a_n r^{-(n+1)}P_n(cos\theta) + \sum_{n=0}^{\infty} b_n r^{n}P_n(cos\theta) \\ &+ \sum_{n=0}^{\infty} c_n r^{-(n+1)}Q_n(cos\theta) + \sum_{n=0}^{\infty} d_n r^{n}Q_n(cos\theta). \end{aligned} \tag{15}$$

2.2 Solution in Cartesian coordinates

The Cartesian coordinate triple $\{x,y,z\}$ is arbitrarily chosen to mean {North direction in the $\{x,y\}$-plane, East direction in the $\{x,y\}$-plane, positive up}. Laplace equation in Cartesian coordinates

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0, \tag{16}$$

is dealt with by

$$V(x,y,z) = X(x)Y(y)Z(z). \tag{17}$$

It separates either into

$$\frac{\partial^2 X}{\partial x^2} + k^2 X = 0, \tag{18a}$$

$$\frac{\partial^2 Y}{\partial y^2} + \ell^2 Y = 0 \quad \text{and} \tag{18b}$$

$$\frac{\partial^2 Z}{\partial z^2} - (k^2 + \ell^2) Z = 0 \tag{18c}$$

or into

$$\frac{\partial^2 X}{\partial x^2} + k^2 X = 0, \tag{19a}$$

$$\frac{\partial^2 Y}{\partial y^2} - \ell^2 Y = 0 \quad \text{and} \tag{19b}$$

$$\frac{\partial^2 Z}{\partial z^2} - (k^2 - \ell^2) Z = 0. \tag{19c}$$

In the first case the solutions of the first two ordinary second-order differential equations are:

$$X(x) = a\exp(-ikx) + b\exp(ikx) \tag{20a}$$
$$Y(y) = a\exp(-i\ell y) + b\exp(i\ell y) \tag{20b}$$

or equivalently written in sin/cos-terms. This leads for $Z(z)$ to

$$Z(z) = a\exp(-\sqrt{k^2+\ell^2}z) + b\exp(\sqrt{k^2+\ell^2}z). \tag{20c}$$

In the second case we obtain

$$X(x) = a\exp(-ikx) + b\exp(ikx), \tag{21a}$$
$$Y(y) = a\exp(-\ell y) + b\exp(\ell y) \quad \text{and} \tag{21b}$$
$$Z(z) = a\exp(-\sqrt{k^2-\ell^2}z) + b\exp(\sqrt{k^2-\ell^2}z). \tag{21c}$$

The complete set of solutions becomes – with k and ℓ assumed to be integers!:

$$V(x,y,z) = \sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} a_{k\ell}\exp[i(kx+\ell y)(-\sqrt{k^2+\ell^2}z)] + \sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} b_{k\ell}\exp[i(kx+\ell y)\sqrt{k^2+\ell^2}z] \tag{22}$$

or, alternatively, from eqs. (21a)-(21c):

$$V(x,y,z) = \sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} a_{k\ell}\exp\left[(ikx+\ell y-\sqrt{k^2-\ell^2}z\right] + \sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} b_{k\ell}\exp\left[ikx+\ell y+\sqrt{k^2-\ell^2}z\right]. \tag{23}$$

Special case – V independent of y For V independent of y two cases are distinguished:

Case 1 :

$$\frac{\partial^2 X}{\partial x^2} + k^2 X = 0 \tag{24a}$$

and

$$\frac{\partial^2 Z}{\partial z^2} - k^2 Z = 0 \tag{24b}$$

with the solutions

$$\begin{aligned} X(x) &= a\exp(-ikx) + b\exp(ikx) \\ Z(z) &= a\exp(-kz) + b\exp(kz). \end{aligned} \tag{25a}$$

Case 2 :

$$\frac{\partial^2 X}{\partial x^2} = \frac{\partial^2 Z}{\partial z^2} = 0 \tag{26}$$

with the solution

$$X(x) = a + bx \quad \text{and} \tag{27a}$$

$$Z(z) = a + bz. \tag{27b}$$

For the complete solutions we find:

Case 1 :

$$\begin{aligned} V(x,z) = &\sum_{k=-\infty}^{\infty} a_k \exp\left[ikx - |k|z\right] \\ &+ \sum_{k=-\infty}^{\infty} b_k \exp\left[ikx + |k|z\right] \end{aligned} \tag{28}$$

and

Case 2 :

$$V(x,z) = a + bx + cz + dxz. \tag{29}$$

2.3 Solution in circular cylinder coordinates

In circular cylinder coordinates $\{r, \lambda, z\}$ = {radial in $\{x,y\}$-plane, longitude in $\{x,y\}$-plane, positive up} Laplace equation becomes

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \lambda^2} + \frac{\partial^2 V}{\partial z^2} = 0. \tag{30}$$

With

$$V(r,\lambda,z) = f(r)g(\lambda)Z(z) \tag{31}$$

it separates into

$$\frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} - (k^2 + \frac{m^2}{r^2})f = 0, \tag{32a}$$

$$\frac{\partial^2 g}{\partial \lambda^2} + m^2 g = 0, \tag{32b}$$

$$\frac{\partial^2 Z}{\partial z^2} + k^2 Z = 0. \tag{32c}$$

The solutions are e.g.

$$f(r) = aI_m(kr) + bK_m(kr) \tag{33a}$$

$$g(\lambda) = a\exp(-im\lambda) + b\exp(im\lambda) \tag{33b}$$

$$Z(z) = a\exp(-ikz) + b\exp(ikz) \tag{33c}$$

with the modified Bessel functions of the first kind I_m and of the second kind K_m; see [Lehner(1996)]. Alternatively, one could separate into

$$\frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + (k^2 - \frac{m^2}{r^2})f = 0, \tag{34}$$

$$\frac{\partial^2 g}{\partial \lambda^2} + m^2 g = 0, \quad \text{and} \tag{35}$$

$$\frac{\partial^2 Z}{\partial z^2} - k^2 Z = 0, \tag{36}$$

with solutions

$$f(r) = aJ_m(kr) + bN_m(kr) \tag{37a}$$

$$g(\lambda) = a\exp(-im\lambda) + b\exp(im\lambda) \tag{37b}$$

$$Z(z) = a\exp(-kz) + b\exp(kz) \tag{37c}$$

with the Bessel function J_m and the Neumann functions (Bessel functions of the second kind) N_m. The complete set of solutions becomes now

$$V(r,\lambda,z) = \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_{km}\exp(i(m\lambda + kz))K_m(kr)\,dk + \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} b_{km}\exp(i(m\lambda + kz))I_m(kr)\,dk \tag{38}$$

or

$$V(r,\lambda,z) = \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_{km}\exp(im\lambda + kz))J_m(kr)\,dk + \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} b_{km}\exp(i(m\lambda + kz))N_m(kr)\,dk. \tag{39}$$

Special case – V independent of z For $V(r,\lambda)$ the two ordinary second-order differential equations become

$$\frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} - \frac{m^2}{r^2} f = 0 \tag{40}$$

and

$$\frac{\partial^2 g}{\partial \lambda^2} + m^2 g = 0 \tag{41}$$

with the well-known solutions

$$f(r) = a r^m + b r^{-m} \tag{42}$$

and

$$g(\lambda) = a\exp(-im\lambda) + b\exp(im\lambda). \tag{43}$$

Special attention has to be paid to the case $m = 0$, for then it is

$$f(r) = a + b\ln r \quad \text{and} \tag{44a}$$
$$g(\lambda) = a. \tag{44b}$$

This results in the complete set of solutions

$$\begin{aligned} V(r,\lambda) =& c_0 + c_1 \ln r + \\ &+ \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} a_m r^{-|m|} \exp(im\lambda) \\ &+ \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} b_m r^{|m|} \exp(im\lambda). \end{aligned} \tag{45}$$

See also [Walter(1971),].

3 Determination of the solution by boundary conditions

In the previous chapter the solution of Laplace equation was given in spherical, Cartesian and circular-cylinder coordinates. Also three special cases are included. Now the mathematical solutions are to be determined with the help of boundary conditions. Only for the spherical case this step shall be discussed explicitly. For the two other coordinate systems and for the special cases the solutions will only be summarized. Throughout only "exterior" problems will be treated with a regularity condition at infinity for V.

3.1 Determination of the solution for the exterior of a sphere with radius $r = R$

The mathematical solution is given in eq. (12). The following assumptions hold:

(α) $\mathcal{D}V(\theta,\lambda,r=R) = f(\theta,\lambda)$
The linear differential operator D applied to the potential V takes the values of the boundary function f on the sphere $S(O, r = R)$.

(β) $\lim_{r\to\infty} V(\theta,\lambda,r) = 0$
The potential V takes the values zero of the hypothetical boundary function $f'(\theta,\lambda) = 0$ on a sphere S' with $\lim r \to \infty$ (regularity condition).

Determination:

- since all $Z_{nm}(\theta, \lambda)$ take the value infinity for $\theta = 0$ (= z-axis), for V in order not to become infinity all coefficients c_{nm} and d_{nm} must be zero:

$$c_{nm} = d_{nm} = 0; \tag{46}$$

- in order to meet boundary condition (β) with $\lim_{r\to\infty} r^n = \infty$ all b_{nm} must be zero:

$$(\beta) \qquad b_{nm} = 0. \tag{47}$$

It remains to determine the coefficients a_{nm}. We discuss three cases of (α): the boundary conditions of type Dirichlet, Neumann and Stokes.

Type Dirichlet ($\mathcal{D} = 1$): Then with

$$\begin{aligned} A_{nm} =& \frac{1}{4\pi} \iint f(\theta, \lambda) Y_{nm}(\theta, \lambda)\, d\sigma \\ =& \langle f, Y_{nm}(\theta, \lambda) \rangle \end{aligned} \tag{48}$$

(we now assume that the spherical harmonics Y_{nm} are orthonormal) and therefore

$$f(\theta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{nm} Y_{nm}(\theta, \lambda), \tag{49}$$

it follows by comparison of coefficients (compare (12) and (49))

$$R^{-(n+1)} a_{nm} = A_{nm} \tag{50}$$

and therefore

$$a_{nm} = R^{n+1} A_{nm}. \qquad (\alpha - I) \tag{51}$$

This could be called the spectral solution of the Dirichlet problem for the exterior of the sphere. Insertion into (12) yields

$$V(\theta, \lambda, r) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} A_{nm} Y_{nm}(\theta, \lambda). \tag{52}$$

With (48) and the addition theorem

$$(2n+1) P_n(\cos\psi) = \sum_{m=-n}^{n} Y_{nm}(\theta, \lambda) Y_{nm}(\theta', \lambda') \tag{53}$$

one arrives at

$$\begin{aligned} V(\theta, \lambda, r) &= \frac{1}{4\pi} \iint \left\{ \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} (2n+1) P_n(\cos\psi) \right\} f(\theta', \lambda')\, d\sigma' \\ &= \frac{1}{4\pi} \iint D(\psi; r, R) f(\theta', \lambda')\, d\sigma'. \end{aligned} \tag{54}$$

This is the closed solution of the spherical Dirichlet problem, in geodesy referred to as *Poisson equation.*

Type Neumann ($\mathcal{D} = \frac{\partial}{\partial r}$):

$$(\alpha) \qquad \left.\frac{\partial V}{\partial r}\right|_{r=R} = f(\theta, \lambda). \tag{55}$$

Then (50) turns into

$$-(n+1)R^{-(n+2)}a_{nm} = A_{nm} \tag{56}$$

and therefore

$$a_{nm} = -\frac{R}{n+1}R^{n+1}A_{nm}. \qquad (\alpha - II) \tag{57}$$

Now insertion into (12) gives

$$V(\theta, \lambda, r) = -\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\frac{R}{n+1}\sum_{m=-n}^{n}A_{nm}Y_{nm}(\theta, \lambda). \tag{58}$$

Again a closed solution is possible:

$$\begin{aligned} V(\theta, \lambda, r) &= -\frac{R}{4\pi}\iint\left\{\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\frac{2n+1}{n+1}P_n(\cos\psi)\right\}f(\theta', \lambda')\,d\sigma' \\ &= \frac{R}{4\pi}\iint H(\psi; r, R)f(\theta', \lambda')\,d\sigma'. \end{aligned} \tag{59}$$

The solution is called *Hotine integral* in geodesy (apart from the minus sign).

Type Stokes ($\mathcal{D} = -(\frac{\partial}{\partial r} + \frac{2}{r})$):

$$(\alpha) \qquad -\left(\frac{\partial V}{\partial r} + \frac{2}{r}V\right)_{r=R} = f(\theta, \lambda). \tag{60}$$

The comparison of coefficients yields

$$(n-1)R^{-(n+2)}a_{nm} = A_{nm} \tag{61}$$

and therefore

$$a_{nm} = \frac{R}{n-1}R^{n+1}A_{nm} \qquad \text{for} \qquad n \neq 1. \tag{62}$$

Here a complications arises due to the singularity for $n = 1$. Thus $f(\theta, \lambda)$ has to meet an additional condition for $n = 1$:

$$A_{1m} = \frac{1}{4\pi}\iint f(\theta, \lambda)Y_{1m}(\theta, \lambda)\,d\sigma = 0, \tag{63}$$

whereas the a_{1m} remain undetermined by $f(\theta, \lambda)$; see e.g. [Rummel(1995)]. The solution reads

$$\begin{aligned} V(\theta, \lambda, r) = A_{00} + \frac{R}{4\pi}\iint St(\psi; r, R)f(\theta', \lambda')\,d\sigma' \\ + a_{1,-1}Y_{1,-1}(\theta, \lambda) + a_{1,0}Y_{1,0}(\theta, \lambda) + a_{1,1}Y_{1,1}(\theta, \lambda). \end{aligned} \tag{64}$$

This is the well-known *Stokes integral* formula.
Analytical expressions of D, H and St are summarized in Table 1.
The determination of special case (15) with V independent of λ is completely analogous to the previous one. Throughout it is

$$b_n = c_n = d_n = 0,$$

and we have

$$f(\theta) = \sum_n A_n P_n(\cos\theta) \Leftrightarrow A_n = \frac{2n+1}{2}\int_0^{\pi} f(\theta)P_n(\cos\theta)\sin\theta\,d\theta.$$

Table 1: Closed expressions of integral kernels for the spherical GBVP .

$D(\psi;r,R)$	$\frac{R(r^2-R^2)}{\ell^3}$	H-M 1-89
$H(\psi;r,R)$	$-\frac{2}{\ell}+\frac{1}{R}\ln\frac{\ell+R-r\cos\psi}{r(1-\cos\psi)}$	PPV 1656
$St(\psi;r,R)$	$\frac{2R}{\ell}+\frac{R}{r}-3\frac{R\ell}{r^2}-\frac{R^2}{r^2}\cos\psi(5+3\ln\frac{r-R\cos\psi+\ell}{2r})$	H-M 2-162

Dirichlet:

$$a_n = R^{n+1}A_n,$$

$$V(\theta,r) = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1} A_n P_n(\cos\theta),$$

$$V(\theta,r) = \int_0^{\pi} D(\theta,\theta')f(\theta')\sin\theta'\,d\theta$$

with

$$D(\theta,\theta') = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\frac{2n+1}{2}P_n(\cos\theta)P_n(\cos\theta'). \tag{65}$$

Type Neumann:

$$a_n = -\frac{R}{n+1}R^{n+1}A_n$$

$$\begin{aligned} V(\theta,r) &= -\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\frac{R}{n+1}A_n P_n(\cos\theta) \\ &= \int_0^{\pi} N(\theta,\theta')f(\theta')\sin\theta'\,d\theta \end{aligned}$$

with

$$N(\theta,\theta') = -R\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\frac{2n+1}{2n+2}P_n(\cos\theta)P_n(\cos\theta'). \tag{66}$$

Type Stokes:

$$a_n = \frac{R}{n-1}R^{n+1}A_n \qquad \text{for} \qquad n \neq 1,$$

$$\begin{aligned} V(\theta,r) &= A_0 + \sum_{n=2}^{\infty}\frac{R}{n-1}\left(\frac{R}{r}\right)^{n+1} A_n P_n(\cos\theta) + a_1\cos\theta \\ &= A_0 + \int_0^{\pi} St(\theta,\theta')f(\theta')\sin\theta'\,d\theta + a_1\cos\theta \end{aligned}$$

with

$$St(\theta,\theta') = R\sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n+1}\frac{2n+1}{2n-2}P_n(\cos\theta)P_n(\cos\theta'). \tag{67}$$

Closed analytical expressions for D, N and St (eqs. (65), (66) and (67)) have not been derived.

3.2 Determination of the solution in rectangular coordinates for the upper half space $z \geq z_0$

Only the solution of the Dirichlet and Neumann problem will be given in this case.

$$(\beta) \qquad \lim_{z\to\infty} V = 0 \qquad \text{regularity}$$

requires in (22)

$$b_{k\ell} = 0.$$

It is assumed that the given boundary functions are periodic (chess board pattern) with the same period T in x and y direction and write

$$f(x,y) = \sum_k \sum_\ell A_{k\ell} \exp[i(kx + \ell y)].$$

Type Dirichlet

$$a_{k\ell} = \exp(\sqrt{k^2+\ell^2}z_0)A_{k\ell},$$

$$V(x,y,z) = \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} A_{k\ell} \exp[i(kx+\ell y) - \sqrt{k^2+\ell^2}(z-z_0)],$$

$$V(x,y,z) = \frac{1}{T^2} \int_0^T \int_0^T D(x,y,z;x',y',z') f(x,y)\, dx\, dy$$

with

$$D(x,y,z;x',y') = -\frac{2(z-z_0)}{((x-x')^2+(y-y')^2+(z-z_0)^2)^{3/2}}.$$

Type Neumann:

$$a_{k\ell} = -\frac{1}{\sqrt{k^2+\ell^2}} \exp(\sqrt{k^2+\ell^2}z_0)A_{k\ell},$$

$$V(x,y,z) = -\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \frac{A_{k\ell}}{\sqrt{k^2+\ell^2}} \exp[i(kx+\ell y) - \sqrt{k^2+\ell^2}(z-z_0)],$$

and (see [Kertz(1973), Table 4]):

$$N(x,y,z;x',y') = \frac{2}{\sqrt{((x-x')^2+(y-y')^2+(z-z_0)^2)}}.$$

Special case: V independent of y

$$(\beta) \quad \lim_{z\to\infty} V = 0$$

requires in (28))

$$b_k = 0.$$

Type Dirichlet:

$$a_k = A_k \exp(|k| z_0)$$

$$V(x,z) = \sum_{k=-\infty}^{\infty} A_k \exp[ikx - |k|(z - z_0)]$$

and

$$D(x,z;x') = \frac{2(z - z_0)}{(x - x')^2 + (z - z_0)^2}.$$

Type Neumann:

$$a_k = -\frac{1}{k} A_k \exp(|k| z_0) \qquad \text{for} \qquad k \neq 0$$

$$V(x,z) = -\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \frac{A_k}{k} \exp[ikx - |k|(z - z_0)]$$

and

$$N(x,z;x') = \ln[(x - x')^2 + (z - z_0)^2].$$

3.3 Determination of the solution in circular cylinder coordinates for the outside of the boundary cylinder $r \geq R$

Also in this case only the solutions for Dirichlet and Neumann are given.

$$(\beta) \qquad \lim_{r\to\infty} V = 0 \qquad \text{regularity}$$

requires in (38) that

$$b_{k\ell} = 0.$$

Type Dirichlet:

$$a_{k\ell} = \frac{1}{K_m(kR)} A_{km}$$

$$V(r,\lambda,z) = \int_k \sum_m A_{km} \frac{K_m(kr)}{K_m(kR)} \exp[i(m\lambda + kz)]\, dk$$

Type Neumann:

$$\left.\frac{\partial V}{\partial r}\right|_{r=R} = \int_k \sum_m a_{km} \exp[i(m\lambda + kz)] \frac{ik}{2} (K_{m-1}(kR) - K_{m+1}(kR))\, dk;$$

see e.g. [Lebedev(1965)]

$$\Rightarrow a_{km} = A_{km} \frac{2}{ik(K_{m-1}(kR) - K_{m+1}(kR))}.$$

Special case: V independent of z (circle)

$$(\beta) \qquad \lim_{r\to\infty} V = 0 \qquad \text{regularity}$$

requires in (45) that

$$c_1 = 0 \quad \text{and} \quad b_m = 0.$$

Type Dirichlet:

$$a_m = R^{|m|} A_m,$$

$$V(r,\lambda) = \sum_{m=-\infty}^{\infty} \left(\frac{R}{r}\right)^{|m|} A_m \exp(im\lambda) \qquad \text{and} \quad a_0 = c_0$$

$$V(r,\lambda) = \frac{1}{2\pi} \int_0^{2\pi} \frac{r^2 - R^2}{\ell^2} f(\lambda)\, d\lambda$$

i.e.

$$D(r,\lambda;r',\lambda') = \frac{r^2 - R^2}{\ell^2}$$

with

$$\ell^2 = r^2 + R^2 - 2rR\cos(\lambda - \lambda').$$

Type Neumann:

$$a_m = -\frac{1}{|m|} R^{-(|m|+1)} A_m \qquad (m \neq 0)$$

$$V(r,\lambda) = -\sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{R}{|m|} \left(\frac{R}{r}\right)^{|m|} \exp(im\lambda)$$

and

$$N(r,\lambda;R,\lambda') = -R \ln \frac{r^2}{r^2 + R^2 - 2rR\cos(\lambda - \lambda')}.$$

4 Discussion

- For the determination of the solutions of the GBVPs in the various coordinate systems boundary conditions of the type

  $$\mathcal{D}V(x^1, x^2, x^3 = const) = f(x^1, x^2)$$

  have to be available on the surface $x^3 = const$. In geodesy this condition often results from the generalized Bruns transformation, as shown in [Grafarend(1980),]. The boundary surface is actually the telluroid (determined by some mapping). As the telluroid is a surface too complicated, it is approximated in practise by an ellipsoid, sphere or a tangent plane.

- The linear differential operator $\mathcal{D}$ requires often to consider boundary value problems different from the classical Dirichlet and Neumann ones. Even the Stokes boundary condition is not a classical boundary value problem of the third kind (Robin or Poincaré). Each of them requires, therefore, careful analysis of their singularities.

- Three special cases have been included. The first one treats V as a function of $\{\theta, r\}$ only: $V(\theta, r)$. It is convenient, because it allows to build up a physical geodesy without λ-dependence. Even satellite trajectories can be included. It has been dealt with in the dissertations by [Gerontopoulos(1978)] and [Van Gelderen(1991)]. It has also the advantage to be extendable to the ellipsoidal case without major complications.

  The second special case is two-dimensional cartesian $V(x, z)$. Thus, the field is assumed to be invariant in y-direction: $\frac{\partial^2 V}{\partial y^2} = 0$. This model is very popular in geophysics and applied there at many instances. It permits to demonstrate all principles of the much harder three-dimensional case but leads to very simple Fourier series. It is applied in a very convincing manner throughout in [Turcotte and Schubert(1982),].

  The third special case is derived from the circular cylinder coordinates: $V(\lambda, r)$. It leads to boundary value problems inside and outside a boundary circle in the plane. Again it is extremely simple but allows to demonstrate essential features. It is employed, for example, by [Walter(1971)].

References

[Van Gelderen(1991)] Gelderen, M. van (1991). The geodetic boundary value problem in two dimensions and its iterative solution. , Netherlands Geodetic Commission, Publications on Geodesy, New Series, **35**.

[Gerontopoulos(1978)] Gerontopoulos, P. (1978). Molodensky's problem in the plane. , Mitteilungen der geodätischen Institute der TU Graz.

[Grafarend(1980)] Grafarend, E. (1980). The Bruns transformation and a dual setup of geodetic observational equations. report NOS 85, NGS 16, NGS, Rockville.

[Heiskanen and Moritz(1967)] Heiskanen, W. and Moritz, H. (1967). *Physical Geodesy.* Freeman and Co, San Fransico.

[Hotine(1969)] Hotine, M. (1969). *Mathematical geodesy.* US Department of Commerce, Washington, D.C.

[Kertz(1973)] Kertz, W. (1973). *Potentialtheorie in die Geophysik.* Vorlesungsmanuskript, Technische Universität Braunschweig.

[Lebedev(1965)] Lebedev, N. (1965). *Special functions and their applications.* Prentice-Hall.

[Lehner(1996)] Lehner, G. (1996). *Elektromagnetische Feldtheorie.* Springer, Berlin.

[Moon and Spencer(1961)] Moon, P. and Spencer, D. E. (1961). *Field theory handbook.* Springer, Heidelberg.

[Pick *et al.*(1973)] Pick, M., Picha, J., and Vyskočil, V. (1973). *Theory of the earth's gravity field.* Elsevier, Amsterdam.

[Rummel(1995)] Rummel, R. (1995). The first degree harmonics of the stokes problem – what are the practical implications? In *Festschrift Erwin Groten*, pages 98–106. München.

[Rummel and Teunissen(1986)] Rummel, R. and Teunissen, P. (1986). Geodetic boundary value problem and linear inference. In *International Symposium Figure and Dynamics of the Earth, Moon and Planets*, pages 227–264, Prague.

[Rummel and Van Gelderen(1992)] Rummel, R. and Van Gelderen, M. (1992). Spectral analysis of the full gravity tensor. *Geophys. J. Int.*, **111**(1), 159–169.

[Rummel and Van Gelderen(1995)] Rummel, R. and Van Gelderen, M. (1995). Meissl scheme – spectral characteristics of physical geodesy. *J. of Geod.*, **20**, 379–385.

[Rummel and Van Gelderen(1999)] Rummel, R. and Van Gelderen, M. (1999). The solution of the general least squares geodetic boundary value problem by least squares. *Submitted to J. of Geodesy.*

[Rummel *et al.*(1989)] Rummel, R., Teunissen, P. J. G., and Van Gelderen, M. (1989). Uniquely and overdetermined geodetic boundary value problems by least squares. *Bull. Géod.*, **63**, 1–33.

[Turcotte and Schubert(1982)] Turcotte, D. and Schubert, G. (1982). *Geodynamics; applications of continuum physics to geological problems.* John Wiley & Sons, New York.

[Walter(1971)] Walter, W. (1971). *Einfürung in die Potentialtheorie.* Bibliographisches Institut, 765/765a, Mannheim.

Quo vadis geodesia? ... Sic erit pars publica

Volker Schäfer

Das Symposium mit der Diskussion zur Zielbestimmung für die wissenschaftliche Geodäsie wäre aus der Sicht eines Angehörigen der amtlichen Vermessungswesens Baden-Württemberg unvollständig, wenn sie dieses nicht auch eines Seitenblicks würdigen würde.
Das amtliche Vermessungswesen mit seiner gesetzlichen Aufgabe, die für ein Staatswesen allgemeingültigen Geo-Bezugssysteme festzulegen und sie auf der Erde vorzuzeigen, ist Partner und Nutznießer der geodätischen Wissenschaft, aber durchaus auch Fordernder.

Die Geschichte des heute noch maßgeblichen Lagebezugssystems in Deutschland kann als bekannt vorausgesetzt werden. Sie war von der möglichsten Einheitlichkeit des Deutschen Hauptdreiecksnetzes im Sinne der Vorstellungen des Deutschen Reiches nach 1870 geprägt. Sie litt unter der Teilung Deutschlands und der Ausrichtung in einem seiner Teile nach Osten.
Das Einheitliche Astronomisch-Geodätische Netz der Staaten des ehemaligen Warschauer Paktes unter Nutzung der Beobachtungen des Reichsdreiecksnetzes mit der Datumsfestlegung Pulkowo und in Verbindung mit dem Referenzellipsoid von Krassowskij (System 42/83) weicht erheblich vom DHDN ab. Die Differenzen, bezogen auf ETRF89, sind laut Ihde, ZfV 1994 Seite 192 ff bereits in den Translationen u, v und w in den Achsen X, Y und Z:

DHDN	->	ETRF	+582 m,	+105 m	und	+414 m
S 42	->	ETRF	+24 m,	-123 m	und	-94 m

Die Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) in der Gestalt der neuen Einheit seit 1990 haben die Chancen genutzt, welche die erforderliche Zusammenführung des DHDN und des Systems 42/83 bot. Sie haben die Technologie der weltweiten Satellitenmessverfahren eingesetzt und in wenigen Jahren den Grundstein für ein deutschlandeinheitliches und vielleicht noch großräumigeres Bezugssystem ETRS89 gelegt.
Zugleich haben sie sich in Kenntnis der Anforderungen und Mühen entschieden, ETRS89 mit den Punkten des ETRF89 zu lokalisieren und zum künftigen Bezugssystem für Landesvermessung und Liegenschaftskataster zu bestimmen. Darüberhinaus ist die AdV 1996 über die Innenministerkonferenz an die Bundesregierung mit der Bitte herangetreten, in Brüssel für ein europaweit einheitliches Bezugssystem ETRS89 und Abbildungssystem UTM einzutreten. Die Bundesregierung hat dem Ersuchen entsprochen; Resonanzen seitens der EU sind noch nicht erkennbar. Spätestens mit dem europäischen Galileo-Satellitennavigationsplan wird ein einheitliches Bezugssystem überfällig.

Baden-Württemberg wird den Beschluss der AdV von 1995 zur Einführung und Anwendung von ETRS89 und UTM umsetzen:

„...

3. Für die Aufgaben der Grundlagenvermessung wird das Bezugssystem ETRS89 Zug um Zug wie folgt eingeführt:

a) Lagefestpunktfeld
Im Lagefestpunktfeld wird das neue Bezugssystem zunächst zusätzlich zu den bestehenden Systemen entsprechend den vom Plenum der AdV bestätigten Festlegungen des Arbeitskreises „Grundlagenvermessung" kurzfristig eingeführt. Die zuständigen Landesvermessungsbehörden stellen im ETRS 89 für hinreichend viele Punkte dreidimensionale Koordinaten bereit, so daß neue und alte Messungen bei Bedarf im ETRS 89 ausgewertet oder alternativ Koordinaten aus den bisherigen Systemen in das ETRS 89 transformiert werden können.

b) Höhenfestpunktfeld und Schwerefestpunktfeld
Bei der Berechnung von Normalhöhen im System des DHHN 92 sind die Parameter des GRS 80 und die Nivellementpunkt-Koordinaten im System des ETRS 89 zu verwenden. Zur Bestimmung von Parametern für 3D-Koordinatentransformationen sind mindestens die Punkte des DREF-Netzes (Hierarchiestufe B), erforderlichenfalls auch die Punkte der Hierarchiestufe C (z.B. BWREF) einzunivellieren.

4. Das derzeit verwendete Blattschnittsystem der analogen Topographischen Kartenwerke bleibt vorerst unverändert bestehen. Das Bezugssystem ETRS 89 wird baldmöglichst in den analogen Topographischen Kartenwerken in den Maßstäben 1:10 000 und kleiner eingeführt, indem das Koordinatengitter des UTM-Systems angegeben wird. Eine Umstellung auf neue runde Blatteckwerte kann erst in Frage kommen, wenn die analogen Kartenausgaben aus ATKIS gewonnen werden können.
5. Es ist erforderlich, die Basisinformationssysteme „Automatisierte Liegenschaftskarte (ALK)", „Automatisiertes Liegenschaftsbuch (ALB)" und „Amtliches Topographisch-Kartographisches Informationssystem (ATKIS)" in demselben Bezugs- und Abbildungssystem zu führen. Länder, in denen die Einführung des ETRS 89 und UTM bereits jetzt geboten ist, können kurzfristig auf diese Systeme umstellen. Die anderen Länder führen ihre Arbeiten bis zur vollständigen Einrichtung von ALK, ALB und ATKIS im bisherigen System fort und stellen spätestens danach die Basisinformationssysteme geschlossen auf ETRS 89 und UTM um.
6. Um Anforderungen überregionaler Nutzer auf Bereitstellung von ALK- und ATKIS-Daten im Bezugssystem ETRS 89 und im Abbildungssystem UTM erfüllen zu können, soll von den Arbeitskreisen „Informations- und Kommunikationstechnik", „Liegenschaftskataster" und „Topographie und Kartographie" untersucht werden, wie diesen Nutzerwünschen entsprochen werden kann.

...".

Wie umfangreich die Aufgabe eines Systemwechsels ist, ergibt sich aus wenigen auf Baden-Württemberg bezogenen Angaben: Im Land existieren jeweils rund 60 000 Punkte des amtlichen Lage- und Höhensystems und 120 DREF- und BWREF-Punkte, dazu kommen rund 700 000 Aufnahmepunkte als Verdichtungsstufe in der Lage. Nur geschätzt werden können die koordinierten Grenzpunkte, Gebäudeeckpunkte und Punkte der Katastertopographie mit Gauß-Krüger-Koordinaten; es sind in etwa 80 Millionen in der staatlichen Vermessungsverwaltung. Mit der Fertigstellung der digitalen Liegenschaftskarte und des Festpunkt-Informationssystems sind sie alle digital verfügbar.

Es ist im Lande zeitweise diskutiert worden, ob der AdV-Beschluss im Hinblick auf die in den letzten Jahrzehnten energisch vorangetriebene Einrichtung des Aufnahmepunktfeldes mit dem Ziel eines flächendeckenden Koordinaten-Katasters nicht finanziell zu aufwendig und sachlich sogar überflüssig ist. Ein Blick auf die Transformation DREF91- DHDN mit dem identischen Punkt Potsdam (Anlage 1 - Verschiebungsvektoren im Maßstab ca. 1 : 300) mit globalen Transformationsparametern zeigt, was der verantwortliche Umgang des Vermessungswesens der Gegenwart mit seiner Zukunft gebietet.
Auch die Verhältnisse in Baden-Württemberg auf der Basis einer Transformation anhand von identischen Punkten (Anlage 2 - Verschiebungsvektoren im Maßstab ca. 1 : 10) bestätigen die Notwendigkeit eines zukunftsorientierten globalen Bezugssystems.

Die Vermessungsverwaltung Baden-Württemberg hat ein digitales Modell der Landestopographie (Digitales Landschaftsmodell DLM) erstellt sowie eine Version im Datenumfang der Topographischen Karte 1 : 25 000 (ohne Gebäude und Höhen) auf den Markt gebracht. Die dritte Dimension (Digitales Höhenmodell DHM) wird derzeit ausgeschrieben und soll möglichst 2000 in Gestalt eines Höhenrasters mit generell 10 m-Abstand und 2 m-genau für Boden sowie Vegetations- und Bebauungsobergrenze vorliegen; auf Anforderung soll der Datenbestand bis auf 0,50 m aufgelöst werden.
In der Liegenschaftskarte ALK sind bereits für 85 % des Flurstücksbestands Eigentum und Nutzung digital dokumentiert; bis 2005 wird auch der Rest umgesetzt sein. Die entsprechende textliche Liegenschaftsbeschreibung liegt schon seit 1994 vollständig digital vor. Die Kommunikation

zwischen Vermessungsverwaltung und Datennutzer wird auf Internet-Technologie umgestellt:
Es bestehen in Baden-Württemberg somit beste Voraussetzungen für das Vorhaben. Dies würde den genannten AdV-Beschluss zwingend erforderlich machen, läge er nicht bereits vor.

Wer gleichwohl noch zweifelte, dem wurde durch den Beschluss der Landesregierung Baden-Württembergs vom Oktober 1998 zur flächendeckenden Einrichtung des Satellitenpositionierungsdienstes der AdV (SAPOS-Bezugssystem) mit vernetzten SAPOS-Festpunkten bis 2001 wohl endgültig deutlich, dass der Weg zu einem zukunftstauglichen Lagebezugssystem konsequent fortgesetzt wird. Und mit den im Juli 1999 von den Landesregierungen von Baden-Württemberg und Bayern beschlossenen Vernetzung und Standortplanung an der Landesgrenze und gegenseitigem Datentransfer entsteht ein, auch wirtschaftlich interessanter, flächendeckender SAPOS-Südblock.

Die Einrichtung von SAPOS liegt vor allem in wirtschafts- und industriepolitischer Hinsicht, daneben aber ebenso im Hinblick auf die Eigenanwendung der Vermessungsverwaltung, im Interesse des Landes:

- Ein satellitengestütztes Navigations- und Ortungssystem mit den Funktionalitäten von SAPOS ist Teil einer modernen informationstechnischen Infrastruktur und daher für den Wirtschaftsstandort Baden-Württemberg unverzichtbar.
- Ein verlässlicher Markt für satellitengestützte Navigationsprodukte in Hard- und Software ist Vorbedingung für die Industrie, in die Entwicklung der nächsten Generation dieser zukunftsträchtigen Navigationstechnologie zu investieren.
- SAPOS wird Teil des amtlichen geodätischen Bezugssystems der Landesvermessung werden und dadurch eine nachhaltige Kostensenkung im Vermessungswesen bewirken. Die Vermessungsverwaltung kommt damit ihrer Verpflichtung zur Aufgabenerledigung mittels moderner, zukunftsweisender Technologie nach.

Die getroffene Entscheidung der Landesregierung sichert zum einen auf Dauer den bislang pilothaft vom Landesvermessungsamt und dem Südwestrundfunk mittels UKW/RDS-Radio betriebenen Datendienst für differentielles GPS im Bereich von wenigen Metern. Von der Verfügbarkeit dieses SAPOS-Dienstes gehen insbesondere die Verkehrsleitsysteme (Telematik und Auto-Navigation), das Flottenmanagement in der Produktelogistik und die Hilfsdienste aus.
Zum anderen wird binnen kurzem ein zentimetergenauer Datendienst mit der Möglichkeit für eine sofortige Auswertung an Ort und Stelle im deutschlandeinheitlichen amtlichen Bezugssystem der Landesvermessung für Süddeutschland dauerhaft garantiert.

Als Folge werden die Gauß-Krüger-Koordinaten (DHDN) von UTM-Koordinaten (DREF) abgelöst. Es wird keine systematische Verdichtung des BWREF erfolgen; das gewohnte herkömmliche Arbeiten „vom Großen ins Kleine" entfällt also. Eine weitergehende flächenhafte und bedarfsunabhängige Verdichtung des Aufnahmepunktfeldes ist eingestellt.

Mit diesen Weichenstellungen einher geht auch die Konsequenz für die Messtechniken der Zukunft: Für viele der derzeit zigtausend markierten Festpunkte entfällt durch die etwa 20 elektronischen SAPOS-Festpunkte die Notwendigkeit zur Unterhaltung. Und auch die altbekannte Hierarchie nach Ordnungen wird sich weitgehend auflösen: TP und AP, für die keine ETRS-Koordinaten benötigt werden, verlieren ihre bis dato praktisch unbegrenzt geltende Verfügbarkeitsgarantie. Daran wird sich die Technik der Positionsbestimmung auszurichten haben.

Mit der Integration von hochgenauen GPS-Bestimmungen in das spannungsbehaftete Landeskoordinatensystem stellt sich konkret die Frage, ob die geodätischen Netze nunmehr nicht dynamisch aufgefasst werden müssen. In den Verwaltungen war bislang unter der Betrachtungsweise der sog. Nachbarschaftstreue die statische Betrachtungsweise bestimmend. Für Baden-Württemberg wurde 1998 die Ausgleichung mit beweglichen Anschlusspunkten als Regelverfahren im Aufnahmepunktfeld eingeführt.
Im amtlichen Höhenbezugssystem für Deutschland ist die Entscheidung für die Normalhöhen

gefallen; sicher ein sich noch fortsetzender Anlass für entschiedenes Pro und Kontra in sachkundiger Bewertung.
Die AdV hat dafür (auszugsweise) folgende Aussagen mit ihrer Entscheidung verbunden:

- Die Ausgleichung erfolgt zwangsfrei in geopotentiellen Koten.
- Das Niveau des DHHN92 wird festgelegt durch Übernahme der geopotentiellen Kote des Knotenpunktes Wallenhorst (bei Osnabrück) aus dem derzeit gültigen europäischen Nivellementnetz REUN86, welches am Nullpunkt des ehemaligen Pegels Amsterdam angeschlossen ist.

Daraus ergeben sich (laut Weber, ZfV 1995 Seite 196 ff) u.a. folgende Vorteile für die Benutzer :

- Normalhöhen sind optimal mit ellipsoidischen Höhen kombinierbar. Dies ergibt sich daraus, dass sie geodätisch bestimmte Abstände von dem Quasigeoid sind, das unter Verwendung von Parametern des GRS80 zu berechnen ist und durch den Nullpunkt des ehemaligen Amsterdamer Pegels verläuft.
- Falls z.B. für den Wasserbau exakte Angaben über den Verlauf einer Niveaufläche benötigt werden, kann auf vorliegende geopotentielle Koten der Nivellementpunkte mit Normalhöhen zurückgegriffen werden.

Für die Nivellementpunkte 1. Ordnung liegen bereits Normalhöhen vor; die für die Berechnung der Normalhöhen für die 2. und 3. Ordnung erforderlichen Schweremessungen werden zügig vorangetrieben. Nach Plänen der AdV sollen ab 2002 die Punkthöhen grundsätzlich in dem neuen deutschlandeinheitlichen Bezugssystem DHHN92 bekanntgegeben werden.

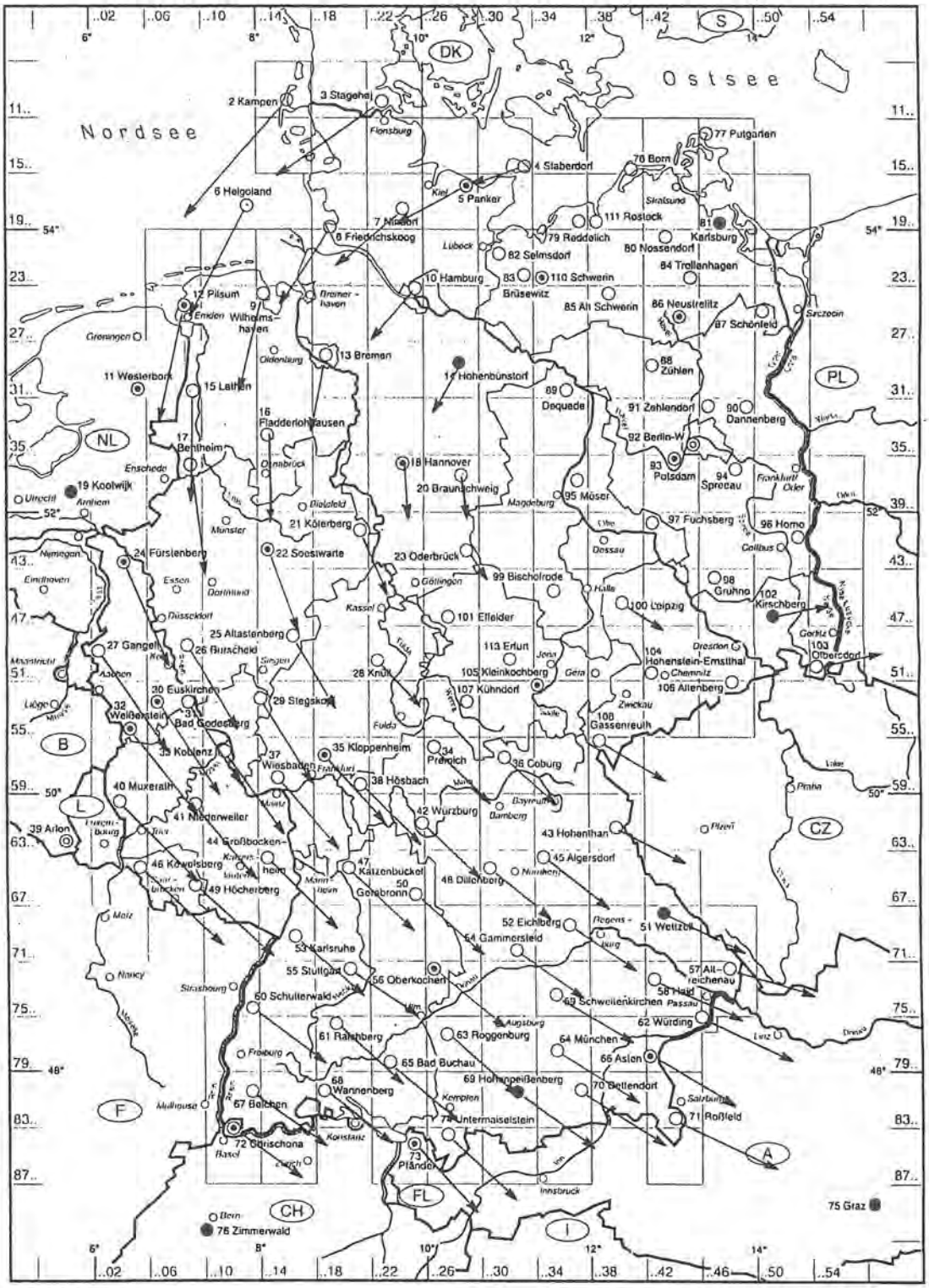
In anderen Angelegenheiten ist die Vermessungsverwaltung zugegebenermaßen durchaus von der Fragestellung „Quo vadis geodesia?" berührt:
Bezüglich wirtschaftlicher und zugleich präziser Höhenmessverfahren blickt die Landesvermessung erwartungsvoll in Richtung der geodätischen Lehrstühle in der Hoffnung auf Unterstützung. Was die Verwaltungen für die Zukunft benötigen, sind Vorgehensweisen zur GPS-gestützten Höhenbestimmung. Die bisherigen Messmethoden sind zu aufwendig und kommen für notwendige, bisher jedoch nicht periodisch mögliche Wiederholungs-Messkampagnen nicht in Frage. Andererseits ist die Kenntnis über Bewegungen der Erdoberfläche und Massenverschiebungen nicht nur in extrem erdbebengefährdeten Gebieten unerlässlich.

Untersuchungen zur Verwendbarkeit der GPS-Messverfahren zur qualifizierten Höhenbestimmung (z.B. Lux, Strauss, AVN 1998 Seite 291 ff) kommen auf Höhengenauigkeiten von unter 1 cm bis 2 cm mit durchaus aber größeren einzelnen Abweichungen. Vorschläge, im Nivellementnetz unterhalb der 2. Ordnung die bisherigen Genauigkeitsanforderungen herabzusetzen und so den Einsatz von GPS zur Gebrauchshöhenbestimmung zu ermöglichen, entschärfen das Problem des Aufwands bei konventionellen Präzisionsnivellements für periodische Wiederholungsmessungen in den Nivellementnetzen 1. und gegebenenfalls 2. Ordnung nicht.
Ein anderer wichtiger Themenkomplex darf ob der technischen Fragestellungen aber keinesfalls unerwähnt bleiben.
Mit einiger Sorge beobachtet die Vermessungsverwaltung von Baden-Württemberg Bestrebungen zur Verlagerung der Gewichte der Lehrinhalte in den Studienordnungen. Es wird nicht in Frage gestellt, dass der Arbeitsmarkt Öffentlicher Dienst für Geodäten bei weitem nicht mehr so aufnahmefähig ist wie in der Vergangenheit. Aber nach Ansicht von Praktikern wird der Bedarf an aufeinander abgestimmten sowie bedarfsorientierten Geodaten günstige Berufschancen für Geodäten eröffnen.
Die Voraussetzung für sich am Markt behauptende „Geo-Datenprovider" sind neben qualifiziertem geodätischen und informationstechnischem Grundlagenwissen insbesondere auch ein ebensolches von Wesen und Aussagekraft amtlicher Geodaten und ihrer Verwendung in Planung, Wirtschaft und Verwaltung. Insofern ähneln sich die Anforderungen an die Studieninhalte bei den Vermessungsverwaltungen und dem allgemeinen Arbeitsmarkt sehr stark.
Die Universitäten sind aufgerufen, den Spagat zwischen Studiendauer und weitgefächertem Berufsfeld zu beherrschen.

Anlage 1

DREF 1991
Deutsches Referenznetz im ETRS 89

"Globale" Transformation vom DREF91 ins DHDN mit dem ident. Punkt Potsdam



Anlage 2

Restklaffungen der Transformation DREF und DHDN



Maßstab der Vektoren ~ 1:10

Reproducing Estimators via Least Squares: An Optimal Alternative to the Helmert Transformation

— dedicated to Erik Grafarend on his 60th birthday —

Burkhard Schaffrin

Abstract

In the tradition of several joint papers by the honoree and myself, including E. Grafarend and B. Schaffrin (1974, 1976, 1988) and E. Grafarend, B. Schaffrin, and E. Knickmeyer (1982), I took up the subject again in B. Schaffrin (1984, 1989), N. Tamim and B. Schaffrin (1995) and B. Schaffrin and J. Cothren (1998), now with the emphasis on network densification methods which are rigorous, but which leave tie points unchanged along with their variances and covariances. Such a requirement led to the notion of "reproducing estimators" as first presented by B. Schaffrin (1997).

It has been known for a long time that the so-called "dynamic adjustment" of geodetic networks can be interpreted as least-squares collocation, according to H. J. Buiten (1978), which would not yield estimates with the "reproducing property." Also the "free net adjustment" followed by a Helmert transformation would not per se guarantee identical tie point coordinate estimates unless we "forget" about the deviations whose norm we had minimized, but usually not to zero. So the only "reproducing" alternative seemed to be given by the formulas of a classical least-squares adjustment with fixed constraint even though the tie points are "fiducial" in their nature, having typically a full dispersion matrix. Therefore, in the formula for the corresponding variance-covariance matrix we find an additional term which vanishes as soon as the "fiducial" points become real "fixed" points; see, for instance, B. Schaffrin (1984, 1989). This apparently has been the method, applied by JPL over the years, when they talked about "fiducial network strategies" as in S. M. Lichten et al. (1989), for instance, without spelling out exactly what they did.
In any case, as could have been expected, the latter one was shown not to be the optimal estimator with the "reproducing property" by B. Schaffrin (1997). Instead, the optimum turned out to be the "forgetful dynamic adjustment" where every densification point is found conventionally through least-squares collocation, but the tie points remain unchanged. Consequently, a free adjustment followed by a "forgetful Helmert transformation" must be considered inferior and may no longer be carried out routinely; see also K. R. Koch (1983) and F. W. O. Aduol (1993). Here we try to employ least-squares methodology to generate that "reproducing estimator" which, as we already know, will be optimal in this class.

References

Aduol, F. W. O. (1993): A static-dynamic model for densification of geodetic networks, *AVN-Intl. Edition* 10, 20–36.

Buiten, H. J. (1978): Junction of nets by collocation, *manus. geodaet.* 3, 253–297.

Grafarend, E., E. Knickmeyer, and B. Schaffrin (1982): Geodetic datum transformations (in German), *ZfV* 107, 15–25.

Grafarend, E. and B. Schaffrin (1974): Unbiased free net adjustment, *Survey Review* 22, 200–218.

Grafarend, E. and B. Schaffrin (1976): Equivalence of estimable quantities and invariants in geodetic networks, *ZfV* 101, 485–491.

Grafarend, E. and B. Schaffrin (1988): From the static to the dynamic treatment of geodetic networks (in German), *ZfV* 113, 79–103.

Koch, K. R. (1983): The choice of the datum of a trigonometric network for the determination of new points (in German), *ZfV* 108, 104–111.

Lichten, S. M., W. I. Bertiger, and U.J. Lindqwister (1989): The effect of fiducial network strategy on high-accuracy orbit and baseline determination; in W.L. Stein and M. Kumar (eds.), Proc. of the 5th International Geodetic Symp. on Satellite Positioning, Las Cruces, New Mexico, 1989, Vol. I, 516–525.

Schaffrin, B. (1984): On the use of prior information in free networks (in German); in K. Rinner et al. (eds.), Ingenieurvermessung '84, Dümmler: Bonn, B6/1–15.

Schaffrin, B. (1989): Fiducial versus fixed points in the GPS network approach; in W.L. Stein and M. Kumar (eds.), Proc. of the 5th International Geodetic Symp. on Satellite Positioning, Las Cruces, New Mexico, 1989, Vol. I, 500–511.

Schaffrin, B. (1997): On suboptimal geodetic network fusion, Poster paper, IAG General Meeting, Rio de Janeiro, Brazil, Sept. 1997.

Schaffrin, B. and J. Cothren (1998): Towards optimizing hierarchical data revisions, Proc. of the International Symp. on GIS: Between Visions and Applications, Stuttgart, Germany, Sept. 1998, 515–521.

Tamim, N. and B. Schaffrin (1995): A Methodology to create a digital cadastral overlay through upgrading digitized cadastral data, *Surveying & LIS* 55, 3–12.

Seasonal and Short-Period Fluctuations of Earth Rotation Investigated by Wavelet Analysis

Harald Schmitz-Hübsch and Harald Schuh

Summary

When compared with classical Fourier or spectral analysis there are several remarkable advantages of the wavelet analysis. In particular, it allows the time localisation of an unstable quasi-harmonic signal within a given data set. This paper concentrates on seasonal variations and on the short-period, i.e. subseasonal fluctuations of Earth rotation. The analysis of length of day (lod) series yields in the high frequency range periods of 28 days, 14 days down to 5.6 days caused by the lunisolar tides and irregular periodic variations between 40 and 150 days. These are mainly excited by global zonal winds. For the semi-annual variations of lod a correlation with El Niño events and with the antarctic circumpolar current can be seen. The main seasonal components of polar motion are the prograde annual and semi-annual variations. Both get maximal amplitudes one to two years after strong El Niño events. Additionally, variable periods in polar motion with periods between two and five months and even down to periods of 8-10 days can be seen in the wavelet spectrum of the short-period range.

1. Introduction

The Earth rotation vector is variable with time, both in its direction and in its length. The directional change of the rotation axis with respect to an Earth-fixed reference system is called polar motion and has a magnitude of several meters on the Earth' surface. Polar motion consists mainly of two periodic components, one is the Chandler wobble with a period of about 1.18 years and the other one is the annual wobble. The Chandlerian part of polar motion is caused by the deviation of the rotation axis from the Earth's polar axis of inertia whereas the annual variation and variations with periods shorter than one year are mainly due to the interactions between the atmosphere, the hydrosphere and the solid Earth. The length of a mean solar day (length of day, lod) depends on the rotation of the Earth, i.e. changes of the Earth' rotational velocity with respect to a quasi constant velocity derived from atomic clocks cause small changes of lod. The lod time series consist mainly of a secular trend, long-period variations and seasonal variations with an annual and a semi-annual period. In the last years the short-period variations of lod with periods down to a few hours have become an important scientific subject.

As the total angular momentum vector of the Earth, the atmosphere and the hydrosphere (including oceans) is temporally constant (apart from minor torques of the sun and the moon), there is a close relation between meteorological and hydrological processes and the rotational behaviour of the Earth. The atmospheric perturbations of polar motion can be expressed by the equatorial components of the atmospheric angular momentum (AAM) excitation functions χ_1, χ_2 and those of the length of day (lod) by the zonal component χ_3, respectively (Barnes et al., 1983).

It has been shown in several publications that a strong correlation exists between lod and the total AAM (Nastula et al., 1992; Rosen, 1993; Kolaczek, 1995; Zatman, 1997). The atmospheric contribution to the variations of polar motion in time scales between several weeks and a few months has also been demonstrated by many authors (Brzezinski, 1987; Eubanks et al., 1988; Eubanks, 1993; Kuehne et al.,

1993; Kosek, 1995; Kosek et al., 1995; Kosek and Kolaczek, 1997; Nastula, 1997; Nastula et al. 1997). In recent time the important contribution of the oceans on the Earth rotation parameters has been reported, too, e.g. by Chao et al. (1995), Chao et al. (1996) and by Chao and Zhou (1999).

In the last years, the wavelet transform as an adaptive time-frequency analysis has been success- fully applied in geodesy and geophysics. Examples are given by Gambis (1992), Popinski and Kosek (1994), Chao and Naito (1995), Baudin et al. (1996), Schmidt (1996), Praveen (1997), Gibert et al. (1998), Schmidt and Schuh (1999), Schmitz-Hübsch (1999), Schuh and Titov (1999). In our investigations the Morlet Wavelet transform will be used to study the time-variability of the instable quasi-periodic oscillations of the Earth rotation parameters for periods shorter than one year. The time-variable wavelet covariance (e.g. Torrence and Compo, 1998; Whitcher, 1998) between the geodetic signals and the atmospheric excitation functions will be regarded, too.

2. Data

We analysed the EOP(IERS) C 04 data series of the pole coordinates and of lod from 01.01.1962 till 31.03.1999 provided by the International Earth Rotation Service (IERS), Paris. These are equidistant daily values with the very short periods filtered out (below 8d till 1988, below 3d till 1993, below 2d since 1994). Atmospheric Angular Momentum (AAM) series were also analysed which were derived from global meteorological data (series *aam.ncep.reanalysis.1958.1999* computed by NCEP, National Center for Environmental Prediction and published by IERS). They represent the atmospheric influence on x_p, y_p and lod due to global wind forcing and global air pressure variations. In our computations both the pole coordinates x_p, y_p and the equatorial atmospheric excitation χ_1, χ_2 were treated in the complex plane $(x_p - i \cdot y_p)$, $(\chi_1 + i \cdot \chi_2)$, respectively.

3. Method of Analysis

The Fourier transform is a suitable tool for analysing periodic signals like ocean tides or Earth tides. The signal is compared with harmonic oscillations of different frequencies. The Fourier transform gives the information about the frequency content of the signal without any time localisation. Thus, time-varying amplitudes and/or periods can hardly be detected.

There are several remarkable advantages of the wavelet transform when compared with 'classical' methods based on the Fourier transform. In the wavelet transform the harmonic oscillation is replaced by a wavelet function which is adapted to the signal to be analysed by shifting it along the time axis and scaling it along the frequency axis. It allows the time localisation of an unstable quasi-harmonic signal within a given data set. Hence, the wavelet transform is an excellent tool for analysing signals with time-varying amplitudes and/or periods. The wavelet transform was introduced by Grossmann and Morlet (1984). With respect to a wavelet by which the data are analysed it has the form:

$$W_\psi \left[f(t)\right](a,b) = \frac{1}{\sqrt{|a|}} \int f(t) \cdot \overline{\psi\left(\frac{t-b}{a}\right)}\, dt$$

$$a \neq 0,\ a, b \in \mathbb{R}$$

with
Ψ (t) – non-stationary part of a wave (= wavelet) and the wavelet basis a and b:

a – dilation (scale parameter)
b – translation (shift parameter)

The analysing wavelet Ψ (t) is a localised oscillating function defined on the real line. Here, the complex Morlet wavelet will be used:

$$\psi(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{ipt} \cdot \left[e^{-t^2/\left(2\sigma_0^2\right)} - \sqrt{2} \cdot e^{-t^2/\sigma_0^2} - (p\cdot\sigma_0)^2/4 \right]$$

with
p - variable (usually $p > 5$; in this paper $p = 2\pi$ was used),
σ_0 - parameter which describes the decay of the Morlet wavelet.

The wavelet cross-spectrum will also be regarded. As it is a complex function, we will compute the *squared* wavelet cross-spectrum between the lod-series f(t) and the AAM-series g(t) according to Schmidt (1999):

$$\left|W_\psi \left[f(t),g(t)\right](a,b)\right|^2 = W_\psi \left[f(t)\right](a,b) \cdot \overline{W}_\psi \left[f(t)\right](a,b) \cdot W_\psi \left[g(t)\right](a,b) \cdot \overline{W}_\psi \left[g(t)\right](a,b)$$

This is equivalent to the cross-wavelet power (Torrence, 1998).

The Morlet wavelet is well-suited for the analysis of Earth rotation parameters which - depending on the excitation mechanism - partly show very clear and stable periods but contain also unstable, spurious, quasi-periodic variations. As described above, the wavelet transform allows the exact localisation by time which usually cannot be seen by the classic Fourier analysis. Furthermore, the choice of different wavelets offers a lot of flexibility for the analyst. Generally it can be said that the localisation by time can only be improved by worsening the localisation by frequency and vice versa. The choice of the wavelet transform and of the specific decay parameter σ_0 to be used depends on the data to be analysed and on the information we are interested in. It's the task of the analyst to find the best compromise.

4. Results

4.1 Variations of length of day (lod)

1. The two time series of lod and of AAM plotted in fig. 1a are analysed by the wavelet transform. In fig. 1b the main components of lod can be seen which are the well-known annual variation and the semi-annual variation, the second one being considerably smaller than the first one.

2. These seasonal variations of lod are strongly correlated with variations of the angular momentum due to the global zonal wind and air pressure field (fig. 1c). A comparison of the energies of the wavelet spectra shows that the annual period of lod is almost totally excited by the global wind and pressure field whereas the semiannual period must be additionally excited by sources other than the winds and the pressure.

3. The main other cause for the semi-annual variation in lod is the Antarctic Circumpolar Current (ACC) (e.g. Dickey et al., 1993). It is responsible for the variation of the semi-annual oscillation of lod because it feeds the cold Humboldt current from South to North along the West coast of South America. During strong El Niño events (marked by thin vertical lines in fig. 1b,c) the Humboldt current is being disturbed by the warm water of the El Niño moving from North to South along the South American West coast. There are many interferences in the currents and as a consequence the semi-annual variation in lod almost vanishes. This is clearly visible on fig. 1b, e.g. during the very strong El Niño event in 1982/83.

4. In the period range shorter than 150 days down to 30 days, the transient variations of lod can be clearly assigned to those related to the global wind and pressure field, i.e. they are almost exclusively caused by atmospheric excitation as can be seen on fig. 2a,b,. The wavelet cross-spectrum (fig. 2c) nicely displays the strong correlation between the two processes in that period range. The average coorelations have increased since 1983 as a consequence of the improved accuracy of lod-measurements by using modern space geodetic techniques.

5. The periods below 30 days in lod are mainly due to the solid Earth tides (Yoder et al., 1981). The strongest variations are those with periods of 13.63 and 13.66 days (fortnightly variations) with a modulation period of 18.6 years and those around 27.6 days (monthly variations) what can be seen on fig. 2a and also on fig. 2d, which is a 3d-representation of fig. 2a rotated by 180°.

6. We have also looked on the very short periods in lod, i.e. on periods between 5 and 8 days. These were not filtered out from the EOP(IERS) C 04 series since 1988. For this period range, the wavelet spectrum shows very clearly the tidally induced variations around 7 days for the *whole* data set (fig. 3). Even the two periods at 6.86 days and at 7.10 days can be separated. This proves that it was not the time-interval of 5 days and 7 days, respectively, between subsequent VLBI-sessions which caused the peaks around 5 and 7 days in lod-spectra (plotted on fig. 3, left side) as was assumed by some authors in the past (e.g. Dehant et al., 1999). If such an effect existed, it would have been made visible by the wavelet analysis with a clear 'step' from 5 to 7 days in April, 1991 when the VLBI observing interval was changed from 5 to 7 days. On fig. 3 we can also see a thin horizontal line at the 5.6 days period till 1991 corresponding to the small tidally induced variation with that period. Later than 1991 this period could not be resolved any more because the time interval between the VLBI sessions had been increased to 7 days.

4.2 Variations of polar motion

The variations of the energy of the Chandler motion and that of the annual term can only be seen by the wavelet analysis if one of the two is filtered out first from the original data series. Then, the following results are obtained from the wavelet analysis (see also Schmitz-Hübsch, 1999):

1. The annual wobble of the pole (with the Chandler wobble removed before) is dominantly a prograde circular motion; their retrograde part is very small (fig. 4a). By contrast, the corresponding annual variation of the angular momentum due to the global air pressure and wind gets a strongly elliptic shape, i.e. with prograde *and* retrograde components of about the same magnitude (fig. 4b).

2. The energy of the annual polar motion is not constant by time. In particular there are strong maxima of the annual variations one to two years after the strongest El Niño events (fig. 4a). Such time lags cannot be seen in the corresponding wavelet spectra of the global AAM excitation functions (fig. 4b). This indicates the important role of the global oceans in the excitation of the annual polar wobble as was also reported by Furuya and Hamano (1998) when looking at the Pacific Ocean, only.

3. For the short periods of polar motion, i.e. below 200 days, we see strong semi-annual prograde circular variations of polar motion (fig. 5a). The maxima of the semi-annual prograde variations are delayed by about one year after the strong El Niños in 1982/83, 1987/88 and 1994/95, respectively. There is no such an effect in AAM (fig. 5b) which contains only weak semi-annual variations. This is an indication that the semi-annual polar motion is mainly caused by the oceans rather than by the atmosphere.This shows again the influence of the global ocean currents and in particular of El Niño events on the Earth rotation parameters.

4. The wavelet spectrum of the very short periods of polar motion, i.e. below 60 days is plotted in fig. 6a. It should be mentioned that whereas the seasonal variations of the pole (annual and semi-annual) are mainly prograde motions (see fig. 4a, 5a), the very short periods of the pole, i.e. below 30 days (as

can be seen on fig. 6a), are mainly retrograde motions. A comparison with the corresponding AAM variations (fig. 6b) shows that the relative energy of the variations from 60 down to 30 days is higher in polar motion than in AAM but the situation gets inverse for the extremely short periods below 30 days. It should be noted that periods shorter than 30 days in polar motion are not visible in the spectrum before the end of 1983. The investigation of the very short periods then became possible due to the increased accuracy of the VLBI results which was achieved by adding the Wettzell station to the international VLBI network in autumn 1983.

5. High-frequency variations with periods shorter than 12 days are visible in polar motion (fig. 7a). These very short periods are also contained in AAM (fig. 7b), they are due to the atmospheric modes in the period range of 8-10 days (Eubanks et al., 1988; Eubanks, 1993). On fig. 7a it can be seen even clearer than on fig. 6a that the very short periods in polar motion are dominantly retrograde motions.

5. Conclusions

The wavelet transform was successfully applied for the investigation of variations in length of day (lod) and polar motion. By the Morlet analysing function transient, quasi-periodic effects could be exactly localised and be brought into relation with other geophysical phenomena such as variations of global wind and pressure and the global oceanic tides and currents.

The seasonal variations of lod are dominated by the variations of the global zonal wind field. An interaction between strong El Niño events, the Antarctic Circumpolar Current and the semi-annual period of lod could be shown. Irregular short-period variations in lod could also be localised and could be clearly brought into relation with AAM variations. The tidally induced lod variations with periods between 5 and 35 days could also be made visible. By the wavelet analysis it could be proved that the variations around 7 days are due to the tides, too, and had not been caused by the time interval of 7 days between subsequent VLBI sessions.

The annual wobble of the pole is mainly a prograde circular motion. Epochs when small elliptic 'deformations' of these usually circular polar motions occured can be identified by the wavelet analysis. The seasonal atmospheric excitation due to variations of the global air pressure and wind contains both prograde and retrograde terms corresponding to an elliptic motion.

Influences of strong El Niño events on polar motion could also be made visible for different periods which demonstrates the important effect of oceanic currents on the pole.

The very short-period variations of the pole are mainly retrograde motions. It can nicely be seen in the wavelet spectra that the increased accuracy of the observed polar motion has allowed since the end of 1983 to detect the periods shorter than 30 days, too.

6. References

Barnes, R.T.H, R. Hide, A.A. White, and C.A. Wilson: Atmospheric angular momentum fluctuations, length-of-day changes and polar motion, Proc. R. London, Vol. A387, 31-73, 1983

Baudin, F., A. Gabriel, D. Gilbert, P.L. Pallé, and C. Régulo: Temporal characteristics of solar P-modes, Astron. Astrophys., Vol. 311, 1024-1032, 1996

Brzeziński, A.: Statistical investigations on atmospheric angular momentum functions and their effects on polar motion, Manuscripta Geodaetica, Vol. 12, 268-281, 1987

Chao, B.F., and I. Naito: Wavelet analysis provides a new tool for studying Earth's rotation, EOS, Trans. AGU, Vol. 76, 161-165, 1995

Chao, B.F., R.D. Ray, and G.D. Egbert: Diurnal/semidiurnal oceanic tidal angular momentum: Topex/Poseidon models in comparison with Earth's rotation rate, Geophys. Res. Lett., Vol. 22, 1993-1996, 1995

Chao, B.F., R.D. Ray, J.M. Gipson, G.D. Egbert, and C. Ma: Diurnal/semidiurnal polar motion excited by oceanic tidal angular momentum, J. Geophys. Res., Vol. 101, 20151-20163, 1996

Chao, B.F., and Y.-H. Zhou: Meteorological excitation of interannual polar motion by the North Atlantic Oscillation, Journ. of Geodynamics, Vol. 27, 61-73, 1999

Dehant, V. et al.: Considerations concerning the non-rigid Earth nutation theory, accept. by Celest. Mech., 1999

Dickey, J.O. et al.: The oceanic contribution to the Earth's seasonal angular momentum budget, Geophys. Res. Lett., Vol. 20, No. 24, 2953-2956, 1993

Eubanks, T.M., J.A. Steppe, J.O. Dickey, R.D. Rosen, and D.A. Salstein: Causes of rapid motions of the Earth's pole, Nature, Vol. 334, 115-119, 1988

Eubanks, T.M.: Variations in the orientation of the Earth, In: Contributions of Space Geodesy to Geodynamics: Earth Dynamics, Geodynamics Series, AGU, Vol. 24, 1-54, 1993

Furuya, M, and Y. Hamano: Effect of the Pacific Ocean on the Earth's seasonal wobble inferred from National Center for Environmental Prediction ocean analysis data, Journ. of Geophys. Res., Vol. 103, No. B5, 10131-10140, 1998

Gambis, D.: Wavelet transform analysis of the length of the day and the El Niño/Southern Oscillation variations at intraseasonal and interannual time scales, Ann. Geophysicae, Vol. 10, 429-437, 1992

Gibert, D., M. Holschneider, and J.L. Le Mouël: Wavelet Analysis of the Chandler Wobble, Journ. Geophys. Res., Vol. 103, No. B11, 27069-27089, 1998

Grossmann, A., and J. Morlet: Decomposition of Hardy functions into square integrable wavelets of constant shape, SIAM J. Math Anal., Vol. 15(4), 723-736, 1984

Kolaczek, B.: Short Period Variations of Earth Rotation, Proc. Journées 1995 Systèmes de Référence Spatio-Temporels, Warsaw, Poland, 18-20, publ. by the Space Res. Centre, PAS, Poland, 147-154, 1995

Kosek, W., J. Nastula, and B. Kolaczek: Variability of polar motion oscillations with periods from 20 to 150 days in 1979-1991, Bull. Géodés., Vol. 69, 308-319, 1995

Kosek, W.: Time Variable Band Pass Filter Spectra of Real and Complex-Valued Polar Motion Series, Artificial Satellites, Planetary Geodesy No 24, Vol. 30, No 1, publ. by the Space Research Centre, Warsaw, 27-43, 1995

Kosek, W., and L. Kolaczek: Semi-Chandler and semiannual oscillation of polar motion, Geophys. Res. Letters, Vol. 24, No. 17, 2235-2238, Sept. 1997

Kuehne, J., S. Johnson, and C.R. Wilson: Atmospheric excitation of nonseasonal polar motion, J. Geophys. Res., Vol. 98, 19973-19978, 1993

Nastula, J.: Short periodic variations in the Earth´s rotation in the period 1984-1990, Ann. Geophysicae, 10, 441-448, 1992

Nastula, J.: The regional atmospheric contribution to the polar motion and EAAM excitation functions, Proceedings of the Int. Symp. on Gravity, Geoid and Marine Geodesy, 1996, Tokyo, 1997

Nastula, J., W. Kosek, and B. Kolaczek: Analysis of zonal atmospheric excitation functions and their correlation with polar motion excitation functions, Space Res. Centre, PAS, Warsaw, Poland, 1997

Popinski, W., and W. Kosek: Wavelet Transform and its Application for Short Period Earth Rotation Analysis, Planetary Geodesy, 22, Vol. 29, 2, 75-86, 1994

Praveen, K.: Wavelet Analysis for Geophysical Applications, Reviews of Geophysics, Vol. 35/4, 385-412, 1997

Rosen, R.D.: The Axial Momentum Balance of Earth and its Fluid Envelope, Surveys in Geophysics, Vol. 14, 1-29, 1993

Schmidt, M.: Moderne Methoden der Signalanalyse, Zeitschrift für Vermessungswesen, ZfV, Vol. 121, 315-325, 1996

Schmidt, M., and H. Schuh: Wavelet-Analyse der mit VLBI beobachteten Nutationsreihen, Zeitschrift für Vermessungswesen, ZfV, Vol. 124 , 25-30, 1999

Schmidt, M.: Grundprinzipien der Wavelet-Analyse und Anwendungen in der Geodäsie, habilitation thesis (preprint), München, 1999

Schmitz-Hübsch, H.: Wavelet Analysis of Earth Rotation, Proc. of the III. Venezuelian Congress of Geodesy, 23.-27.02.1998, Maracaibo, in press, 1999

Schuh, H., and O. Titov: Short-period variations of the Earth rotation parameters as seen by VLBI, Proc.

of the 13th Working Meeting on European VLBI for Geodesy and Astrometry, Viechtach, Feb. 1999, ed. by W. Schlüter and H. Hase, 172-177, 1999

Torrence, C., and G.P. Compo: A practical guide to wavelet analysis, Bull. of the American Meteorological Society, Vol. 79(1), 61-78, 1998

Whitcher, B.: Assessing Nonstationary Time Series Using Wavelets, Ph. D. thesis, Univ. of Washington, 1998

Yoder, C.E., J.G. Williams, and M.E. Parke: Tidal variations of Earth rotation, J. Geophys. Res., Vol. 86, 881-891, 1981

Zatman, S., and J. Bloxham: The phase difference between length of day and atmospheric angular momentum at subannual frequencies and the possible role of core-mantle coupling, Geophys. Res. Letters, Vol. 24, No. 14, 1799-1802, 1997

Figure Captions

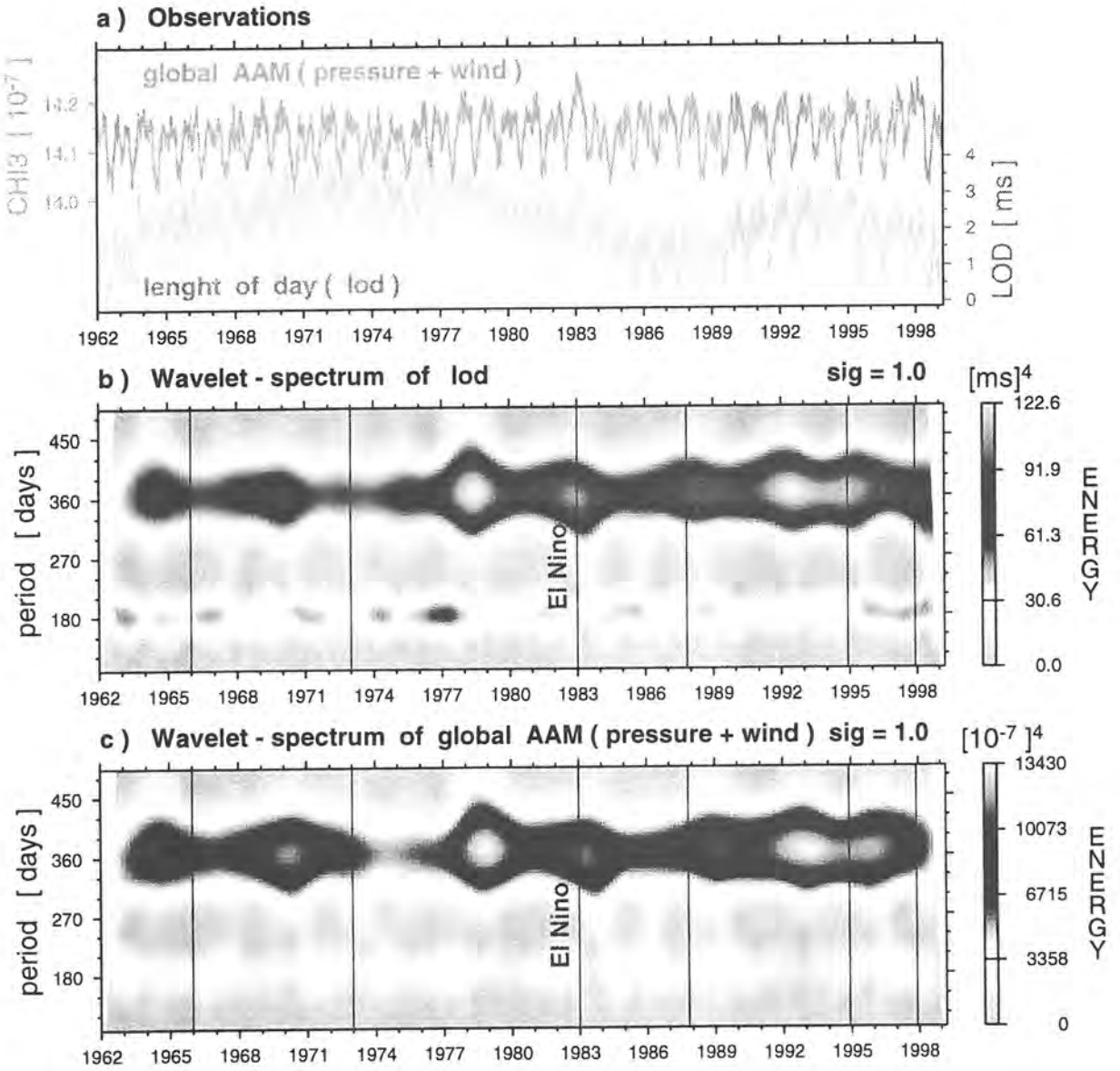


Fig. 1: Observed length of day (lod) from 1962 till 1999 and atmospheric angular momentum (AAM) excitation functions due to global pressure and wind (a) and wavelet spectra of the two series showing the seasonal variations (b,c)

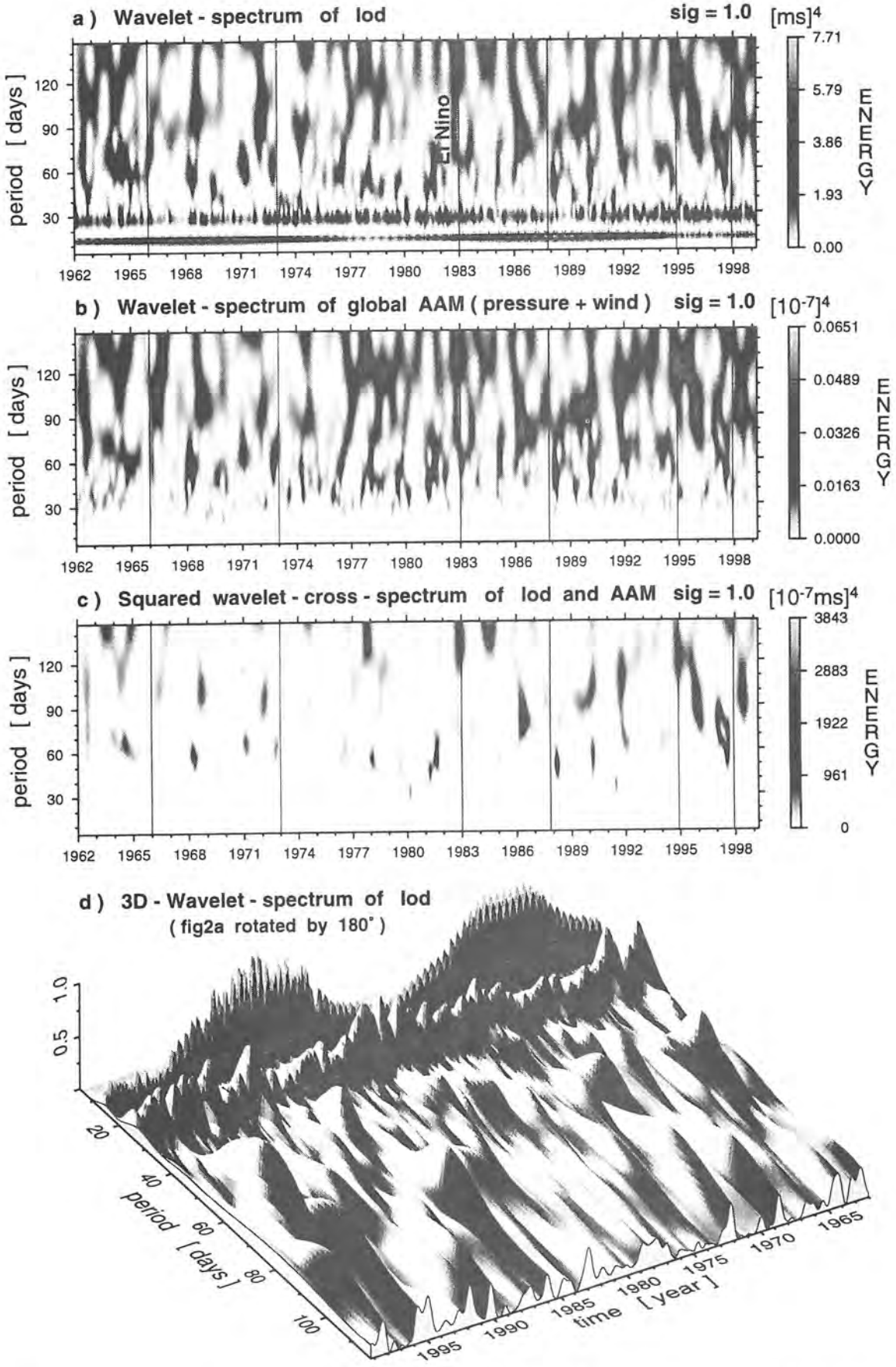


Fig. 2: Wavelet spectra of lod and of AAM for periods from 10 to 150 days (a,b), the wavelet cross spectrum (c) and a 3D-representation of fig. 2(a) rotated by 180° (d)

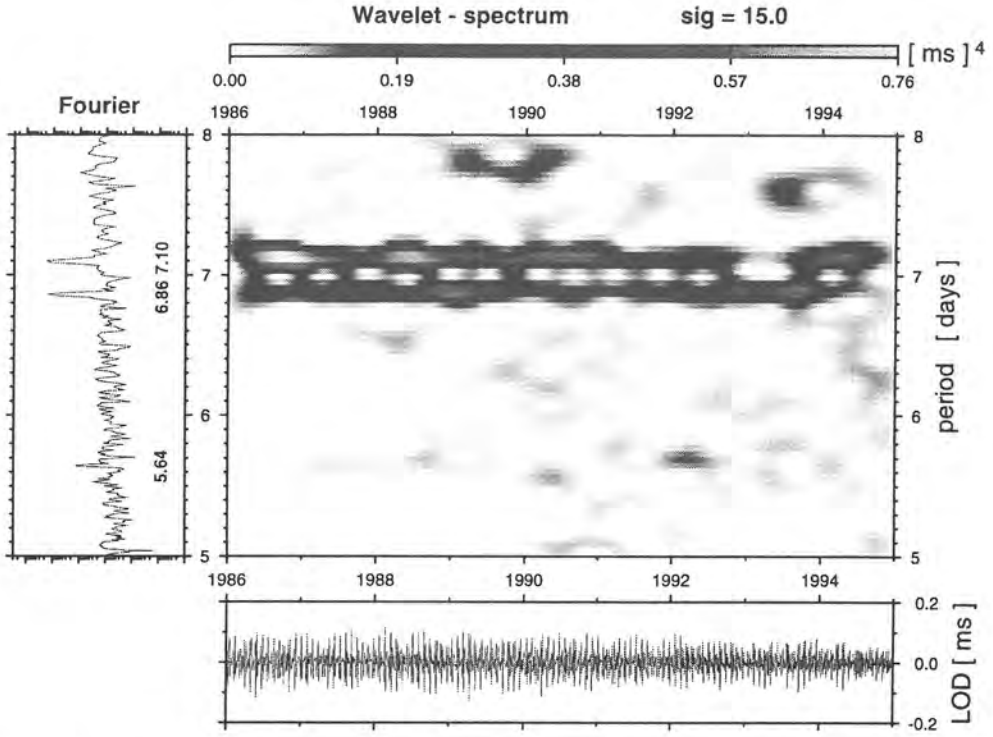


Fig. 3: Wavelet spectrum of lod for the very short periods between 5 and 8 days. The tidally induced variations around 7 days can be seen clearly

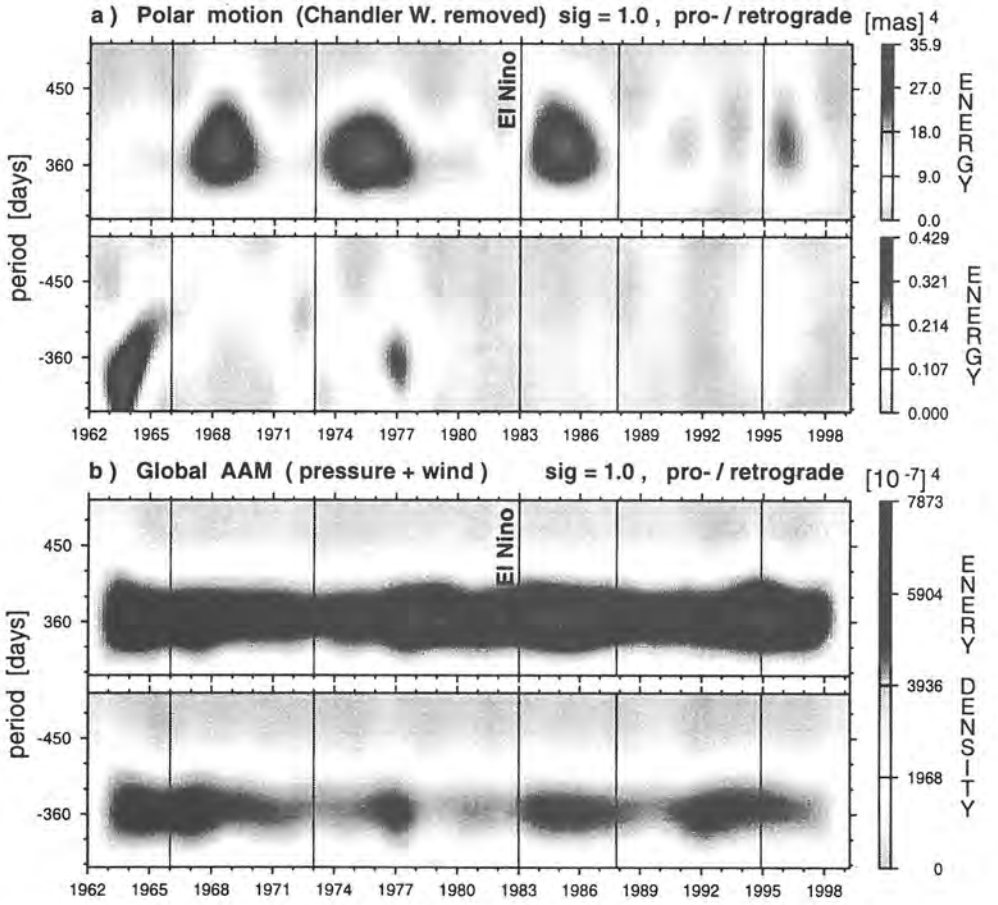


Fig. 4: Wavelet spectra of polar motion (a) and of the global AAM excitation functions (b) around the annual period (the Chandler Wobble at 1.18 years had been removed before)

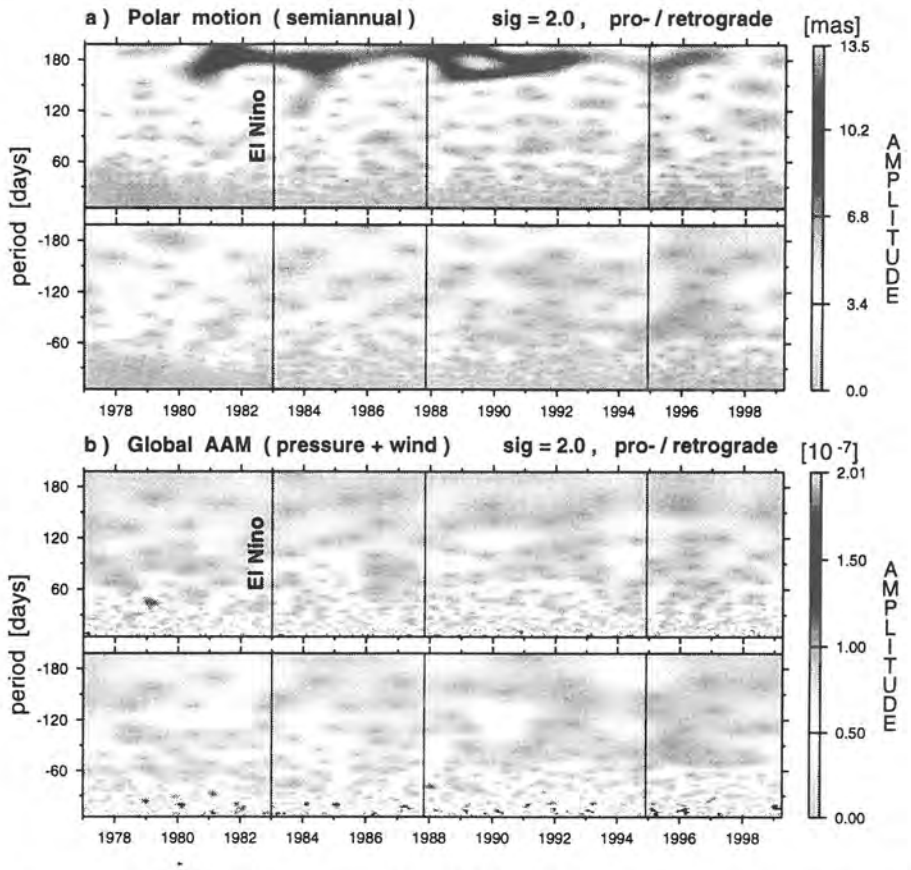


Fig. 5: Wavelet spectra of polar motion (a) and of AAM (b) for periods shorter than 200 days

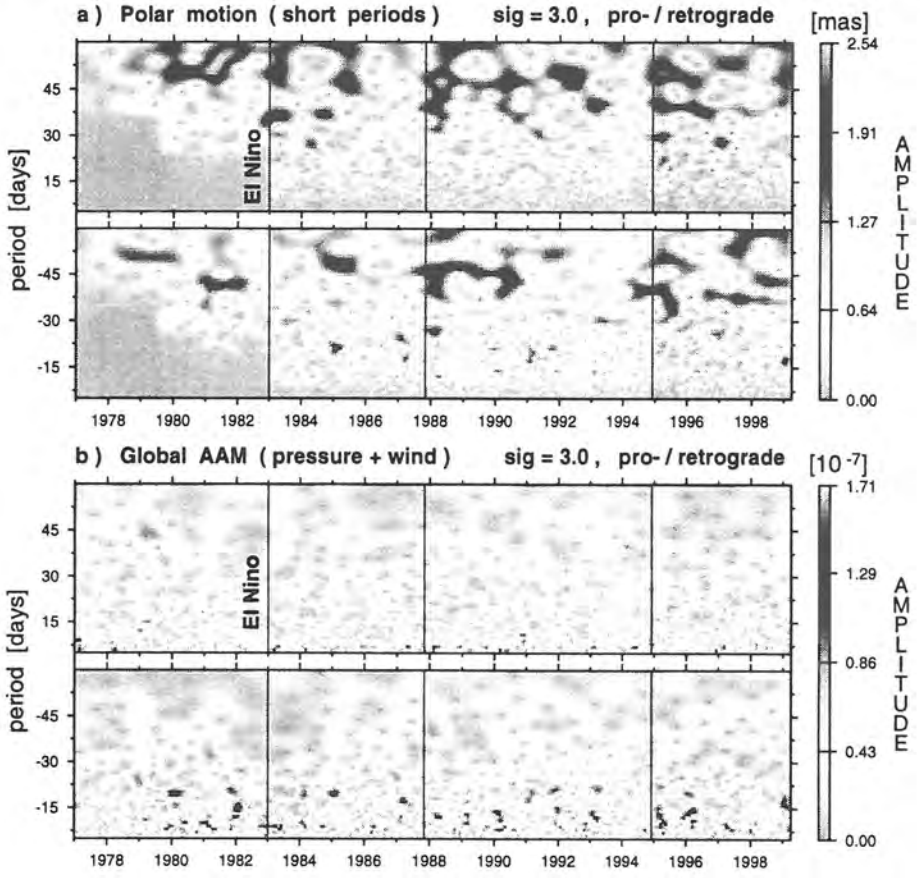


Fig. 6: Wavelet spectra of polar motion (a) and of AAM (b) for periods shorter than 60 days

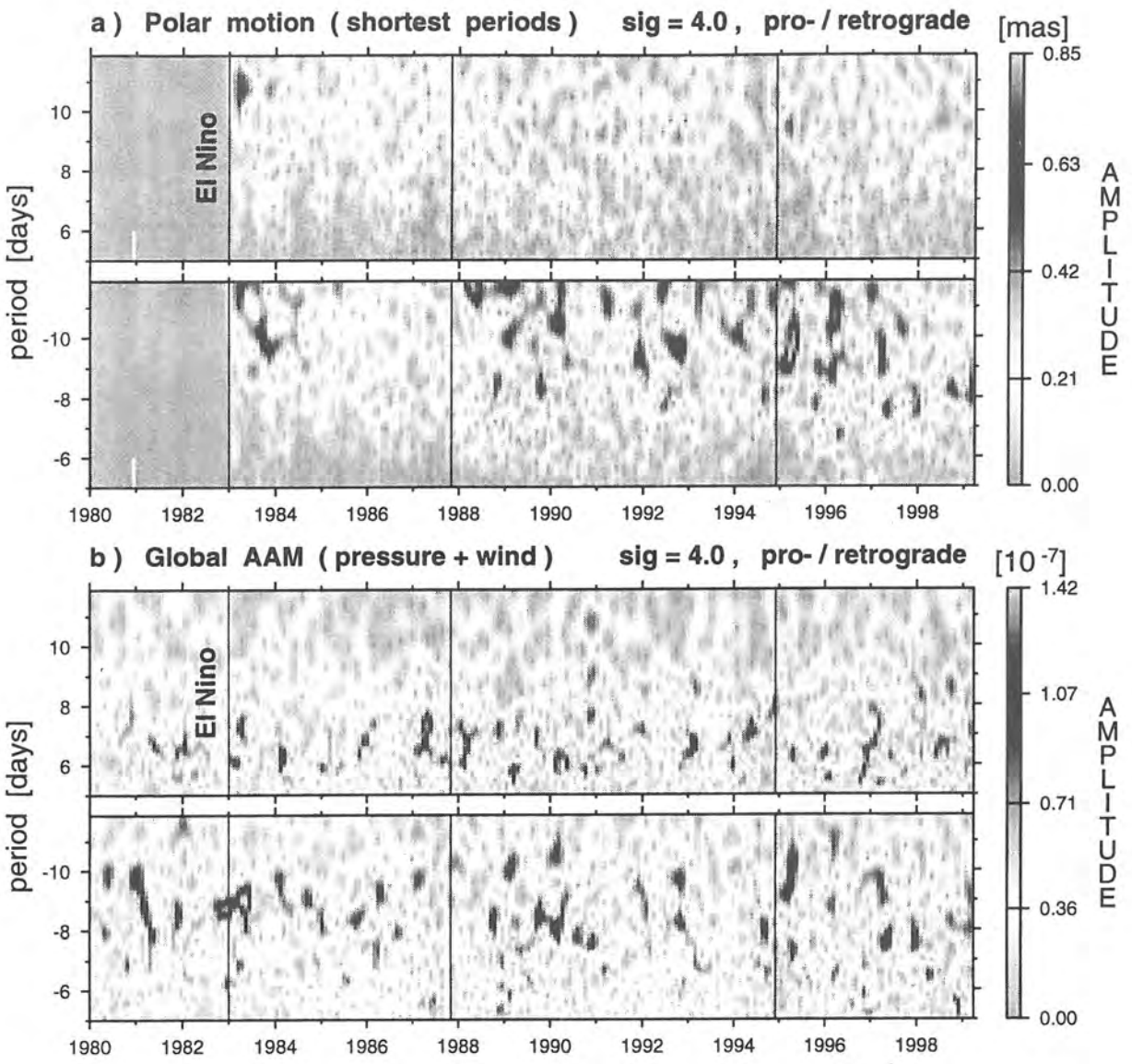


Fig. 7: Wavelet spectra of polar motion (a) and of AAM (b) for periods between 5 and 12 days

Charakterisierung des C-II Ringlasers

Ulrich Schreiber, Manfred Schneider, Geoffrey E. Stedman, Clive H. Rowe, Wolfgang Schlüter

Abstract:

Intertialsensoren spielen für die hochgenaue Vermessung der Erdrotation in der Geodäsie noch keine Rolle, da ihre technische Realisierung noch nicht weit genug vorangetrieben worden ist. Mechanische Kreisel, Faserkreisel und Laserkreisel finden zwar eine breite Anwendung für Aufgaben in der Navigation und Orientierung von beweglichen mechanischen Strukturen. Allerdings sind die Genauigkeitsanforderungen an Auflösungsvermögen und Stabilität für diese Anwendungen sehr viel geringer als für geowissenschaftliche Fragestellungen. Mit dem Canterbury Ringlaser C-I konnte gezeigt werden, daß durchaus die Möglichkeit besteht einen großen Ringlaser zu bauen. Von einem Gerät (G) mit einer umspannten quadratischen Fläche von 16 m^2 kann man erwarten in den geowissenschaftlich interessanten Genauigkeitsbereich von 10^{-9} oder besser der Erdrotation vorzudringen [3, 4]. Ausgehend von den Erfahrungen mit dem C-I wurde der Prototyp C-II entworfen und gebaut, der alle notwendigen Eigenschaften des Großringes G erproben soll. Sein Konzept und die erzielten Ergebnisse werden hier vorgestellt.

1 Das Ringlaser Prinzip

Umläuft Licht einer gegebenen Wellenlänge eine beliebig geformte Fläche gegensinnig, so ergibt sich eine Phasenverschiebung, wenn diese Fläche gegen das lokale Fermisystem [6] rotiert. Diese Phasenverschiebung zwischen den beiden Umlaufsinnen ist proportional zur Rotationsgeschwindigkeit des Flächenstückes und verschwindet, wenn die Drehrate zu Null wird. Diese Eigenschaft wurde von G. Sagnac 1913 experimentell beobachtet und heißt Sagnac-Effekt. Führt man diese Fläche beispielsweise als Quadrat (A) aus und fügt ein Verstärkungsmedium (z.B. ein Gasgemisch aus Helium und Neon) hinzu, so ergibt sich ein Laseroszillator mit ringförmigem Resonator und die meßtechnisch schlecht bestimmbare Phasendifferenz des passiven Sagnacinterferometers wird zu einer Frequenzdifferenz [1]. Somit ergibt sich die Ringlasergleichung zu

$$\delta f = \frac{4\mathbf{A} \cdot \mathbf{\Omega_e}}{\lambda L} \tag{1}$$

wobei L der Umfang des Ringes ist und $\lambda = 633$ nm die Wellenlänge des HeNe Lasers, während $\mathbf{\Omega}_e$ die Rotationsgeschwindigkeit der Erde beschreibt.
Für den Betrieb eines Gyroskopes ist dies jedoch nicht die einzige Bedingung. Man muß ferner sicherstellen, daß nur eine einzige longitudinale Lasermode, bevorzugt TEM(0,0), pro Umlaufsinn anschwingt. Für Navigationskreisel mit deutlich geringerer Winkelauflösung ist dies kein Problem, da deren Resonatorlängen mit L < 60 cm nur eine Longitudinalmode verstärken kann, bei einem großen Ring wie dem C-II mit einem freien Spektralbereich (FSR = c/L = 74.948 MHz; für L = 4 m) sind jedoch mehrere longitudinale Moden anregbar. Durch gezieltes Herabsetzen der Verstärkung des Lasermediums bleibt letzten Endes nur noch eine Lasermode übrig, die sich aufgrund eines differentiell geringfügig höheren Verstärkungsfaktors auf Kosten der benachbarten Moden behauptet.

2 Eigenschaften des C-II

Um größtmögliche Stabilität zu gewährleisten ist der Resonator des Lasergyroskopes in monolithischer Bauweise aus dem formstabilen Werkstoff Zerodur ausgeführt worden. Er besteht aus einem quadratischen Block von 1,2 m Kantenlänge und einer Dicke von 18 cm. Alle 4 Ecken sind abgeschrägt, so daß Bohrungen mit einem Durchmesser von 3 cm parallel zu den Kanten, 10 cm vom Rand entfernt in der neutralen Ebene des Blockes ein Quadrat von 1 m^2 bilden. Dort wo die Bohrungen an den abgeschrägten Enden aufeinandertreffen sind Umlenkspiegel mit Hilfe von Glasträgern optisch angesprengt worden, so daß das Rohrleitungssystem evakuierbar ist. Bohrungen entlang der Diagonalen des quadratischen Rezipienten mit einer Öffnung an die Blockoberseite stellen den Anschlußpunkt für eine Vakuumapparatur her. An zwei gegenüberliegenden Seiten wurde noch eine buchtenförmige Aussparung von 30 cm Länge und 20 cm Tiefe angelegt. Eine dieser Buchten wird zur Erzeugung der Laserstrahlen verwendet, indem hier eine Kapillare von 4 mm Querschnitt eingesetzt wurde. Über 2 Elektroden wird die Leistung eines Hochfrequenzsenders eingekoppelt und ein Plasma angeregt. Die andere Bucht ist durch ein Glasrohr als Verbindungsstück wieder luftdicht verschlossen worden. Sie ist für experimentelle Zwecke vorgesehen, bei denen ein allseitiger Zugang zum Laserstrahl erforderlich ist. Abbildung 1 zeigt den C-II nach der Fertigstellung des Gerätes durch die Fa. Carl Zeiss an seinem Aufstellungsort in der Cashmere- Höhle in Christchurch (Neuseeland).



Abbildung 1: Der C-II Ringlaser an seinem Aufstellungsort in Christchurch

Die Rotationsbewegung eines Ringlasers kann man erfassen, indem die trotz der geringen Transmission von weniger als 1 ppm durch die Spiegel durchtretende Laserstrahlung in einem Kösterprisma überlagert und die Schwebungsfrequenz mit einem Photodetektor (Photomultiplier, PIN-Diode) registriert wird. Die am Ausgang des Detektors als Wechselspannung erscheinende Sagnacfrequenz des Ringlasers wird anschließend spannungsverstärkt, digitalisiert mit Zusatzinformationen der meteorologischen Station ergänzt und gespeichert. Da der meßbare Anteil der Sagnacfrequenz von der Orientierung des Ringlasers relativ zur Rotationsachse der Erde abhängt, läßt sich Gleichung (1) auch schreiben als

$$\delta f = \frac{4\mathbf{A}\cdot\mathbf{\Omega_e}}{\lambda L} = \frac{f_0 \Omega cos\Lambda}{4F}. \qquad (2)$$

Hierbei ist die Erdrotationsrate $\Omega = 2\pi/86164.1 = 7.2921 \times 10^{-5}$ rad/s, wobei die Zahl im Nenner der siderischen Tageslänge in Sekunden entspricht, $f_0 = 4.73612215 \times 10^{14}$ Hz der optischen Frequenz des Lasers (genauer die Komponente i des I_2), $F = 74.948$ MHz dem freien Spektralbereich und $\Lambda = 43.5679^o = 43°34'37''$ S dem Breitengrad des Aufstellungsortes bei Christchurch. Wenn eine Zeitreihe der Sagnacfrequenz erfaßt ist, kann sie mit den Methoden der digitalen Spektralanalyse weiter untersucht werden. Durch Fouriertransformation erhält man das dazugehörige Frequenzspektrum. Abbildung 2 zeigt hierzu ein Beispiel. Für den Betrieb des Ringlasers als Erdrotationssensor ist es wichtig, daß die zirkulierende Laserleistung möglichst hoch und damit die gemessene Linienbreite so klein wie möglich ist. Dies wird zum einen durch die Länge der aufgenommenen Zeitreihe und zum anderen durch die mögliche Verstärkung im

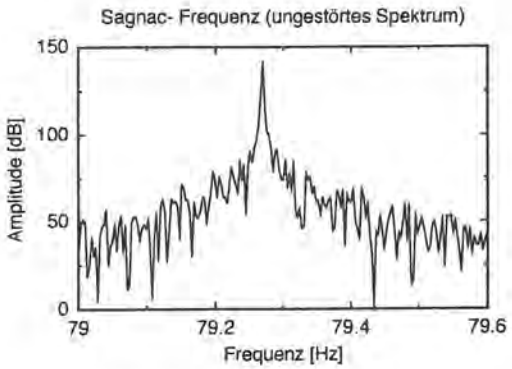


Abbildung 2: Beispiel für ein Sagnacspektrum des C-II Ringlasers. Die Linienbreite beträgt weniger als 200 μHz FWHM und der Signal-/Rauschabstand ist größer als 90 dB.

Lasermedium bis zum Auftreten weiterer Longitudinalmoden begrenzt. Für den C-II erhält man die minimal erreichbare Linienbreite mit 200 μHz. Sie ergibt sich sowohl aus der Abschätzung der Leistungsbilanz, wie auch aus Messungen. Da auch bei einer gegebenen Halbwertsbreite die Linienmitte noch mit höherer Genauigkeit bestimmt werden kann, ist das eigentliche Kriterium für die Qualität der Apparatur die Stabilität der Linienmitte. Hier ist es erforderlich, daß apparativ verursachte Stabilitätsschwankungen geringer sind als die Variation in der Eingangsdrehrate.

3 Fehlerquellen im Aufbau von Ringlasern

Obwohl es immer wieder möglich ist, die Drift der Apparatur über einige Stunden hinweg so gering zu halten, daß sie deutlich unterhalb der schrotrauschbegrenzten Linienbreite der Sagnaclinie liegt, ist dies über größere Zeiträume hinweg bislang noch nicht gelungen. Ursache hierfür ist eine Schwankung im Betrag des Skalenfaktors in der Ringlasergleichung. Schreibt man Gleichung (1) in der Form

$$\delta f = \frac{4A}{\lambda L}\mathbf{n} \cdot \mathbf{\Omega_e} \tag{3}$$

mit $\mathbf{n}$ dem normierten Flächennormalenvektor, so ist der Skalenfaktor definiert als

$$Sk = \frac{4A}{\lambda L} \tag{4}$$

und abhängig von der Größe des Instrumentes. Auch wenn durch die Verwendung des selbst unter dem Einfluß moderater Temperaturschwankungen formstabilen Werkstoffes Zerodur geometrische Einflüsse auf die Größe der Fläche (A) und Länge des Umfanges (L) ausgeschlossen sind, so ergibt sich durch die Rückstreuung von Licht aus einem Umlaufsinn in den jeweils gegenläufigen Sinn ein Mechanismus für die Störung der Sagnacfrequenz [5]. Die Größe der Rückstreuung an einem Spiegel gibt eine sogenannte Einrastfrequenz ω_L vor, bei der kein Sagnacsignal mehr beobachtet werden kann, wenn die Eingangsdrehrate für den Kreisel so gering wird, daß die daraus resultierende Sagnacfrequenz die Einrastfrequenz unterschreitet. Im einfachsten Fall [2] gilt:

$$\delta f = SK\sqrt{\omega^2 - \omega_L^2} \tag{5}$$

Da ω_L nicht mit der Zeit konstant ist, sondern sich mit der Phasenlage des effektiv rückreflektierten Signales ändert, ist es wichtig, daß die geometrische Beziehung aller Rückstreuquellen im Strahlengang konstant bleibt. Experimentell wurde beobachtet, wie bereits Schwankungen

in den Spiegelabständen von unter 5 nm bedeutende Auswirkungen zeigen. Das entspricht einer Resonatorlängenänderung von weniger als $\lambda/100$. Solche Änderungen ergeben sich im gegenwärtigen Aufbau aus dem Einfluß von Druck- und Temperaturschwankungen auf den Zerodurblock, wie auch durch eine Membranwirkung der Spiegelträger.

4 Der Betrieb des C-II

Für den Betrieb des Ringlasers hat sich die Amplitudenstabilisation als wichtig erwiesen. Sie erfüllt im wesentlichen zwei Aufgaben. Zum Einen verhindert sie, daß sich der Arbeitspunkt des Ringlasers so weit verlagert, daß mehr als eine longitudinale Mode pro Umlaufsinn verstärkt wird. Zum Anderen unterdrückt sie Schwankungen in dem Verstärkungsfaktor des Laserprozesses, welche sich auf den Betrag der gemessenen Sagnacfrequenz auswirken. Realisiert wird dieser Regelkreis dadurch, daß ein Photomultiplier die durch den am weitesten vom Plasma entfernten Umlenkspiegel des im Uhrzeigersinn umlaufenden Laserstrahles durchtretende Lichtintensität aufnimmt, über eine invertierende Operationsverstärkerschaltung verstärkt und damit die Verstärkung des Hochfrequenzsenders regelt, der für die Plasmaanregung sorgt. Da die Kennlinie für die erzeugte Intensität des Laserlichtes in Abhängigkeit von der eingekoppelten Hochfrequenzleistung in der Nähe der Laserschwelle sehr steil ist, erhält man hiermit eine recht konstante Intensität des Ringlasers für beide Umlaufsinne.

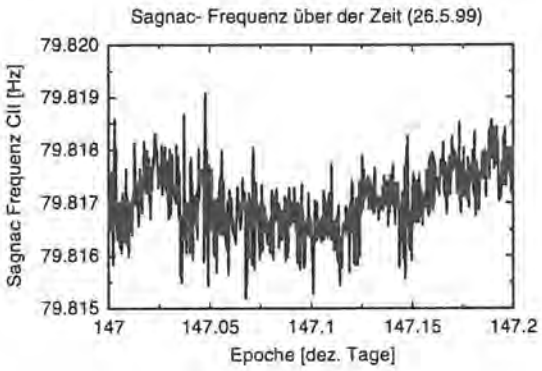


Abbildung 3: Darstellung der Sagnacfrequenz über der Zeit.

Dennoch muß an dieser Stelle ein großer Aufwand getrieben werden, da selbst minimale nderungen in der Lichtintensität von weniger 0,1% sich bereits auf die Sagnacfrequenz auswirken. Abbildung 3 zeigt einen Ausschnitt aus einer Zeitreihe gemessener Sagnacfrequenzen des C-II. Dabei zeigt die Streuung der Messwerte die durch das Schrotrauschen gegebene Unsicherheit in der Bestimmung der Sagnacfrequenz an, während die systematischen Variationen in der Zeitreihe dem gegenwärtigen Verständnis nach auf eine unzureichende Stabilisierung der eingekoppelten Hochfrequenzleistung zurückzuführen ist. Nach den Umgebungseinflüssen wie Luftdruck und Temperatur ist dies wohl die nächstgrößte Instabilitätsquelle in dem Ringlasersystem. Mit dem Regelkreis zur Stabilisierung der Hochfrequenzanregung des Laservorganges und der Digitalisierung des Sagnacsignales ist die Ringlasersteuerung noch nicht komplett. Zusätzlich werden auch noch die Umgebungsparameter Luftdruck, Luftfeuchte und Temperatur des Ringes mitbestimmt. Der Luftfeuchte kommt hierbei eine untergeordnete Bedeutung als Überwachungsgröße für die Benutzbarkeit des Labores zu, während Luftdruck und Ringtemperatur Einfluß auf die Sagnacfrequenz nehmen.

5 Meßergebnisse

Das Datenerfassungsprogramm des C-II ist so aufgebaut worden, daß es jeweils nach 30 Minuten eine Messung aufzeichnet und die momentane Sagnacfrequenz über diese Zeitspanne bestimmt. Damit erhält man über längere Zeiträume hinweg Darstellungen über die Stabilität der Apparatur. Abbildung 4 zeigt ein solches Beispiel. Deutlich sieht man die durch die Skalenfaktorschwan-

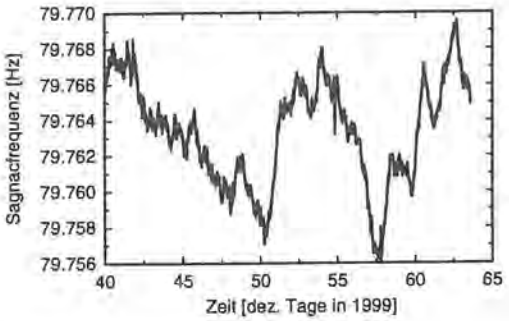


Abbildung 4: Darstellung einer Zeitreihe der Sagnacfrequenz wie sie vom C-II gemessen wurde..

kungen verursachten Systematiken auf der Zeitreihe des Sagnacsignales. Diese korrelieren mit den Schwankungen im Luftdruck, wie man der Abbildung 5 entnehmen kann. Auch Schwankun-

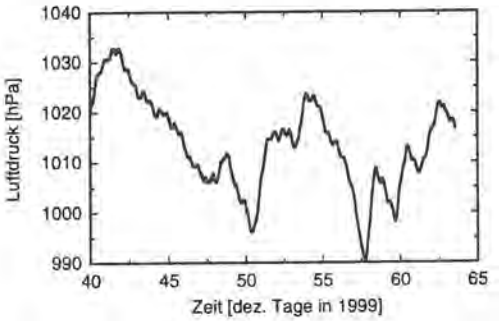


Abbildung 5: Der Verlauf des Luftdruckes während der Messungen.

gen in der Temperatur nehmen einen großen Einfluß, so daß sie für eine Korrektur berücksichtigt werden müssen. Den Temperaturverlauf zeigt die Abbildung 6.
Es wurde nun versucht durch Berücksichtigung des Luftdruck- und Temperaturverlaufes die Skalenfaktorvariationen zu modellieren. Auf diese Weise ist der korrigierte Verlauf der Sagnacfrequenz über der Zeit in Abbildung 7 entstanden. Man sieht eine deutliche Verbesserung durch

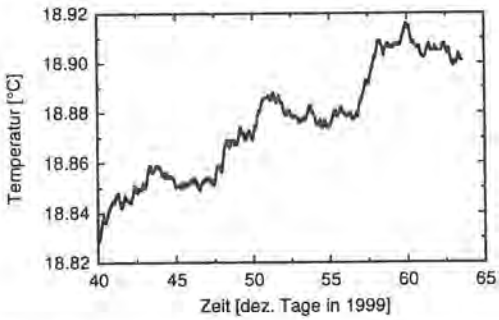


Abbildung 6: Der Verlauf der Temperatur während der Messungen.

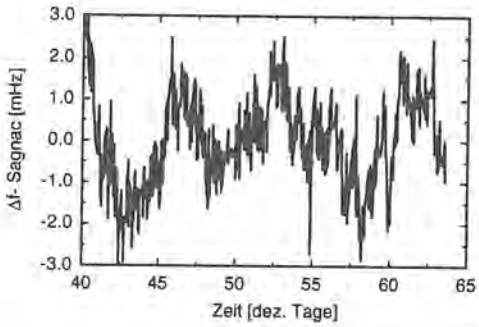


Abbildung 7: Der nach dem Einfluß von Luftdruck und Temperatur korrigierte Verlauf der Sagnacfrequenz über der Zeit.

ein Korrekturmodell [7] mit linearem Ansatz für die Drift durch Luftdruck und Temperatur. Nach den ausgeführten Korrekturen ergibt sich eine Stabilität des C-II Ringlasers für die Messung der Erdrotationsrate von bis auf 0,05 - 0,3 Winkelsekunden pro Stunde über einen Messzeitraum von 24 Tagen hinweg. Die verbliebene periodische Systematik auf der Zeitreihe der Sagnacfrequenz kann gegenwärtig noch nicht erklärt werden, wird aber auf Veränderungen in der Plasmaanregung zurückführbar sein. Das Hauptaugenmerk richtet sich zur Zeit darauf, wie die Skalenfaktorschwankungen für die Rohmessungen durch apparative Maßnahmen reduziert werden können.

6 Danksagung

G E Stedman und C H Rowe danken für die Unterstützung durch den Marsden Fund UOC 513 und den Projektmitteln der University of Canterbury 2096611, 2098611 welche für die Installation und Inbetriebnahme des C-II gewährt wurden. Die übrigen Autoren danken der Deutschen Forschungsgemeinschaft für die Förderung des Vorhabens (Kennzeichen: SCHN240/6-3).

Literatur

[1] G E Stedman, *Ring laser tests of fundamental physics and geophysics* , Rep. Progr. Phys. **60** 615-688 (1997).

[2] F. Aronowitz; *The Laser Gyro*, Laser Applications, M. Ross (ed), Academic Press, New York (1971), **1**, 133-200

[3] H.R. Bilger, G.E. Stedman, Z. Li, U. Schreiber, M. Schneider; *Ring Lasers for Geodesy*; IEEE Transactions on Instrumentation and Measurement, Vol. 44, 468-470 (April 1995).

[4] H.R Bilger, U. Schreiber, G.E. Stedman; *Design and Application of Large Perimeter Ring Lasers* ; Proceedings of the Symposium Gyro Technology, Stuttgart, Sept. 1996 8.0 - 8.24 (1996).

[5] R. Rodloff; *A Laser Gyro with Optimized Resonator Geometry*, IEEE Journal of Quantum Electronics, **QE-23**, 438-445 (April 1987)

[6] M. Schneider; *Himmelsmechanik, Band III Gravitationstheorie*, Spektrum Akademischer Verlag, Heidelberg, 1996

[7] U. Schreiber; *Ringlaser Technologie für geowissenschaftliche Anwendungen*, Habilitationsschrift, Techn. Universität München, 1998 (im Druck)

Ellipsoidal and topographical effects in the scalar free geodetic boundary value problem

Kurt Seitz

Abstract

In the formulation of the *scalar free boundary value problem* we solve for the gravity potential in the external space outside the earth's surface and for the vertical position of the boundary surface. After linearization the reduced boundary condition refers to the *Telluroid* $s \ni p$, and the new difference quantity δw, the disturbing potential, is introduced. To represent the unknown disturbing potential in the global basis of spherical harmonics a *harmonic analysis* has to be applied to the given boundary data. In this context the boundary data have to be *(downward) continued* from s to a reference surface which shows a rotational symmetry with respect to the earth's mean rotational axis. In general a sphere K or the surface E of an ellipsoid of revolution is selected.

After the analytical continuation of the evaluation operator E_s the boundary condition can be split in two parts. The main component is covered by the *isotropic term* which corresponds to the Stokes–problem. The residual part consists of the *ellipsoidal* and *topographical* components which are functionals of δw. Therefore an iterative solution strategy is appropriate. First numerical evaluations indicate that this iterative process converges for boundary data continued to an ellipsoid, but diverges if the boundary data is continued to a sphere.

Introduction

Geodetic boundary value problems represent idealized situations in geodetic data analysis. Neither the geodetic observations are continuous, nor they are given on the whole surface of the earth. Nevertheless the formulation in the framework of boundary value problems has two important aspects: First, in studying idealized problems in an "ideal" form, mathematical tools can be used which never can applied to real situations, providing deep results concerning the mathematical analysis, which can be generalized to more difficult situations. Second, for special data distributions analytical solutions can be derived which directly can be applied in data evaluation.

In recent years the challenge in the field of the geodetic boundary value problem has been directed to formulations approaching more and more the real data situation in Geodesy. The surface of the earth is now considered as a non–spherical surface, influenced by topography and ellipticity. Mixed boundary value problems have been investigated as well as problems induced by satellite and airborne data which will be available in near future. In this review paper we will restrict to "classical" boundary value problems on the background of geodetic data given on the earth's surface; also over–determined problems will not be considered here.

To determine both the external gravity potential W and the geometry of the earth's surface S, various *boundary value problems* (bvp) can be formulated. They depend on the utilized *observables* L and whether the boundary is supposed to be known or unknown. If the geometry of S is already determined by the classical terrestrial techniques or by methods of satellite geodesy,

then the *fixed boundary value problem* (Klees, 1992, 1995) is under consideration. Otherwise the resulting bvp is of free type. In Grafarend et al. (1985) the *vectorial free bvp* is discussed which differs in many respects from the scalar free bvp, first of all tackled by Sacerdote and Sansò (1986). The relation between the unknowns W, S and the observables L is given by *boundary conditions*. Generally, they are of non–linear structure. In the following we want to focus on the *scalar free bvp*.

The scalar free boundary value problem

In the formulation of the *scalar free bvp* the horizontal position of each point P on the boundary $S \ni P$ is assumed to be given. So we have to introduce beside the external gravity potential W the vertical position P as a geometrical unknown. To solve for both unknowns two types of observables have to be given on the boundary surface in continuous form. We can assume, that the gravity potential $W(P)$ as well as gravity values $\Gamma(P) = |grad\, W(P)|$ on $S \ni P$ are given as boundary data. Assuming that the earth is rigid, rotating with the constant angular velocity ω and giving rise to the non–harmonic centrifugal potential Z, we can formulate the **non–linear scalar free bvp**: *Suppose the boundary data $W(P)$ and $\Gamma(P)$ to be given on $S \ni P$. The unknown gravity potential $W(\mathbf{x})$ has to fulfill the extended Laplace equation in the mass free space Ω_a outside S, and the gravitational potential V tends to zero if the geocentric distance $r = |\mathbf{x}|$ tends to infinity (radiation condition):*

$$\begin{array}{lcll} Lap\, W(\mathbf{x}) & = & 2\omega^2\,, & \mathbf{x} \in \Omega_a \\ V(\mathbf{x}) & \sim & \frac{1}{r} + O\left(\frac{1}{r^3}\right), & r \to \infty \\ W(P) & = & V(P) + Z(P)\,, & P \in S \\ \Gamma(P) & = & |grad\, W(P)|\,, & P \in S\,. \end{array} \tag{1}$$

To deal with small quantities, we have to introduce suitable approximations for the unknowns. The potential $W(r, \beta, \lambda)$ will be approximated through the analytical normal potential $w(r, \beta, \lambda)$. Here the gravity field of an equipotential reference ellipsoid (*Somigliana–Pizzetti* normal field) is often introduced. This potential is analytically easy to handle and shows a symmetry with respect to the earth's mean rotational axis:

$$\begin{array}{lcl} w(r,\beta) & = & v(r,\beta) + z(r,\beta) \\ v(r,\beta) & = & \dfrac{\mu^v}{r}\left[1 - \sum\limits_k \left(\dfrac{a}{r}\right)^k J_k P_k(\sin\beta)\right], \text{ with } \quad k \in \{2,4,6,\ldots,N^v\}\,. \end{array} \tag{2}$$

Alternatively we can combine the well known centrifugal potential

$$z(r,\beta) \;=\; Z(r,\beta) \;=\; \frac{1}{2}\omega^2 r^2 \cos^2\beta \;=\; \frac{1}{2}\frac{\mu^v}{r}\overline{m}\left(\frac{r}{a}\right)^3 \cos^2\beta\,, \tag{3}$$

containing the small constant $\overline{m} = \omega^2 a^3/\mu^v$, together with a truncated spherical harmonic expansion of the gravity potential V (for example up to degree and order $N^v = 20$). In the latter case the difference quantities become smaller, but the analytical amount of the whole problem increases.

To approximate the boundary surface S a telluroid mapping $S \ni P \to s \ni p$ is defined. The most natural choice amongst several possible mappings (Hirvonen, 1960; Grafarend, 1978; Heck, 1986) is Molodensky's telluroid definition with the mapping equations:

$$\begin{array}{lcl} B(p) & = & B(P) \\ L(p) & = & L(P) \\ w(p) - w(p_o) & = & W(P) - W(P_o)\,. \end{array} \tag{4}$$

The telluroid point p lies on the same ellipsoidal normal as P, parameterized through the geographical latitude $B(P)$ and the geographical longitude $L(P)$. It is fixed along this ellipsoidal normal $\mathbf{n}$ in such a way, that p has the same geopotential number in the normal field, as P in the actual field. P_o denotes a global height reference point, e.g. a tide gaugue, while p_o is the corresponding point on the ellipsoid.

Now we are able to formulate difference quantities, which act as new unknowns: The difference of the actual potential $W(Q)$ and the normal potential $w(Q)$ in the same point $Q \in \Omega_a$

$$\delta w \quad = \quad W(Q) - w(Q) \tag{5}$$

is called the *disturbing potential.* And the second (geometrical) unknown, the *height anomaly*

$$\Delta h \quad = \quad h(P) - h(p) \tag{6}$$

is the difference between the ellipsoidal height of P and the ellipsoidal height of the telluroid point p respectively. In the same way both boundary data are redefined: The *potential anomaly*

$$\Delta w \;=\; W(P) - w(p) \tag{7}$$

equals zero if the absolute potential W_o and w_o are selected in the framework of a proper datum definition (Heck and Rummel, 1990; Rummel and Ilk, 1995) in such a way that $\Delta w_o = W(P_o) - w(p_o) = 0$ holds. This fact becomes obvious if we compare equation (4). The scalar *gravity anomaly* $\Delta\gamma$ is defined as follows:

$$\Delta\gamma \;=\; \Gamma(P) - \gamma(p)\,. \tag{8}$$

After these considerations we can reformulate the still non–linear problem: *Suppose the boundary data Δw and $\Delta\gamma$ to be given on S. The unknown disturbing potential $\delta w(\mathbf{x})$ has to fulfill the Laplace equation in the mass free space Ω_a outside S, and the disturbing potential δw tends to zero if the geocentric distance r tends to infinity:*

$$\begin{array}{lcll} Lap\,\delta w(\mathbf{x}) & = & 0\,, & \mathbf{x} \in \Omega_a \\ \delta w(\mathbf{x}) & \sim & \frac{1}{r} + O\left(\frac{1}{r^3}\right)\,, & r \to \infty \\ \Delta w & = & W(P) - w(p) = \delta w(P) + w(P) - w(p) & \\ \Delta\gamma & = & \Gamma(P) - \gamma(p) = |grad\,[\delta w(P) + w(P)]\,| - \gamma(p)\,. & \end{array} \tag{9}$$

Linearization of the reduced boundary condition

In the formulation of the non–linear problem (9) the normal potential w, that occurs in the boundary condition, must be calculated in p and in the boundary point P. Since the vertical position of P is unknown, this formulation of the boundary conditions is unsuitable. Therefore a Taylor series expansion for the disturbing potential δw and the normal potential is set up in the known telluroid point p (Heck, 1988):

$$\begin{array}{lcl} \delta w(P) & = & \delta w(p) + (\partial_i \delta w \cdot n_i)\,\Delta h + \ldots \\ w(P) & = & w(p) + (\partial_i w \cdot n_i)\,\Delta h + \frac{1}{2}\,(\partial_{ij} w \cdot n_i \cdot n_j)\,(\Delta h)^2 + \ldots\,. \end{array} \tag{10}$$

The Taylorstep $\Delta h = h(P) - h(p)$ runs along the ellipsoidal normal $\mathbf{n}$ due to the used telluroid definition of Molodensky (4) and is equal to the height anomaly. The symbols n_i and ∂_i denote the Cartesian coordinates of the ellipsoidal unit normal vector $\mathbf{n}$ and the partial derivatives with respect to these coordinates, referring to an earth–fixed equatorial reference frame. The partial derivatives of second order $\partial_{ij} w$ can be understood as the elements of the Marussi matrix $\mathbf{M}$.

Substituting $\delta w(P)$ and $w(P)$ in the boundary condition Δw by its Taylor series (10) we get the expanded boundary condition, where the terms on the right hand side refer to the telluroid point p:

$$\Delta w \;=\; \delta w + < grad\ w, \mathbf{n} > \Delta h + \cdots . \tag{11}$$

Assuming that $\Delta w = 0$ holds, and neglecting the non–linear terms, we end up with a relationship for the height anomaly (Brun's formula):

$$\Delta h \;=\; -\frac{\delta w}{<grad\ w, \mathbf{n}>} . \tag{12}$$

Substitution of the vectorial gravity disturbance $\delta\boldsymbol{\gamma}(P) = grad\ \delta w(P)$ and the normal gravity vector $\boldsymbol{\gamma}(P) = grad\ w(P)$

$$\begin{aligned} \delta\boldsymbol{\gamma}(P) &= \delta\boldsymbol{\gamma}(p) + (\partial_i \delta\boldsymbol{\gamma} \cdot n_i)\,\Delta h + \ldots \\ \boldsymbol{\gamma}(P) &= \boldsymbol{\gamma}(p) + (\partial_i \boldsymbol{\gamma} \cdot n_i)\,\Delta h + \tfrac{1}{2}\,(\partial_{ij}\boldsymbol{\gamma} \cdot n_i \cdot n_j)\,(\Delta h)^2 + \ldots \end{aligned} \tag{13}$$

in the boundary condition for the gravity anomaly $\Delta\gamma$ results in the linearized boundary condition:

$$\Delta\gamma \;= < \frac{\boldsymbol{\gamma}}{\gamma}, grad\ \delta w > + < \frac{\boldsymbol{\gamma}}{\gamma}, \mathbf{M}\ \mathbf{n} > \Delta h + \ldots . \tag{14}$$

The right hand side of (14) is related to $p \in s$. Now the height anomaly Δh can be eliminated in (14) by use of (12). Neglecting the non–linear terms we end up with the linear reduced boundary condition for the scalar free bvp (Heck, 1989):

$$\Delta\gamma \;= < \frac{\boldsymbol{\gamma}}{\gamma}, grad\ \delta w > + < \frac{\boldsymbol{\gamma}}{\gamma}, \mathbf{M}\ \mathbf{n} > \frac{\delta w}{<\boldsymbol{\gamma}, \mathbf{n}>} . \tag{15}$$

In Heck and Seitz (1991, 1993, 1995) and Seitz et al. (1994) the effects of the non–linear terms have been studied. If a *Somigliana–Pizzetti* field is used as normal potential the effects on the vertical position due to the non–linear terms in the reduced boundary condition are less than 4mm in the vicinity of the earth's surface. They are decreasing to 2mm if a truncated spherical harmonic model (max degree 20) is used to model the normal gravitational potential.

Now the *linearized scalar free bvp* can be formulated where the boundary condition refers to the telluroid $s \ni p$: *Suppose the boundary data $\Delta\gamma$ to be given. The unknown disturbing potential $\delta w(\mathbf{x})$ has to fulfill the Laplace equation in the mass free space Ω_a outside s, and the disturbing potential δw tends to zero if the geocentric distance r tends to infinity – this corresponds to the postulate of regularity at infinity:*

$$\begin{aligned} Lap\ \delta w(\mathbf{x}) &= 0\,, & & \mathbf{x} \in \Omega_a \\ \delta w(\mathbf{x}) &\sim \tfrac{1}{r} + O\left(\tfrac{1}{r^3}\right), & & r \to \infty \\ \Delta\gamma &= < \tfrac{\boldsymbol{\gamma}}{\gamma}, grad\ \delta w > + < \tfrac{\boldsymbol{\gamma}}{\gamma}, \mathbf{M}\ \mathbf{n} > \tfrac{\delta w}{<\boldsymbol{\gamma}, \mathbf{n}>}\,, & & \text{on } s\,. \end{aligned} \tag{16}$$

The boundary condition (15) in the formulation (16) can be expressed in operator style (Heck, 1991):

$$\Delta\gamma \;=\; B_s\{\delta w\} \;=\; E_s \circ D\{\delta w\}\,. \tag{17}$$

The boundary operator B_s maps the disturbing potential δw at the telluroid s onto the scalar free gravity anomaly $\Delta\gamma$. B_s is composed of the differential operator D and the evaluation operator E_s which are applied subsequently. The evaluation operator E_s relates the differential $D\{\delta w\}$ to the boundary surface s. The linear operator D, applied to the spatial function δw, can be represented by the identity operator I and the partial derivatives with respect to the spherical coordinates r, β and λ:

$$D \;=\; d_o I + d_r \frac{\partial}{\partial r} + d_\beta \frac{\partial}{\partial \beta} + d_\lambda \frac{\partial}{\partial \lambda}\,. \tag{18}$$

The coefficients d_o, d_r, d_β and d_λ of the differential operator are functionals of the normal field w:

$$\begin{aligned} d_o &= -<\frac{\boldsymbol{\gamma}}{\gamma}, \mathbf{M}\,\mathbf{n}>\frac{1}{<\boldsymbol{\gamma},\mathbf{n}>} \\ d_r &= \frac{\gamma_1}{\gamma} \\ d_\beta &= \frac{1}{r}\frac{\gamma_2}{\gamma} \\ d_\lambda &= \frac{1}{r\cos\beta}\frac{\gamma_3}{\gamma} \equiv 0, \text{ for } w(r,\beta). \end{aligned} \tag{19}$$

In (19), the components of the normal gravity vector $\boldsymbol{\gamma}(p)$ are denoted by γ_i and refer to the local orthonormal system $\{p; \mathbf{g}_i\}$. The basis vector $\mathbf{g}_1$ is parallel to the geocentric position vector $\mathbf{x}(p)$, $\mathbf{g}_2$ points to the north and $\mathbf{g}_3$ completes this right handed system. A survey about the linearization of boundary value problems is given in Heck (1997).

Series expansion of the differential operator

To obtain an analytical representation of the coefficients of the differential operator D, the normal potential w has to be described analytically. Therefore we introduce the following approximation w_a for the normal potential w:

$$w_a(r,\beta) = \frac{\mu^v}{r}\left[1 - J_2\left(\frac{a}{r}\right)^2 P_2(\sin\beta) - J_4\left(\frac{a}{r}\right)^4 P_4(\sin\beta) + \frac{1}{2}\overline{m}\left(\frac{r}{a}\right)^3\cos^2\beta\right]. \tag{20}$$

Under this approximation the coefficients d_o, d_r, d_β and d_λ are represented in the following analytical way. This new **second order approximation** of the differential operator D is reached after extensive analytical manipulations. For details see Seitz (1997).

$$\begin{aligned} d_o &= -\frac{2}{r}\Big\{1 - 3J_2P_2 + \tfrac{3}{2}\overline{m}\cos^2\beta \\ &\quad - \tfrac{1}{4}e^2\sin^2\beta\left[2e^2\cos^2\beta - 12J_2\left(2 - 3\sin^2\beta\right) + 13\overline{m}\cos^2\beta\right] \\ &\quad + \tfrac{3}{2}\tfrac{h}{a}\left(4J_2P_2 + 3\overline{m}\cos^2\beta\right) - \tfrac{9}{4}J_2^2\left(1 - 10\sin^2\beta + 13\sin^4\beta\right) \\ &\quad - \tfrac{3}{4}J_2\overline{m}\cos^4\beta + \tfrac{3}{2}\overline{m}^2\left(1 - 3\sin^2\beta + 2\sin^4\beta\right) - 10J_4P_4 + O(e^6)\Big\} \\ d_r &= -1 + \tfrac{1}{2}\sin^2\beta\cos^2\beta\left(3J_2 + \overline{m}\right)^2 + O(e^6) \\ d_\beta &= -\tfrac{1}{r}\sin\beta\cos\beta\Big\{3J_2 + \overline{m} \\ &\quad + 3\left(\tfrac{1}{2}e^2\sin^2\beta - \tfrac{h}{a}\right)\left(2J_2 - \overline{m}\right) + 9J_2^2P_2 \\ &\quad + \tfrac{3}{2}J_2\overline{m}\left(1 + \sin^2\beta\right) + \overline{m}^2\cos^2\beta - \tfrac{5}{2}J_4\left(3 - 7\sin^2\beta\right) + O(e^6)\Big\} \\ d_\lambda &= O(e^6), \quad [\equiv 0, \text{ if } w = w(r,\beta)]. \end{aligned} \tag{21}$$

The absolute error of the neglected terms in the boundary condition is less than $1 \cdot 10^{-10} ms^{-2}$ if a *Somigliana–Pizzetti* normal field is used as reference field. If a truncated spherical harmonic expansion ($N^v = 20$) is used as normal gravitational potential v, the absolute error increases to $5 \cdot 10^{-8} ms^{-2}$.

The coefficient d_o can be split off into a dominant term, caused by the isotropic part μ^v/r of the normal potential, and the so–called *ellipsoidal terms.* The anisotropy of v and the influence of the centrifugal potential z contribute to the *ellipsoidal term* δd_o:

$$d_o = -\frac{2}{r} + \delta d_o. \tag{22}$$

The same decomposition can be applied for the radial term

$$d_r = -1 + \delta d_r. \tag{23}$$

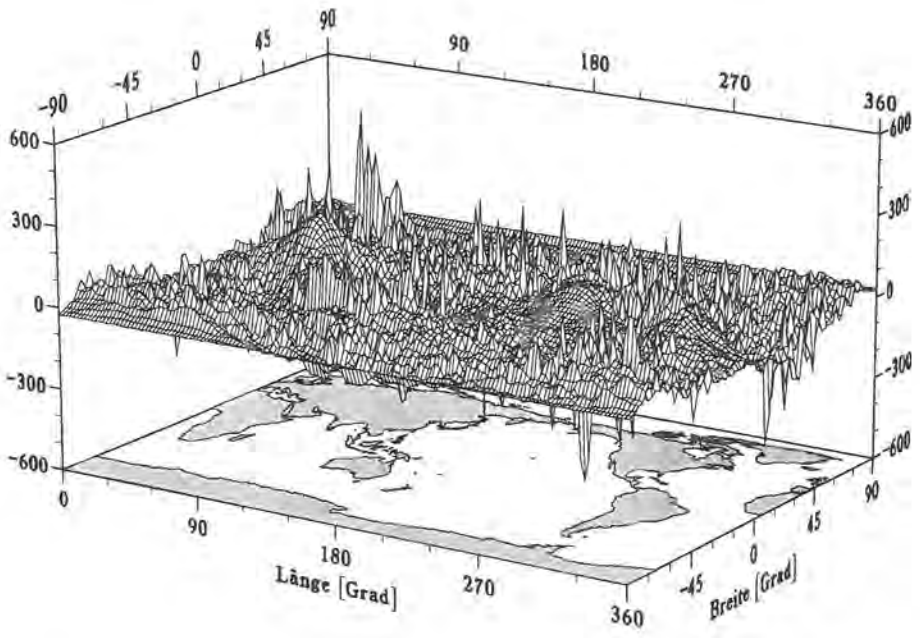


Figure 1: Ellipsoidal terms $[10^{-8} ms^{-2}]$ in the reduced boundary condition of the scalar free bvp. V: OSU91a1f, w: GRS80.

The linear boundary condition (15), (17) can now be written as

$$\Delta\gamma \;=\; E_s\left\{-\tfrac{2}{r}\delta w - \tfrac{\partial \delta w}{\partial r}\right\} + E_s \circ \delta D\left\{\delta w\right\} . \tag{24}$$

The linear differential operator D is here decomposed in an isotropic part and the ellipsoidal increments:

$$\begin{aligned} D &= -\tfrac{2}{r}I - \tfrac{\partial}{\partial r} + \delta d_o I + \delta d_r \tfrac{\partial}{\partial r} + d_\beta \tfrac{\partial}{\partial \beta} + d_\lambda \tfrac{\partial}{\partial \lambda} \\ &= -\tfrac{2}{r}I - \tfrac{\partial}{\partial r} + \delta D . \end{aligned} \tag{25}$$

If only the isotropic term μ^v/r is considered and – pay attention to this fact – the centrifugal potential is omitted, one deals with the **isotropic** or **radial approximation** of the differential operator D. This leads to the *fundamental equation of Physical Geodesy* (Heiskanen and Moritz, 1967):

$$\Delta\gamma \;=\; -\tfrac{2}{r}\delta w - \tfrac{\partial \delta w}{\partial r} . \tag{26}$$

In this rough approximation of the boundary condition the ellipsoidal terms δd_o, δd_r, d_β and d_λ are neglected. The ellipsoidal terms in the linear reduced boundary condition of the scalar free bvp are given with (25) through the expression:

$$E_s \circ \delta D\left\{\delta w\right\} \;=\; E_s\left\{\delta d_o \delta w + \delta d_r \frac{\partial \delta w}{\partial r} + d_\beta \frac{\partial \delta w}{\partial \beta} + d_\lambda \frac{\partial \delta w}{\partial \lambda}\right\} . \tag{27}$$

With the spherical harmonic model OSU91a1f from the Ohio State University (Rapp et al., 1991) representing the actual gravitational field of the earth and the Geodetic Reference System 1980 (GRS80) as reference field, the total ellipsoidal terms (27) are illustrated in figure 1. The maximum values are in the range of $\pm 600 \cdot 10^{-8} ms^{-2}$.

In the geodetic literature since Jekeli (1981), Cruz (1986) or Pavlis(1988) the ellipsoidal correction terms ε_γ and ε_h are customary. They are applied to the boundary data. The corrected gravity anomalies are now related to the isotropic approximation of the boundary condition

(24). The correction terms ε_γ and ε_h (Jekeli, 1981; there was a misprint in the sign for ε_h) are proportional to the coefficients δd_o and d_β of the differential operator:

$$\begin{aligned} \varepsilon_\gamma &= \delta d_o\, \delta w &&= \tfrac{1}{r}\left[6J_2\left(\tfrac{a}{r}\right)^2 P_2(\sin\beta) - \tfrac{3\omega^2 r^3}{\mu^v}\cos^2\beta\right]\delta w \\ \varepsilon_h &= d_\beta\, \tfrac{\partial \delta w}{\partial \beta} &&= -\tfrac{1}{r}e^2 \sin\beta\cos\beta\, \tfrac{\partial \delta w}{\partial \beta}\,. \end{aligned} \tag{28}$$

The ellipsoidal term $E_s\{\delta d_o \delta w\}$ has a very smooth behaviour which can be seen in figure 2. The ellipsoidal term $E_s\{d_\beta \frac{\partial \delta w}{\partial \beta}\}$ is depicted in figure 3.

The usually applied correction terms ε_γ and ε_h are a **first order approximation**. The term ε_γ describes the influence of the difference between the isotropic field and the exact normal field in the boundary condition. One can assess that $|\varepsilon_\gamma| \leq 230 \cdot 10^{-8} ms^{-2}$ holds in the vicinity of the earth's surface. In the first order approximation the term ε_h corrects for the fact, that the partial derivative with respect to the geocentric distance r instead of the derivative in direction to the ellipsoidal normal is applied to the disturbing potential δw in the boundary condition. The terms ε_γ and ε_h (28) are often further simplified by taking advantage of the relations (Heiskanen and Moritz, 1967, p78):

$$\begin{aligned} J_2 &= \tfrac{1}{3}\left(e^2 - \overline{m}\right) + \cdots \\ \overline{m} &= \tfrac{e^2}{2} + \cdots\,, \end{aligned} \tag{29}$$

which are first order approximations. It should be noticed that these relations (29) are only valid when a normal field of *Somigliana–Pizzetti*–type is used! With these assumptions the representation

$$\begin{aligned} \varepsilon_\gamma &\approx -\tfrac{1}{r}e^2\left(2 - 3\sin^2\beta\right)\delta w \\ \varepsilon_h &\approx -\tfrac{1}{r}e^2 \sin\beta\cos\beta\, \tfrac{\partial \delta w}{\partial \beta} \end{aligned} \tag{30}$$

results that is mostly referred to in geodetic literature in the context of ellipsoidal corrections (Lelgemann, 1970; Pellinen, 1982; Martinec, 1995).

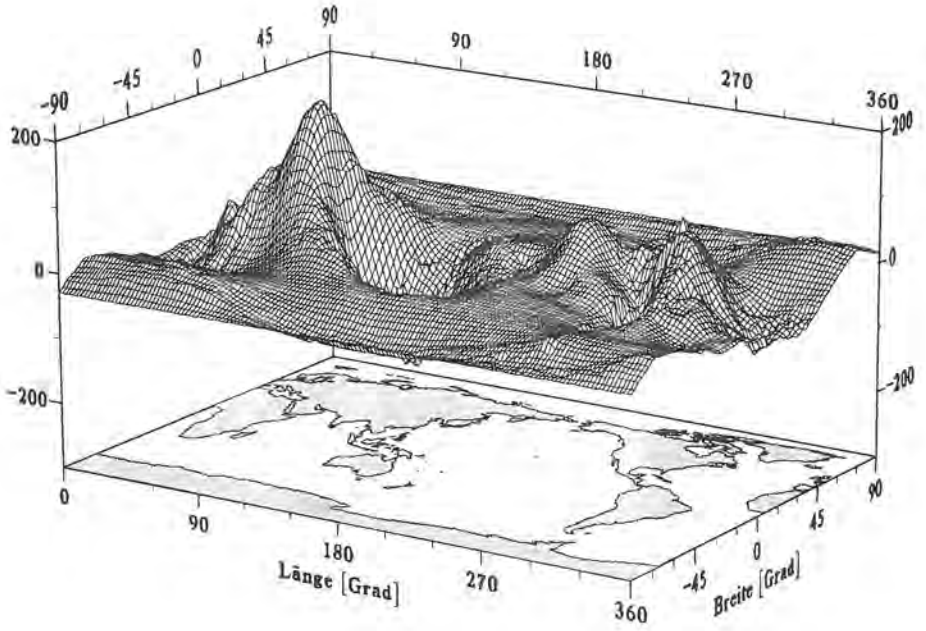


Figure 2: Ellipsoidal term $E_s\{\delta d_o \delta w\}$ $[10^{-8} ms^{-2}] \sim \varepsilon_\gamma$ (Jekeli, 1981; Cruz, 1986). V: OSU91a1f, w: GRS80.

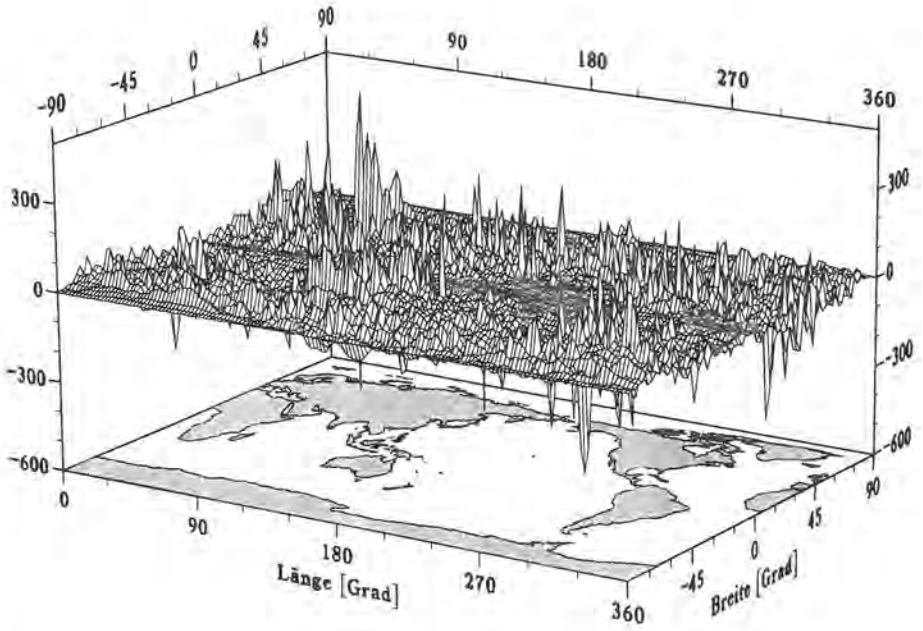


Figure 3: Ellipsoidal term $E_s\{d_\beta \frac{\partial \delta w}{\partial \beta}\}$ $[10^{-8} ms^{-2}] \sim \varepsilon_h$ (Jekeli, 1981; Cruz, 1986). V: OSU91a1f, w: GRS80.

Formulation of an isotropic problem

The object of our efforts is the determination of the harmonic coefficients $\overline{C}_{nm}^{\delta w}$, $\overline{S}_{nm}^{\delta w}$ which represent the disturbing potential δw. This task can be performed for example by harmonic analysis of boundary data that meets an isotropic boundary condition on a surface, axi-symmetric with respect to the earth's mean rotational axis. The simplest choice of such a surface is a sphere of radius a. On a sphere $K \ni k$ the corresponding boundary condition reads:

$$\Delta\gamma_k \;=\; B_k\{\delta w\} \;=\; E_k \circ D_i\{\delta w\} \;:=\; \Delta\gamma - \Delta_k \,. \tag{31}$$

The linear differential operator D_i consists of the identity operator I and the partial derivative with respect to the geocentric distance r. The differential $D_i\{\delta w\}$ is restricted to the surface of the sphere by applying the evaluation operator E_k. The resulting spherical bvp

$$\begin{array}{lcll} Lap\ \delta w(\mathbf{x}) & = & 0\,, & \mathbf{x} \in \Omega_a \\ \delta w(\mathbf{x}) & \sim & \frac{1}{r} + O\left(\frac{1}{r^3}\right), & r \to \infty \\ \Delta\gamma_k & := & \Delta\gamma - \Delta_k \;=\; -\frac{2}{a} E_k\{\delta w\} - E_k\left\{\frac{\partial \delta w}{\partial r}\right\} & \end{array} \tag{32}$$

is formally the third boundary value problem on a sphere. The boundary data we have to deal with is the original boundary data $\Delta\gamma$ (computed with the full normal field) reduced by the **ellipsoidal and topographic correction** term Δ_k.

In a similar way we can formulate the boundary condition on the surface of an ellipsoid of revolution $E \ni e$:

$$\Delta\gamma_e \;=\; B_e\{\delta w\} \;=\; E_e \circ D_i\{\delta w\} \;:=\; \Delta\gamma - \Delta_e \,. \tag{33}$$

The terms Δ_k or Δ_e correct for the anisotropy of the normal field and the difference between the telluroid s and the surface of a sphere or an ellipsoid respectively. In the following we will

restrict ourselves to the isotropic boundary value problem on the surface of an ellipsoid:

$$\begin{aligned} Lap\ \delta w(\mathbf{x}) &= 0, && \mathbf{x} \in \Omega_a \\ \delta w(\mathbf{x}) &\sim \tfrac{1}{r} + O\left(\tfrac{1}{r^3}\right), && r \to \infty \\ \Delta\gamma_e &:= \Delta\gamma - \Delta_e = -E_e\left\{\tfrac{2}{r}\delta w\right\} - E_e\left\{\tfrac{\partial \delta w}{\partial r}\right\}. \end{aligned} \tag{34}$$

The unknown disturbing potential δw still has to fulfill the requirements (16). The reduced boundary data $\Delta\gamma_e$ on the surface of an ellipsoid $E \ni e$ has to fulfill the isotropic boundary condition on E. It is obvious, that the ellipsoidal and topographic correction terms Δ_e are functionals of δw which we solve for. Therefore an iterative procedure is required.

Analytical continuation of the boundary condition onto an ellipsoid of revolution

To come up with a representation of the boundary condition on the surface of an ellipsoid (33) we have to analytically continue the boundary condition from the telluroid onto E. This is done by a formal Taylor series expansion of the evaluation operator E_s. To that end we select the Taylorstep in radial direction:

$$r(p) - r(e) = h\left(1 + \frac{1}{2}e^4 \sin^2\beta\cos^2\beta\right) + a \cdot O(e^8). \tag{35}$$

If (35) is divided by a we have the representation

$$\frac{r(p) - r(e)}{a} = \frac{h}{a}\left(1 + \frac{1}{2}e^4 \sin^2\beta\cos^2\beta\right) + O(e^8), \tag{36}$$

where the expression h/a is of the order of e^2. The formal Taylor expansion of E_s or E_e between the telluroid point $p \in s$ and the corresponding point $e \in E$ on the surface of the ellipsoid can be performed with the Taylorpoint situated either on E or s:

If we set up the Taylor expansion for E_e, which we need for a formulation like (34), in $p \in s$ we get from (36) the representation

$$E_e = E_s + \sum_{n=1}^{\infty} \frac{a^n}{n!}\left(\frac{r(e) - r(p)}{a}\right)^n E_s \circ \frac{\partial^n}{\partial r^n}. \tag{37}$$

Rearranging (37) with respect to the zero order term E_s and substituting this representation from (24) we have the new representation of the boundary condition

$$\Delta\gamma = B_s\{\delta w\} = \underbrace{E_e\left\{-\frac{2}{r}\delta w - \frac{\partial \delta w}{\partial r}\right\}}_{\text{isotropic term}} + \Delta_e, \tag{38}$$

The whole ellipsoidal and topographical components are included in the term

$$\Delta_e = c^{se}_{r_0\beta_1} E_s\left\{\frac{\partial \delta w}{\partial \beta}\right\} + \sum_{i=0}^{8} c^{se}_{r_i\beta_0} E_s\left\{\frac{\partial^i \delta w}{\partial r^i}\right\}. \tag{39}$$

The partial derivatives of the disturbing potential have to be computed at the telluroid. The coefficients $c^{se}_{r_i\beta_j}$ have been derived in Seitz (1997).

Table 1. Error $[10^{-5} ms^{-2}]$ in the analytical approximation of Δ_k depending on the order of the Taylor series. V: OSU91a1f, w: GRS80, $w_a(J_0, J_2, J_4, \overline{m})$.

max. order of $\frac{\partial^i}{\partial r^i}$	Extreme values of the approximation error	
8	−0.001	0.001
7	−0.008	0.010
6	−0.053	0.073
5	−0.326	0.463
4	−1.922	2.562
3	−9.865	11.974
2	−41.280	45.350
1	−133.227	129.352

Table 2. Error $[10^{-5} ms^{-2}]$ in the analytical approximation of Δ_e depending on the order of the Taylor series. V: OSU91a1f, w: GRS80, $w_a(J_0, J_2, J_4, \overline{m})$.

max. order of $\frac{\partial^i}{\partial r^i}$	Extreme values of the approximation error	
8	−0.001	0.001
7	−0.001	0.001
6	−0.001	0.001
5	−0.001	0.001
4	−0.005	0.005
3	−0.127	0.153
2	−2.523	1.435
1	−16.819	28.744

As an alternative to this procedure, related to the question raised by Sansò and Sona (1995), Sansò (1995) about the correct choice of the Taylorpoint we also expanded the ellipsoidal and topographical terms with the Taylorpoint in $e \in E$. Here we get directly the representation of the evaluation operator E_s which we have to substitute from (24)

$$E_s = E_e + \sum_{n=1}^{\infty} \frac{a^n}{n!} \left(\frac{r(p) - r(e)}{a} \right)^n E_e \circ \frac{\partial^n}{\partial r^n} . \tag{40}$$

The resulting ellipsoidal and topographical terms

$$\Delta_e = \sum_{i=0}^{4} c^{es}_{r_i \beta_1} E_e \left\{ \frac{\partial^{i+1} \delta w}{\partial r^i \partial \beta} \right\} + \sum_{i=0}^{8} c^{es}_{r_i \beta_0} E_e \left\{ \frac{\partial^i \delta w}{\partial r^i} \right\} . \tag{41}$$

are now related to the surface of the ellipsoid. It is obvious that the coefficients $c^{es}_{r_i \beta_j}$ are different from $c^{se}_{r_i \beta_j}$, also in there signs. The upper limits of the Taylor series in the alternative developments (38), (39) and (40), (41), respectively, have been chosen such that the same absolute error level of $\pm 1 \cdot 10^{-8} ms^{-2}$ is achieved, which was verified by numerical calculations on the basis of OSU91a1f in Seitz (1997). The ellipsoidal and topographical terms Δ_e are in the range of $\pm 20 \cdot 10^{-5} ms^{-2}$ as can be seen in figure 4. The effect of neglecting the ellipsoidal and topographical terms – using the boundary data $\Delta\gamma$ without applying a correction for the anisotropy of the normal potential and for the geometrical distance between the telluroid and the ellipsoid – on the vertical position of equipotential surfaces in the vicinity of the earth's surface is plotted in figure 5. The total effect can amount up to nearly $2m$.
In a similar way the evaluation operator E_s is continued in Seitz (1997) onto a sphere. The resulting approximation errors for Δ_k and Δ_e are listed in the tables 1 and 2 respectively. They

are also given for different maximum orders of the partial derivative in radial direction. To achieve a maximum error of $\pm 1 \cdot 10^{-8} ms^{-2}$ in case of the continuation to a sphere one has to perform the Taylor expansion up to the order $k = 8$. The Taylor expansion for the evaluation operator can be truncated after the 5th order in case of the ellipsoidal boundary.

A flow chart of the whole process starting from the non–linear boundary condition, the linearization, the different levels of approximation for the differential operator and the analytical continuation of the boundary condition onto the surface of an ellipsoid is given in table 3.

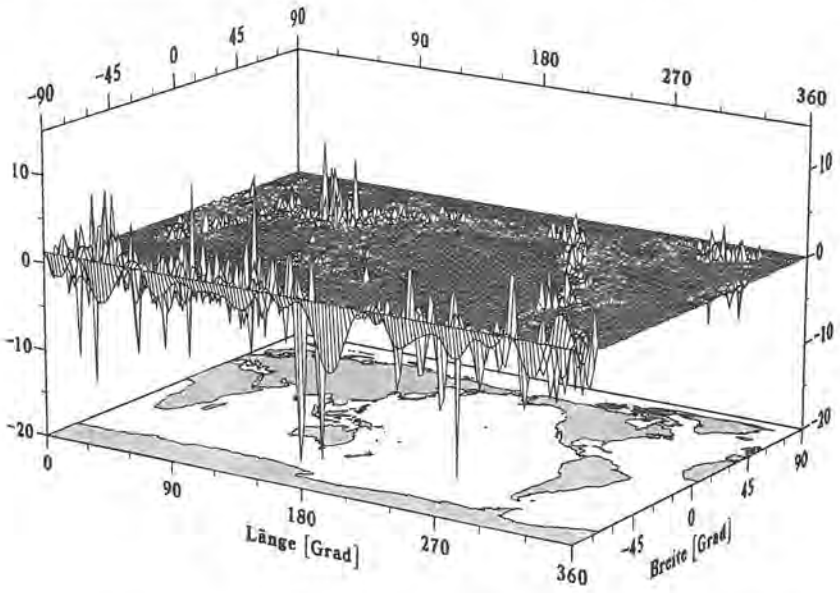


Figure 4: Ellipsoidal and topographical terms Δ_e $[10^{-5} ms^{-2}]$. V: OSU91a1f, w: GRS80.

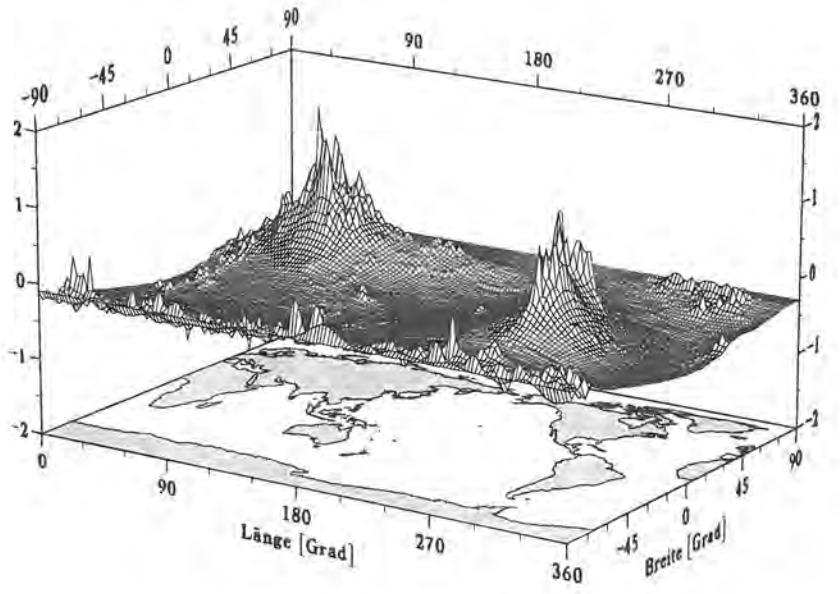
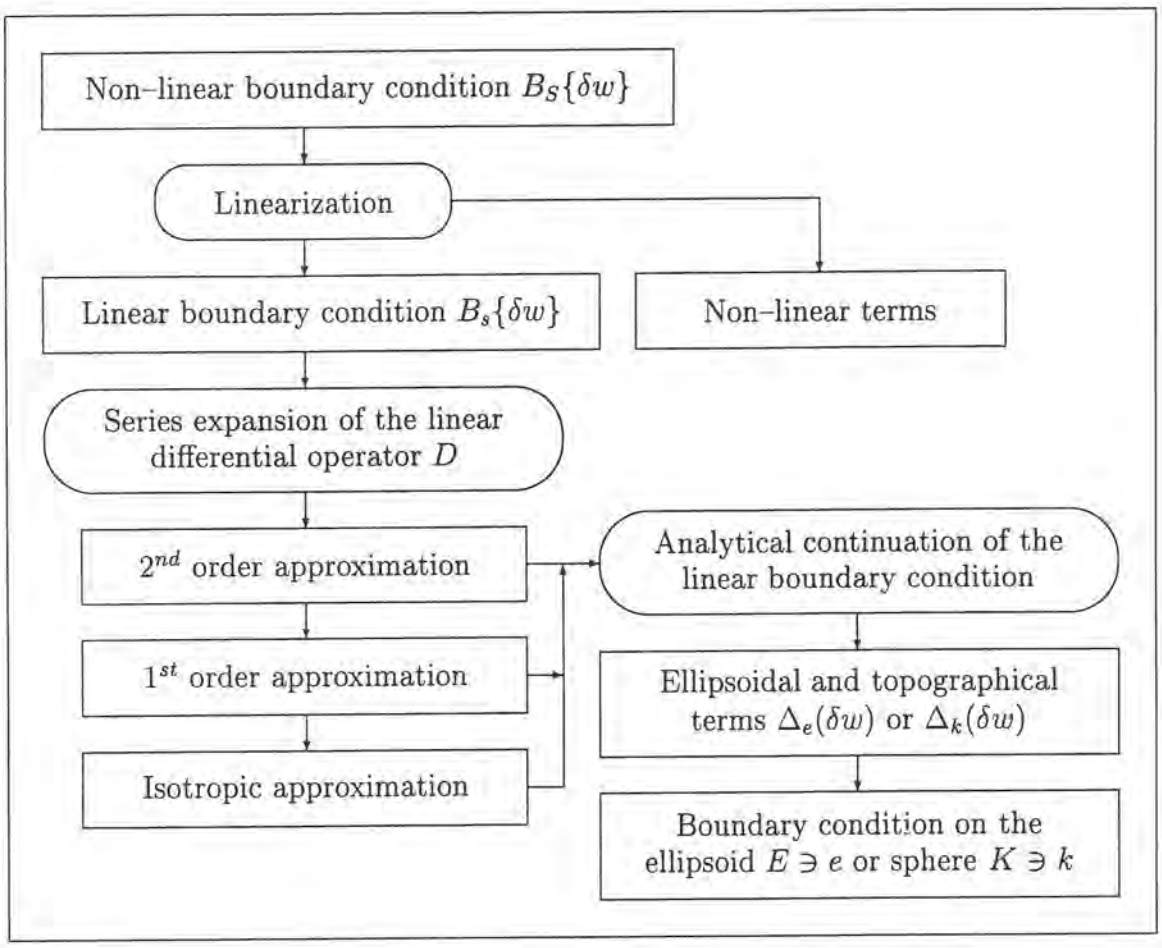


Figure 5: Effect $[m]$ due to Δ_e on the vertical position of equipotential surfaces. V: OSU91a1f, w: GRS80.

Table 3. Flow chart for the analytical evaluation of the boundary condition for the scalar free bvp.



An iterative solution

As already mentioned the ellipsoidal and topographical correction terms ((39) or (41)) are functionals of the unknown disturbing potential which we solve for. If we try to solve the bvp by a harmonic analysis of the reduced boundary data (33) on the surface of an ellipsoid we have to set up an iterative procedure. First numerical tests indicated that the whole procedure diverges if we continue the data onto a sphere. It converges when we use a surface of an ellipsoid, on which we perform the harmonic analysis. Further results will be given in Seitz and Heck (1999).

References

Cruz JY (1986) Ellipsoidal Corrections to Potential Coefficients Obtained from Gravity Anomaly Data on the Ellipsoid. Dept. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio, Rep. No. 371.

Grafarend EW (1978) The Definition of the Telluroid. Bull Géod 52: 25–37.

Grafarend EW, Heck B, Knickmeyer EH (1985) The free versus fixed geodetic boundary value problem for different combinations of geodetic observables. Bull Géod 59: 11–32.

Heck B (1986) A Numerical Comparison of Some Telluroid Mappings. In: Sansò F (ed.), Proc. I. Hotine–Marussi Symposium on Mathematical Geodesy, Roma June 3–6 1985. Milano 1986, 19–38.

Heck B (1988) The non–linear geodetic boundary value problem in quadratic approximation. Manuscr Geod 13: 337–348.

Heck B (1989) A contribution to the scalar free boundary value problem of physical geodesy. Manuscr Geod 14: 87–99.

Heck B (1991) On the linearized boundary value problem of physical geodesy. Dept. of Geodetic Science

and Surveying, The Ohio State Univ., Columbus, Ohio, Rep. No. 407.

Heck B (1997) Formulation and Linearization of Boundary Value Problems: From Observables to a Mathematical Model. In: Sansò F, Rummel R (eds.), Lecture Notes of the international Summer School *Boundary Value Problems and the Modeling of the Earth's Gravity Field in View of the One Centimeter Geoid*, Como 1996. Lecture Notes in Earth Sciences No. 65, Springer Verlag, 1997, 121–160.

Heck B, Rummel R (1990) Strategies for solving the vertical datum problem using terrestrial and satellite geodetic data. In: Sünkel H, Baker T (eds.), Sea Surface Topography and the Geoid. IAG Symposia No. 104, Springer Int., 116–128.

Heck B, Seitz K (1991) Non–linear effects in the scalar free geodetic boundary problem – Spherical harmonic representation of the effects on the reduced boundary condition, the disturbing potential and the vertical position correction –. Technical Report No. 14, Inst. of Geodesy, Univ. of Stuttgart, 71pp.

Heck B, Seitz K (1993) Effects of non–linearity in the geodetic boundary value problems. Deutsche Geodätische Kommission, Reihe A, Heft Nr. 109, Munich, 74pp.

Heck B, Seitz K (1995) Non–linear Effects in the Geodetic Version of the Free GBVP Based on Higher Order Reference Fields. In: Sansò F (ed.), Geodetic Theory Today, Proc. III. Hotine–Marussi Symposium on Mathematical Geodesy, L'Aquila May 30 – June 3 1994, Italy. IAG Symposium No. 114, Springer 1995, 332-339.

Heiskanen WH, Moritz H (1967) Physical geodesy. WH Freeman & Co., San Francisco and London.

Hirvonen RA (1960) New Theory of the Gravimetric Geodesy. Publications of the Isostatic Institute of the IAG, 32, Helsinki.

Jekeli C (1981) The Downward Continuation to the Earth's Surface of Truncated Spherical and Ellipsoidal Harmonic Series of the Gravity and Height Anomalies. Dept. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio, Rep. No.323.

Klees R (1992) Lösung des fixen geodätischen Randwertproblems mit Hilfe der Randelementmethode. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 382, München, 314pp.

Klees R (1995) Perturbation Expansion for Solving the Fixed Gravimetric Boundary Value Problem. In: Sansò F (ed.), Geodetic Theory Today, Proc. III. Hotine–Marussi Symposium on Mathematical Geodesy, L'Aquila May 30 – June 3 1994, Italy. IAG Symposium No. 114, Springer 1995, 340-349.

Lelgemann D (1970) Untersuchungen zu einer genaueren Lösung des Problems von Stokes. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 155, München.

Martinec Z (1995) Boundary value problems for gravimetric determination of a precise geoid. Department of Geophysics, Faculty of Mathematics and Physics, Charles University Prague, Czech Republic, 159pp.

Pavlis NK (1988) Modeling and Estimation of a Low Degree Geopotential Model from Terrestrial Gravity Data. Dept. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio, Rep. No. 386.

Pellinen LP (1982) Effects of the earth ellipticity on solving geodetic boundary value problem. Boll. Geod. Sci. Affini, 89-103.

Rapp RH, Wang YM, Pavlis NK (1991) The Ohio State 1991 Geopotential and Sea Surface Topography Harmonic Coefficient Models. Dept. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio, Rep. No. 410.

Rummel R, Ilk KH (1995) Height datum connection - the ocean part. Allgemeine Vermessungsnachrichten, Heft 8/9, Wichmann Verlag, Heidelberg, 321–330.

Sacerdote F, Sansò F (1986) The scalar boundary value problem of physical geodesy. Manuscr Geod 11: 15–28.

Sansò F (1995) The long road from measurements to boundary value problems in physical geodesy. Manuscr Geod 20: 326–344.

Sansò F, Sona G (1995) Gravity reductions versus approximate B.V.P.'s. In: Sansò F (ed.), Geodetic Theory Today, Proc. III. Hotine–Marussi Symposium on Mathematical Geodesy, L'Aquila May 30 – June 3 1994, Italy. IAG Symposium No. 114, Springer 1995, 304-314.

Seitz K, Schramm B, Heck B (1994) Non–linear Effects in the Scalar Free Geodetic Boundary Value Problem Based on Reference Fields of Various Degrees. Manuscr Geod 19: 327–338.

Seitz K (1997) Ellipsoidische und topographische Effekte im geodätischen Randwertproblem. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 483, München, 140pp.

Seitz K, Heck B (1999) Comparison of the behaviour of an iterative solution of the scalar free bvp on a sphere and ellipsoid. (in preparation)

Ellipsoidal corrections for the inverse Hotine/Stokes formulas

Michael G. Sideris, Zhiling L. Fei and J.A. Rod Blais

ABSTRACT

The ellipsoidal corrections respectively for the spherical gravity disturbance (computed using the inverse Hotine formula) and the spherical gravity anomaly (computed using the inverse Stokes formula) are derived. The corrections consist of two parts: the simple analytical function part and the integral part. The input data are respectively the spherical gravity disturbances and the spherical gravity anomalies and the disturbing potentials, which are already available in some local areas and can be computed globally from the geopotential models such as EGM96. Further discussions on the integral part such as the singularity, the input data and the expansion into a series of spherical harmonics are included.

Keywords: Gravity; Disturbing potential; Inverse Stokes's formula; Inverse Hotine's formula; Ellipsoidal correction

1. Introduction

The satellite altimetry technique provides direct measurements of sea surface heights with respect to the reference ellipsoid, the geometrical reference surface of the Earth. Since 1973, a series of altimetry satellites such as SKYLAB, GEOS-3, SEASAT, GEOSAT, ERS-1 and TOPEX have been launched and have collected data over the oceans. Owing to instrument improvement, geophysical and environmental correction improvement and radial orbit error reduction, the precision of satellite altimetry measurements has improved from the 3-metre to the 2-centimetre level. The resolution of satellite altimeter data along the tracks has also come down from 70 km to 20 km (see Zhang 1993). Tremendous amounts of satellite altimeter data with very high precision have been collected since the advent of the satellite altimetry. After subtracting the dynamic sea surface topography, satellite altimetry can provide an estimation of the geoidal height N in ocean areas with a level of precision of about 10 cm [Rummel and Haagmans (1990)]. These geoidal height data can be used to recover the gravity disturbances and gravity anomalies over the oceans.

Papers reporting current results on recovering the gravity information from satellite altimeter data, and in some cases, a review of prior work, include those of Zhang and Blais (1995), Hwang and Parsons (1995), Olgiati et al. (1995), Sandwell and Smith (1996) and Kim (1996). The models employed for recovering the gravity information from the satellite altimeter data are mainly the spherical harmonic expansion of the disturbing potential, the Hotine/Stokes formulas and the inverse Hotine/Stokes formulas. The gravity disturbances/anomalies obtained via these models might as be called the spherical gravity disturbances/anomalies since these models are valid under the spherical approximation. In these models, the input and output data are supposed to be given on a sphere, the mean sphere. Unfortunately, the geoidal height N (disturbing potential T) from the altimetry and the gravity disturbance/anomaly $\delta g/\Delta g$ to be computed from N refer to the geoid which is very close to the reference ellipsoid S_e. They satisfy the following relations:

$$\Delta T(P) = 0 \qquad (P \text{ is outside } S_e) \tag{1.1}$$

$$T(P) = O(\frac{1}{r_P^3}) \qquad (P \text{ is at infinity}) \tag{1.2}$$

$$\frac{\partial}{\partial h_P} T(P) = -\delta g(P) \qquad (P \text{ is on } S_e) \tag{1.3}$$

$$\Delta g(P) = \delta g(P) + \frac{1}{\gamma_P}\frac{\partial \gamma_P}{\partial h_P} \qquad (P \text{ is on } S_e) \tag{1.4}$$

$$T(P) = \gamma_P N(P) \qquad (P \text{ is on } S_e) \tag{1.5}$$

where r_P is the radius of point P and $\frac{\partial}{\partial h_P}$ is the derivative along the ellipsoidal normal direction of P.

The maximum difference between S_e and the mean sphere is about 100 m. So we can treat the data given on the geoid as the data on the reference ellipsoid. The relative error caused by doing so is about the order of 10^{-4}. However, the relative error of substituting the reference ellipsoid by the mean sphercial surface is about the order of 3×10^{-3}. The effects of this error on the gravity anomaly and gravity disturbance, which are also called the effects of the Earth's flattening, may reach about 0.3 mGal. When the aim of the satellite altimetry is to recover the gravity information with accuracy less than 1 mGal, the effects of the Earth's flattening should be considered.

In order to reduce the effects of the Earth's flattening on the gravity anomaly, Wang (1999) proposed to add an ellipsoidal correction term to the spherical gravity anomaly recovered from the altimetry data via the inverse Stokes formula. The ellipsoidal correction is expressed by the integral formulas and in series of spherical harmonic expansions. In the integral formulas, an auxiliary function χ is needed for computing the ellipsoidal correction Δg^1 from the disturbing potential T, that is:

$$T \xrightarrow{\text{global integral}} \chi \xrightarrow{\text{global integral}} \Delta g^1$$

In this paper, we will derive new ellipsoidal correction formulas respectively to the spherical gravity disturbances and the spherical gravity anomalies. They consist of two parts: a simple analytical part and an integral part. The input data are respectively the spherical gravity disturbances and the spherical gravity anomalies and the disturbing potentials, which are already computed from altimetry data in some ocean areas with a high accuracy or are computed approximately from the Earth Models.

2. Formulas for the ellipsoidal corrections to the spherical gravity disturbance and the spherical gravity anomaly

In this section, we will

(a) establish an integral equation, which shows the relation between the geoidal heights and the gravity disturbances on the reference ellipsoid;
(b) solve the integral equation to get the formula for the ellipsoidal correction to the inverse Hotine's formula (the spherical gravity disturbance);
(c) derive the formula for the ellipsoidal correction to the inverse Stokes formula (the spherical gravity anomaly) from the result of (b);

Establishment of the integral equation

It is easy to prove that for an arbitrarily point P_0 given inside S_e, the function

$$F(Q,P_0)\equiv\frac{r_Q^2-r_{P_0}^2}{r_{P_0}l_{QP_0}^3}=2\frac{\partial}{\partial r_Q}(\frac{1}{l_{PP_0}})-\frac{1}{r_{P_0}}(\frac{1}{l_{QP_0}}) \tag{2.1}$$

satisfies

$$\begin{cases}\Delta F(Q,P_0)=0 & (Q \text{ is outside } S_e)\\ \lim\limits_{Q\to\infty}F(Q,P_0)=0 & \\ F(Q,P_0) \text{ is continuously differentiable on and outside } S_e\end{cases} \tag{2.2}$$

According to Green's second identity (Heiskanen and Moritz, 1962), we obtain that for an arbitrary function V that is harmonic and regular outside S_e and continuously differentiable on and outside S_e,

$$\int_{S_e}V(Q)\frac{\partial F(Q,P_0)}{\partial h_Q}dS_{eQ}=\int_{S_e}\frac{\partial}{\partial h_Q}V(Q)F(Q,P_0)dS_{eQ} \tag{2.3}$$

Let V in (2.3) be the disturbing potential T and V_a defined by (A1-1) in the Appendixes respectively. Then from (1.1), (1.2), (1.3), (A1-2) and (A1-9), we obtain

$$\int_{S_e}T(Q)\frac{\partial F(Q,P_0)}{\partial h_Q}dS_{eQ}=-\int_{S_e}\delta g(Q)F(Q,P_0)dS_{eQ} \tag{2.4}$$

$$\int_{S_e}\frac{\partial F(Q,P_0)}{\partial h_Q}dS_{eQ}=-\int_{S_e}\frac{1}{R}[1+e^2(\frac{1}{6}-\frac{1}{2}\cos^2\theta_Q)+O(e^4)]F(Q,P_0)dS_{eQ} \tag{2.5}$$

For a given P on S_e, we obtain from (2.4) and (2.5) that

$$\begin{aligned}&\int_{S_e}[T(Q)-T(P)]\frac{\partial F(Q,P_0)}{\partial h_Q}dS_{eQ}\\&\quad=\int_{S_e}\{-\delta g(Q)+\frac{T(P)}{R}[1+e^2(\frac{1}{6}-\frac{1}{2}\cos^2\theta_Q)+O(e^4)]\}F(Q,P_0)dS_{eQ}\end{aligned} \tag{2.6}$$

According to (2.1) and the properties of the single-layer potential, we obtain by letting $P_0\to P$ in (2.6) and neglecting the quantities of the order of $O(e^4)$ that

$$\begin{aligned}\int_{S_e}[T(Q)-T(P)]M(Q,P)dS_{eQ}&=4\pi\{-\delta g(P)+\frac{T(P)}{R}[1+e^2(\frac{1}{6}-\frac{1}{2}\cos^2\theta_P)]\}\cos(r_P,h_P)\\&\quad+\int_{S_e}\{-\delta g(Q)+\frac{T(P)}{R}[1+e^2(\frac{1}{6}-\frac{1}{2}\cos^2\theta_Q)]\}F(Q,P)dS_{eQ}\end{aligned} \tag{2.7}$$

where

$$F(Q,P)\equiv\frac{r_Q^2-r_P^2}{r_Pl_{QP}^3} \tag{2.8}$$

and

$$M(Q,P) \equiv \frac{\partial}{\partial h_Q} F(Q,P) = \frac{2r_Q}{r_P l_{QP}^3} \frac{\partial r_Q}{\partial h_Q} - \frac{3(r_Q^2 - r_P^2)}{r_P l_{QP}^4} \frac{\partial l_{QP}}{\partial h_Q} \tag{2.9}$$

The kernel functions M(Q,P) and F(Q,P) are singular when Q→P. Their singularities for Q→P will be discussed in the Section 3.1.

The equation (2.7) is the integral equation from which the inverse Hotine formula and its ellipsoidal correction will be obtained.

2.2. Inverse Hotine' formula and its ellipsoidal correction

Denoting the projection of the surface element dS_{eQ} onto the unit sphere σ by $d\sigma_Q$, we have

$$dS_e = r_Q^2 \sec\beta_Q d\sigma_Q \tag{2.10}$$

where β_Q is the angle between the radius vector of Q and the surface normal of the surface S_e at point Q. With R the mean radius of the ellipsoid ($R = \sqrt[3]{a^2 b}$) and e the first eccentricity of the reference ellipsoid, and θ_P and θ_Q respectively the geocentric co-latitudes of the points P and Q on S_e, we have

$$r_P = R[1 + \frac{1}{2}e^2(\sin^2\theta_P - \frac{2}{3}) + O(e^4)] \tag{2.11a}$$

$$r_Q = R[1 + \frac{1}{2}e^2(\sin^2\theta_Q - \frac{2}{3}) + O(e^4)] \tag{2.11b}$$

$$l_{QP} = 2R\sin\frac{\psi_{QP}}{2}[1 + \frac{1}{4}e^2(\sin^2\theta_Q + \sin^2\theta_P - \frac{4}{3}) + O(e^4)] \tag{2.11c}$$

$$r_Q^2 \sec\beta_Q = R^2[1 + e^2(\sin^2\theta_Q - \frac{2}{3}) + O(e^4)] \tag{2.11d}$$

Furthermore, from Molodensky et al. (1962), we have

$$\frac{\partial r_Q}{\partial h_Q} = \cos(r_Q, h_Q) = 1 + O(e^4) \tag{2.12a}$$

$$\frac{\partial l_{QP}}{\partial h_Q} = \sin\frac{\psi_{QP}}{2}[1 - \frac{1}{4}e^2(3\cos^2\theta_Q + \cos^2\theta_P - \frac{(\cos\theta_Q - \cos\theta_P)^2}{\sin^2\frac{\psi_{QP}}{2}}) + O(e^4)] \tag{2.12b}$$

$$-\frac{1}{\gamma_Q}\frac{\partial\gamma_Q}{\partial h_Q} = \frac{2}{R}[1 + e^2(\cos^2\theta_Q - \frac{1}{6}) + O(e^4)] \tag{2.12c}$$

It then follows from (2.8) and (2.9) that

$$F(Q,P) r_Q^2 \sec\beta_Q = f(\psi_{QP}, \theta_Q, \theta_P)[e^2 + O(e^4)] \tag{2.13}$$

$$M(Q,P) r_Q^2 \sec\beta_Q = \frac{M(\psi_{QP})}{R}[1 + e^2(\frac{1}{2}\cos^2\theta_P - \frac{1}{6}) + O(e^4)] \tag{2.14}$$

where

$$f(\psi_{QP},\theta_Q,\theta_P)=\frac{1}{8}\frac{\sin^2\theta_Q-\sin^2\theta_P}{\sin^3\frac{\psi_{QP}}{2}} \tag{2.15a}$$

$$M(\psi_{QP})=\frac{1}{4\sin^3\frac{\psi_{QP}}{2}} \tag{2.15b}$$

Let

$$\delta g(Q)=\delta g^0(Q)+\delta g^1(Q)e^2+O(e^4) \tag{2.16}$$

Inserting (2.10), (2.13), (2.14) and (2.16) into (2.7) and neglecting the quantities of order of $O(e^4)$, we obtain

$$\begin{aligned}\int_\sigma[T(Q)-T(P)]\frac{M(\psi_{QP})}{R}[1+e^2(\frac{1}{2}\cos^2\theta_P-\frac{1}{6})]d\sigma_Q\\ =4\pi\{-\delta g^0(P)+\frac{T(P)}{R}+e^2[-\delta g^1(P)+\frac{T(P)}{R}(\frac{1}{6}-\frac{\cos^2\theta_P}{2})]\}\\ -e^2\int_\sigma[\delta g^0(Q)-\frac{T(P)}{R}]f(\psi_{QP},\theta_Q,\theta_P)d\sigma_Q\end{aligned} \tag{2.17}$$

Noting (A2-4) in the Appendix and (2.15), it follows that

$$\delta g^0(P)=\frac{T(P)}{R}-\frac{1}{4\pi R}\int_\sigma[T(Q)-T(P)]M(\psi_{QP})d\sigma_Q \tag{2.18}$$

$$\delta g^1(P)=\delta g_1^1(P)+\delta g_2^1(P) \tag{2.19}$$

where

$$\delta g_1^1(P)=(\frac{\cos^2\theta_P}{2}-\frac{1}{6})\delta g^0(P) \tag{2.19a}$$

$$\delta g_2^1(P)=-\frac{1}{4\pi}\int_\sigma\delta g^0(Q)f(\psi_{QP},\theta_Q,\theta_P)d\sigma_Q \tag{2.19b}$$

The formula (2.18) is the inverse Hotine formula, from which the spherical gravity disturbance is computed, and (2.19) is the ellipsoidal correction for the inverse Hotine formula.

2.2. Inverse Stokes' formula and its ellipsoidal correction

According to (1.4) and noting (2.12c), (2.18) and (2.19), we have that

$$\begin{aligned}\Delta g(P)&=\delta g(P)+(\frac{1}{\gamma_P}\frac{\partial\gamma_P}{\partial h_P})T(P)\\ &=\delta g^0(P)-\frac{2T(P)}{R}+e^2[\delta g^1(P)-\frac{2T(P)}{R}(\cos^2\theta_P-\frac{1}{6})]+O(e^4)\end{aligned} \tag{2.20}$$

Let

$$\Delta g(P)=\Delta g^0(P)+\Delta g^1(P)e^2+O(e^4) \tag{2.21}$$

then

$$\begin{aligned}\Delta g^0(P) &= \delta g^0(P) - \frac{2T(P)}{R} \\ &= -\frac{T(P)}{R} - \frac{1}{4\pi R}\int_\sigma [T(Q) - T(P)]M(\psi_{QP})d\sigma_Q\end{aligned} \tag{2.22}$$

$$\begin{aligned}\Delta g^1(P) &= \delta g^1(P) - \frac{2T(P)}{R}(\cos^2\theta_P - \frac{1}{6}) \\ &= (\frac{\cos^2\theta_P}{2} - \frac{1}{6})\delta g^0(P) - \frac{1}{4\pi}\int_\sigma \delta g^0(Q) f(\psi_{QP},\theta_Q,\theta_P)d\sigma_Q - \frac{2T(P)}{R}(\cos^2\theta_P - \frac{1}{6}) \\ &= \Delta g_1^1(P) + \Delta g_2^1(P)\end{aligned} \tag{2.23}$$

where

$$\Delta g_1^1(P) = -\frac{T(P)\cos^2\theta_P}{R} + (\frac{\cos^2\theta_P}{2} - \frac{1}{6})\Delta g^0(P) \tag{2.23a}$$

$$\Delta g_2^1(P) = -\frac{1}{2\pi R}\int_\sigma T(Q) f(\psi_{QP},\theta_Q,\theta_P)d\sigma_Q - \frac{1}{4\pi}\int_\sigma \Delta g^0(Q) f(\psi_{QP},\theta_Q,\theta_P)d\sigma_Q \tag{2.23b}$$

The formula (2.22) is the inverse Stokes formula, from which the spherical gravity anomaly is computed, and (2.23) is the ellipsoidal correction for the inverse Stokes formula.

3. Practical considerations for the integrals in the formulas

In the above section, we obtained the closed formulas (2.19) and (2.23) of the ellipsoidal corrections δg^1 and Δg^1 respectively to the inverse Hotine formula (2.18) (the spherical gravity disturbance δg^0) and the inverse Stokes formula (2.22) (the spherical gravity anomaly Δg^0) from the basic integral equation (2.7). The formula (2.19) (formula (2.23)) is expressed as a sum of a simple analytical function and an integral about δg^0 (Δg^0 and T). Obviously, the first part of δg^1 (Δg^1) is easy to compute from δg^0 (Δg^0 and T). In the following, we will discuss the integral parts (2.19b) and (2.23b).

3.1. Singularities

The integrals in the formulas (2.7), (2.18), (2.19b), (2.22) and (2.23b) are singular because their kernel functions M(Q, P), F(Q, P) and $M(\psi_{QP})$, $f(\psi_{QP},\theta_Q,\theta_P)$ are singular when Q→P or $\psi_{QP} \to 0$.

The singularity of the integral in the inverse Stokes (or Hotine) formula (2.22) (or (2.18)) has been discussed in many references such as Heiskanen and Moritz (1967) and Zhang (1993). Here we discuss the singularities of the integrals in (2.7), (2.19b) and (2.23b).

According to (2.13), we know that the integral in the left side of (2.7) and the integrals in the inverse Stokes formula (2.18) and the inverse Stokes formula (2.22) have the same form. So the integral in the left side of (2.7) can be treated with the same method used in processing the inverse Stokes (Hotine) formula.

Similarly according to (2.14), the integral in the left side of (2.7) and the integrals in (2.19b) and (2.23b) have the same form. So in the following, we only discuss the method to treat the singularity of the integral in (2.19b).

Obviously, we only need to consider the integral in the innermost spherical cap area σ_0 with the center at the computation point P and the radius ψ_0, which is so small that the spherical cap area can be treated as a plane. That is we discuss the following integral

$$\overline{\delta g}(P)=\frac{1}{4\pi}\int_{\sigma_0}\delta g^0(Q)f(\psi_{QP},\theta_Q,\theta_P)d\sigma_Q \tag{3.1}$$

From (A2-2) in the Appendix and (2.15a), and noting that σ is a unit sphere, we have

$$\begin{aligned}\overline{\delta g}(P)=\frac{1}{4\pi}\int_{\psi_{QP}=0}^{\psi_0}\int_{\alpha_{QP}=0}^{2\pi}\delta g^0(Q)\frac{1}{8\sin^3\frac{\psi_{QP}}{2}}\{4\sin^2\frac{\psi_{QP}}{2}\cos^2\frac{\psi_{QP}}{2}[\cos^2\theta_P-\sin^2\theta_P\cos^2\alpha_{QP}]\\ -2\cos\theta_P\cos\psi_{QP}\sin\theta_P\sin\psi_{QP}\cos\alpha_{QP}\}\sin\psi_{QP}d\alpha_{QP}d\psi_{QP}\\ =\frac{1}{4\pi}\int_{l_{QP}=0}^{l_0}\int_{\alpha_{QP}=0}^{2\pi}\delta g^0(Q)\{(1-\frac{l_{QP}^2}{4})[\cos^2\theta_P-\sin^2\theta_P\cos^2\alpha_{QP}]\\ -\frac{1}{l_{QP}}(1-\frac{l_{QP}^2}{2})(1-\frac{l_{QP}^2}{4})^{\frac{1}{2}}\sin 2\theta_P\cos\alpha_{QP}\}d\alpha_{QP}dl_{QP}\end{aligned} \tag{3.2}$$

where

$$l_0=2\sin\frac{\psi_0}{2}. \tag{3.3}$$

For Q in σ_0, we expand $\delta g^0(Q)$ into a Taylor series at the computation point P:

$$\delta g^0(Q)=\delta g^0(P)+x\delta g_x^0(P)+y\delta g_y^0(P)+\cdots \tag{3.4}$$

where the rectangular coordinates x, y are defined by

$$x=l_{QP}\cos\alpha_{QP};\quad y=l_{QP}\sin\alpha_{QP} \tag{3.5}$$

so that the x-axis points North, and

$$\delta g_x^0(P)=\frac{\partial\delta g^0}{\partial x}(P);\ \delta g_y^0(P)=\frac{\partial\delta g^0}{\partial y}(P) \tag{3.6}$$

The Taylor series (3.4) may also be written as

$$\delta g^0(Q)=\delta g^0(P)+[\delta g_x^0(P)\cos\alpha_{QP}+\delta g_y^0(P)\sin\alpha_{QP}]l_{QP}+\cdots \tag{3.7}$$

Inserting this into (3.2), performing the integral with respect to α_{QP} first, noting (A2-3) in the Appendix and neglecting the quantities of $O(l_0^2)$, we have

$$\overline{\delta g}(P)=\frac{l_0}{4}[\delta g^0(P)(3\cos^2\theta_P-1)+\delta g_x^0(P)] \tag{3.8}$$

We see that the effect of the innermost spherical cap area on the integral (2.19b) depends, to a first approximation, on $\delta g^0(P)$ and $\delta g_x^0(P)$. The value of $\delta g_x^0(P)$ can be obtained from the map of δg^0. It is the inclination of North-South profile through P.

3.2. Input data

In (2.19b) and (2.23b), the input data are respectively δg^0, and Δg^0 and T. These data are available only in some ocean areas. Here we give a little modification on the input data.

According to (2.16) and (2.21), we have

$$\delta g^0(Q) = \delta g(Q) - \delta g^1(Q)e^2 + O(e^4) \tag{3.9}$$

$$\Delta g^0(Q) = \Delta g(Q) - \Delta g^1(Q)e^2 + O(e^4) \tag{3.10}$$

In addition, the disturbing potential T(P) on the reference ellipsoid can be expressed as

$$T(P) = T^0(P) + e^2 T^1(P) \tag{3.11}$$

where $T^0(P)$ is the spherical approximation of T(P). Since δg^1 should be multiplied by e^2 before it is added to δg^0, we obtain by inserting respectively above formulas into the integrals in (2.19) and (2.22) and neglecting the quantities of order of $O(e^2)$ that

$$\delta g_2^1(P) = \frac{1}{4\pi}\int_\sigma \delta g(Q) f(\psi_{QP}, \theta_Q, \theta_P) d\sigma_Q \tag{3.12}$$

$$\Delta g_2^1(P) = \frac{1}{4\pi}\int_\sigma [\Delta g(Q) + \frac{2T^0(Q)}{R}] f(\psi_{QP}, \theta_Q, \theta_P) d\sigma_Q \tag{3.13}$$

where δg is the gravity disturbance which can be computed approximately from the global geopotential models, Δg and T^0 are respectively the gravity anomaly and the spherical disturbing potential which are already available globally with the resolutions of less than 1 degree and the accuracy of a few metres and locally with higher resolutions and higher accuracy.

3.3. Spherical harmonic expansions of the integrals

In the following, we will expand $\delta g_2^1(P)$ and $\Delta g_2^1(P)$ into series of spherical harmonics so that they can be computed from the global geopotential models.

According to Chapter 2-14 of Heiskanen and Moritz (1967), under the spherical approximation, we have

$$\delta g(\theta, \lambda) = \frac{1}{R}\sum_{n=2}^{\infty}(n+1)T_n(\theta, \lambda) \tag{3.14}$$

where $T_n(\theta, \lambda)$ is Laplace's surface harmonics of the disturbing potential T:

$$T_n(\theta,\lambda)=\sum_{m=0}^{n}[c_{nm}R_{nm}(\theta,\lambda)+d_{nm}S_{nm}(\theta,\lambda)] \tag{3.15}$$

Let

$$\delta g(\theta,\lambda)\cos^2\theta=\frac{1}{R}\sum_{n=2}^{\infty}(n+1)X_n(\theta,\lambda) \tag{3.16}$$

From (2.22) and the definitions of $\delta g_2^1(P)$ and $\Delta g_2^1(P)$, we know that these two integrals are equal. So according to (1-102) of Heiskanen and Moritz (1967), we have from (3.12) that

$$\Delta g_2^1(P)=\delta g_2^1(P)=\frac{1}{4\pi}\int_\sigma\frac{\delta g(Q)\cos^2\theta_Q-\delta g(P)\cos^2\theta_P}{8\sin^3\frac{\psi_{QP}}{2}}d\sigma_Q-\frac{\cos^2\theta_P}{4\pi}\int_\sigma\frac{\delta g(Q)-\delta g(P)}{8\sin^3\frac{\psi_{QP}}{2}}d\sigma_Q$$

$$=\frac{1}{2R}[-\sum_{n=2}^{\infty}n(n+1)X_n(\theta_P,\lambda_P)+\sum_{n=2}^{\infty}n(n+1)T_n(\theta_P,\lambda_P)\cos^2\theta_P] \tag{3.17}$$

According to (A3-4) and (A3-5),

$$X_n(\theta,\lambda)=\sum_{m=0}^{n}[\delta E_{nm}R_{nm}(\theta,\lambda)+\delta F_{nm}S_{nm}(\theta,\lambda)] \tag{3.18}$$

$$\sum_{n=2}^{\infty}n(n+1)T_n(\theta,\lambda)\cos^2\theta=\sum_{n=2}^{\infty}n(n+1)\sum_{m=0}^{n}[\delta G_{nm}R_{nm}(\theta,\lambda)+\delta H_{nm}S_{nm}(\theta,\lambda)] \tag{3.19}$$

where $\{\delta E_{nm},\delta F_{nm}\}$ and $\{\delta G_{nm},\delta H_{nm}\}$ are defined as follows

$$\begin{Bmatrix}\delta E_{nm}\\ \delta F_{nm}\end{Bmatrix}=\frac{n-1}{n+1}\alpha_{n-2}^m\begin{Bmatrix}c_{n-2m}\\ d_{n-2m}\end{Bmatrix}+\beta_n^m\begin{Bmatrix}c_{nm}\\ d_{nm}\end{Bmatrix}+\frac{n+3}{n+1}\gamma_{n+2}^m\begin{Bmatrix}c_{n+2m}\\ d_{n+2m}\end{Bmatrix} \tag{3.20}$$

$$\begin{Bmatrix}\delta G_{nm}\\ \delta H_{nm}\end{Bmatrix}=\frac{(n-2)(n-1)}{n(n+1)}\alpha_{n-2}^m\begin{Bmatrix}c_{n-2m}\\ d_{n-2m}\end{Bmatrix}+\beta_n^m\begin{Bmatrix}c_{nm}\\ d_{nm}\end{Bmatrix}+\frac{(n+2)(n+3)}{n(n+1)}\gamma_{n+2}^m\begin{Bmatrix}c_{n+2m}\\ d_{n+2m}\end{Bmatrix} \tag{3.21}$$

with

$$\alpha_n^m=\frac{(n-m+1)(n-m+2)}{(2n+1)(2n+3)} \tag{3.22}$$

$$\beta_n^m=\frac{2n^2-2m^2+2n-1}{(2n+3)(2n-1)} \tag{3.23}$$

$$\gamma_n^m=\frac{(n+m)(n+m-1)}{(2n+1)(2n-1)} \tag{3.24}$$

So we obtain from (3.17) that

$$\Delta g_2^1(P)=\delta g_2^1(P)=\frac{1}{R}\sum_{n=2}^{\infty}\sum_{m=0}^{n}[\delta A_{nm}R_{nm}(\theta_P,\lambda_P)+\delta B_{nm}S_{nm}(\theta_P,\lambda_P)] \tag{3.25}$$

where

$$\begin{Bmatrix} \delta A_{nm} \\ \delta B_{nm} \end{Bmatrix} = -\frac{(n-1)(n-m-1)(n-m)}{(2n-1)(2n-3)} \begin{Bmatrix} c_{n-2m} \\ d_{n-2m} \end{Bmatrix} + \frac{(n+3)(n+m+1)(n+m+2)}{(2n+5)(2n+3)} \begin{Bmatrix} c_{n+2m} \\ d_{n+2m} \end{Bmatrix} \quad (3.26)$$

Thus we express $\delta g_2^1(P)$ and $\Delta g_2^1(P)$ by a series of spherical harmonics. The input data $\{c_{nm}, d_{nm}\}$ are the spherical harmonic coefficients of the disturbing potential.

5. Conclusions

This paper gives the ellipsoidal corrections $\delta g^1(P)$ and $\Delta g^1(P)$ for the inverse Hotine formula, the spherical gravity disturbance $\delta g^0(P)$, and the inverse Stokes formula, the spherical gravity anomaly $\Delta g^0(P)$, respectively.

- By adding the ellipsoidal corrections to their spherical solutions, the error of the gravity disturbance and the gravity anomaly decrease from $O(e^2)$ to $O(e^4)$, which is sufficient for most practical purposes.

- $\delta g^1(P)$ is expressed as a sum of a simple analytical function of $\delta g^0(P)$ and an integral in terms of δg^0. In the practical computation of the integral, the input data δg^0 can be substituted by the gravity disturbance δg, which can be approximately computed from the global geopotential models. The integral part of $\delta g^1(P)$ can also be computed directly from the global geopotential models via the formula (3.25).

- $\Delta g^1(P)$ is expressed as a sum of a simple analytical function of $\Delta g^0(P)$ and T(P) and an integral in terms of Δg^0 and T. In the practical evaluation of the integral, the input data Δg^0 and T(P) can be substituted respectively by the gravity anomaly Δg and the spherical disturbing potential T^0, which are already available globally with resolutions better than 1 degree and accuracy within a few metres, and locally with higher resolutions and higher accuracy. The integral part of $\Delta g^1(P)$ can also be computed directly from the global geopotential models via the formula (3.25).

- Comparing to the ellipsoidal correction to gravity anomaly given in Wang (1999), $\Delta g^1(P)$ is simpler not only in the formulation but also in the auxiliary data Δg^0 (or Δg), in comparison to the auxiliary data χ used in Wang (1999).

References

Heiskanen, W.A. and H. Moritz (1967), Physical Geodesy, W.H. Freeman and Co., San Francisco and London.

Hwang C , Parsons B (1995) Gravity anomalies derived from Seasat, Geosat, ERS-1 and TOPEX/POSEIDON altimetry and ship gravity: a case-study over the Reykjanes Ridge. Geophys J Int 122: 551-568.

Kim JH (1996) Improved recovery of gravity anomalies from dense altimeter data. Report N. 444, Department of Geodetic Science and Surveying, OSU.

Molodensky MS, Eremeev VF, Yurkina MI (1962) Methods for study of the external gravitation field and figure of the Earth. Transl. from Russian (1960), Jerusalem, Israel Program for Scientific Translations

Moritz, H. (1980), Advanced Physical Geodesy, Herbert Wichmann Verlag Karlsruhe, Abacus Press, Tunbridge Wells Kent.
Olgiati A, Balmino G, Sarrilh M, Green CM (1995) gravity anomalies from satellite altimetry: Comparisons between computation via geoid heights and vis deflections of the vertical. Bull Geod 69: 252-260.
Rummel R, Haagmans R (1990) Gravity gradients from satellite altimetry. Marine Geodesy 14: 1-12.
Sandwell DT, Smith WHF (1997) Marine gravity anomaly from Geosat and ERS-1 satellite altimetry. J Geophys Res 102: 10039-10054
Wang YM (1999) On the ellipsoidal corrections to gravity anomalies computed using the inverse Stokes integral. Journal of Geodesy 73(1): 29-34
Zhang C (1993) Recovery of gravity information from satellite altimetry data and associated forward geopotential models. UCGE Rep 20058, The University of Calgary, Calgary, Canada.
Zhang C, Blais JAR (1995) Comparison of methods for marine gravity determination from satellite altimetry data in the Labrador Sea. Bull Geod 69: 173

Appendix

1. Denote respectively the ellipsoidal coordinates and the spherical coordinates of a point P by (u_P, β_p, λ_P) and (r_P, θ_P, λ_P). According to Chapter 1-20 of Heiskanen and Moritz (1967), we know that

$$V_a(P) \equiv \frac{Q_0(i\frac{u_P}{E})}{Q_0(i\frac{b}{E})} \tag{A1-1}$$

is harmonic and regular outside S_e and continuously differentiable on and outside S_e, and for Q on S_e,

$$V_a(Q)=1 \tag{A1-2}$$

From Chapter 2-7, 2-8 and 2-9 of Heiskanen and Moritz (1967), we have

$$Q_0(i\frac{u_P}{E}) = -i\tan^{-1}\frac{E}{u_P} \tag{A1-3}$$

$$\frac{\partial V_a}{\partial h_P}(P) = -\sqrt{\frac{u_P^2+E^2}{u_P^2+E^2\sin^2\beta_P}}\frac{\partial V_a}{\partial u_P}(P) \tag{A1-4}$$

and

$$\tan^{-1}\frac{E}{b} = \frac{E}{b}[1-\frac{1}{3}(\frac{E}{b})^2+O((\frac{E}{b})^4)] = e'[1-\frac{1}{3}e'^2+O(e'^4)] \tag{A1-5}$$

So for Q on S_e,

$$\frac{\partial V_a}{\partial h_Q}(Q) = -\sqrt{\frac{b^2+E^2}{b^2+E^2\sin^2\beta_Q}}\frac{E}{\tan^{-1}\frac{E}{b}\sqrt{b^2+E^2}} = -\frac{1}{a\sqrt{1+e'^2\sin^2\beta_Q}}\frac{1}{1-\frac{1}{3}e'^2+O(e'^4)}$$

$$= -\frac{1}{a}[1+e'^2(\frac{1}{3}-\frac{1}{2}\sin^2\beta_Q)+O(e'^4)] \tag{A1-6}$$

Since ,

$$a = R(1+\frac{1}{6}e^2);\ e'^2 = e^2 + O(e^4) \tag{A1-7}$$

and

$$\sin^2\beta_Q = \frac{\tan^2\beta_Q}{1+\tan^2\beta_Q} = \frac{a^2\cos^2\theta_Q}{b^2\sin^2\theta_Q + a^2\cos^2\theta_Q} = \cos^2\theta_Q[1+e^2\sin^2\theta_Q + O(e^4)]$$
$$= \cos^2\theta_Q + O(e^2) \tag{A1-8}$$

Equation (A-6) can be rewritten as

$$\frac{\partial V_a}{\partial h_Q}(Q) = -\frac{1}{R}[1+e^2(\frac{1}{6}-\frac{1}{2}\cos^2\theta_Q)+O(e^4)] \tag{A1-9}$$

2. From the spherical triangle of Figure 1, we have

$$\cos\theta_Q = \cos\psi_{QP}\cos\theta_P + \sin\psi_{QP}\sin\theta_P\cos\alpha_{QP} \tag{A2-1}$$

and hence

$$\sin^2\theta_Q - \sin^2\theta_P = \cos^2\theta_P - \cos^2\theta_P\cos^2\psi_{QP} - \sin^2\theta_P\sin^2\psi_{QP}\cos^2\alpha_{QP}$$
$$-2\cos\theta_P\cos\psi_{QP}\sin\theta_P\sin\psi_{QP}\cos\alpha_{QP}$$
$$= 4\sin^2\frac{\psi_{QP}}{2}\cos^2\frac{\psi_{QP}}{2}[\cos^2\theta_P - \sin^2\theta_P\cos^2\alpha_{QP}]$$
$$-2\cos\theta_P\cos\psi_{QP}\sin\theta_P\sin\psi_{QP}\cos\alpha_{QP} \tag{A2-2}$$

Noting that

$$\int_0^{2\pi} d\alpha_{QP} = 2\pi;\ \int_0^{2\pi}\cos\alpha_{QP}d\alpha_{QP} = 0;\ \int_0^{2\pi}\cos^2\alpha_{QP}d\alpha_{QP} = \pi \tag{A2-3}$$

we obtain that

$$\int_\sigma \frac{\sin^2\theta_Q - \sin^2\theta_P}{\sin^3\frac{\psi_{QP}}{2}}d\sigma = \int_0^\pi\int_0^{2\pi}\frac{\sin^2\theta_Q - \sin^2\theta_P}{\sin^3\frac{\psi_{QP}}{2}}4\sin\frac{\psi_{QP}}{2}d\sin\frac{\psi_{QP}}{2}d\alpha_{QP}$$
$$= 16\pi(2\cos^2\theta_P - \sin^2\theta_P)\int_0^1(1-x^2)dx$$
$$= 16\pi(2\cos^2\theta_P - \frac{2}{3}) \tag{A2-4}$$

3. According to Chapter 2-14 of Heiskanen and Moritz (1967), under the spherical approximation, we have

$$\delta g(\theta,\lambda) = \frac{1}{R}\sum_{n=2}^{\infty}(n+1)T_n(\theta,\lambda) \tag{A3-1}$$

and

$$\Delta g(\theta,\lambda) = \frac{1}{R}\sum_{n=2}^{\infty}(n-1)T_n(\theta,\lambda) \tag{A3-2}$$

where $T_n(\theta, \lambda)$ is Laplace's surface harmonics of the disturbing potential T:

$$T_n(\theta,\lambda) = \sum_{m=0}^{n}[c_{nm}R_{nm}(\theta,\lambda) + d_{nm}S_{nm}(\theta,\lambda)] \tag{A3-3}$$

Let

$$\sum_{n=2}^{\infty}(n+1)T_n(\theta,\lambda)\cos^2\theta \equiv \sum_{n=2}^{\infty}(n+1)\sum_{m=0}^{n}[\delta E_{nm}R_{nm}(\theta,\lambda) + \delta F_{nm}S_{nm}(\theta,\lambda)] \tag{A3-4}$$

$$\sum_{n=2}^{\infty}n(n+1)T_n(\theta,\lambda)\cos^2\theta \equiv \sum_{n=2}^{\infty}n(n+1)\sum_{m=0}^{n}[\delta G_{nm}R_{nm}(\theta,\lambda) + \delta H_{nm}S_{nm}(\theta,\lambda)] \tag{A3-5}$$

From (A11) of Wang (1999) (Note: there is a printing error in that formula) and (A3-3), we know that

$$\begin{aligned} T_n(\theta,\lambda)\cos^2\theta = \sum_{m=0}^{n}\{&c_{nm}[\alpha_n^m R_{n+2m}(\theta,\lambda) + \beta_n^m R_{nm}(\theta,\lambda) + \gamma_n^m R_{n-2m}(\theta,\lambda)] \\ &+ d_{nm}[\alpha_n^m S_{n+2m}(\theta,\lambda) + \beta_n^m S_{nm}(\theta,\lambda) + \gamma_n^m S_{n-2m}(\theta,\lambda)]\} \end{aligned} \tag{A3-6}$$

where

$$\alpha_n^m = \frac{(n-m+1)(n-m+2)}{(2n+1)(2n+3)} \tag{A3-7}$$

$$\beta_n^m = \frac{2n^2 - 2m^2 + 2n - 1}{(2n+1)(2n-1)} \tag{A3-8}$$

$$\gamma_n^m = \frac{(n+m)(n+m-1)}{(2n+1)(2n-1)} \tag{A3-9}$$

So

$$\begin{aligned} \begin{Bmatrix}\delta E_{nm}\\ \delta F_{nm}\end{Bmatrix} &= \sum_{k=2}^{\infty}\frac{k+1}{n+1}N_n^m\int_{\sigma_0}T_k(\theta,\lambda)\cos^2\theta\begin{Bmatrix}R_{nm}(\theta,\lambda)\\ S_{nm}(\theta,\lambda)\end{Bmatrix}d\sigma_0 \\ &= \frac{n-1}{n+1}\alpha_{n-2}^m\begin{Bmatrix}c_{n-2m}\\ d_{n-2m}\end{Bmatrix} + \beta_n^m\begin{Bmatrix}c_{nm}\\ d_{nm}\end{Bmatrix} + \frac{n+3}{n+1}\gamma_{n+2}^m\begin{Bmatrix}c_{n+2m}\\ d_{n+2m}\end{Bmatrix} \end{aligned} \tag{A3-10}$$

$$\begin{aligned} \begin{Bmatrix}\delta G_{nm}\\ \delta H_{nm}\end{Bmatrix} &= \sum_{k=2}^{\infty}\frac{k(k+1)}{n(n+1)}N_n^m\int_{\sigma_0}T_k(\theta,\lambda)\cos^2\theta\begin{Bmatrix}R_{nm}(\theta,\lambda)\\ S_{nm}(\theta,\lambda)\end{Bmatrix}d\sigma_0 \\ &= \frac{(n-2)(n-1)}{n(n+1)}\alpha_{n-2}^m\begin{Bmatrix}c_{n-2m}\\ d_{n-2m}\end{Bmatrix} + \beta_n^m\begin{Bmatrix}c_{nm}\\ d_{nm}\end{Bmatrix} + \frac{(n+2)(n+3)}{n(n+1)}\gamma_{n+2}^m\begin{Bmatrix}c_{n+2m}\\ d_{n+2m}\end{Bmatrix} \end{aligned} \tag{A3-11}$$

Similarly letting

$$\sum_{n=2}^{\infty}(n-1)T_n(\theta,\lambda)\cos^2\theta \equiv \sum_{n=2}^{\infty}(n-1)\sum_{m=0}^{n}[\Delta E_{nm}R_{nm}(\theta,\lambda) + \Delta F_{nm}S_{nm}(\theta,\lambda)] \tag{A3-12}$$

$$\sum_{n=2}^{\infty}n(n-1)T_n(\theta,\lambda)\cos^2\theta \equiv \sum_{n=2}^{\infty}n(n-1)\sum_{m=0}^{n}[\Delta G_{nm}R_{nm}(\theta,\lambda) + \Delta H_{nm}S_{nm}(\theta,\lambda)] \tag{A3-13}$$

then

$$\begin{Bmatrix}\Delta E_{nm}\\ \Delta F_{nm}\end{Bmatrix}=\frac{n-3}{n-1}\alpha_{n-2}^{m}\begin{Bmatrix}c_{n-2m}\\ d_{n-2m}\end{Bmatrix}+\beta_{n}^{m}\begin{Bmatrix}c_{nm}\\ d_{nm}\end{Bmatrix}+\frac{n+1}{n-1}\gamma_{n+2}^{m}\begin{Bmatrix}c_{n+2m}\\ d_{n+2m}\end{Bmatrix} \quad \text{(A3-14)}$$

$$\begin{Bmatrix}\Delta G_{nm}\\ \Delta H_{nm}\end{Bmatrix}=\frac{(n-2)(n-3)}{n(n-1)}\alpha_{n-2}^{m}\begin{Bmatrix}c_{n-2m}\\ d_{n-2m}\end{Bmatrix}+\beta_{n}^{m}\begin{Bmatrix}c_{nm}\\ d_{nm}\end{Bmatrix}+\frac{(n+2)(n+1)}{n(n-1)}\gamma_{n+2}^{m}\begin{Bmatrix}c_{n+2m}\\ d_{n+2m}\end{Bmatrix} \quad \text{(A3-15)}$$

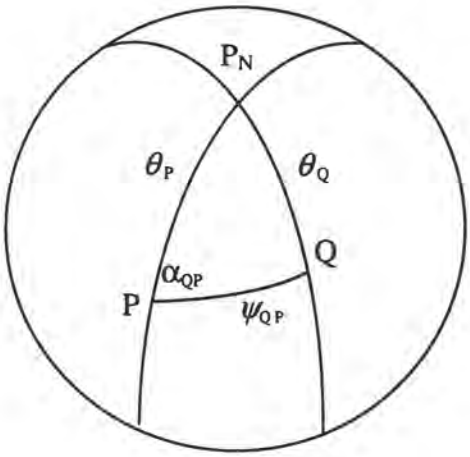


Figure 1. Spherical triangle

Triple frequency GPS for precise positioning

Lars E. Sjöberg

Abstract

In this paper we dwell upon the possibility to determine the GPS phase ambiguities from double difference triple frequency GPS phase and code data more or less instantaneously. We take advantage of the well known fact, that the widelane ambiguity is easily fixed from such observables. It is shown that this holds also for a third signal of carrier wavelength (λ_3) in the range 14.3 cm $\leq \lambda_3 \leq$ 30.0 cm. At the limits λ_3 = 14.3 cm (2100.6 MHz) and λ_3 = 30.0 cm (997.1 MHz) the base ambiguities are easily fixed as soon as the widelane ambiguities have been determined. For other choices of λ_3 the method is less optimal.

Recently the US Department of Defence announced that GPS satellites launched after December 2004 will be equipped with three civil GPS signals, where L_1 and L_2 are the same as today's signals, and the third signal will operate in the frequency 1176.45 MHz (λ_3=25.44 cm). This design will allow rapid precise position over long baselines with significant ionospher biases.

Key-words: Ambiguity, GPS, triple frequency.

1 Introduction

Precise positioning with the Global Positioning System is related with reliable fixing of the signal phase ambiguities. The success in real time positioning is dependent on the fast ambiguity fixing. Various methods have been developed for fast and reliable ambiguity estimation over short baselines, but for long baselines the needed time to fix ambiguities increases drastically due to the influence of various systematic effects, in particular the ionosphere bias. However, one well-known exception is the widelane ambiguity, which can be quickly determined also for long baselines from a linear combination of phase and code data. Sjöberg (1996) and (1997, 1998, 1999) took advantage of this fact to solve for the base ambiguities. It turned out (Sjöberg,1998, 1999 and Almgren, 1998) that this method works very well for short baselines, but mainly the ionosphere effect restricts its success for long baselines.

Recently the US Department of Defence (DoD) has announced, that it plans to introduce a second civil frequency identical with the current L_2 frequency, and later, after December 2004, a third civil frequency at 1176.45 MHz (λ_3=25.44 cm) is planned to operate on all new GPS satellites.

The goal of this paper is to take advantage of all three signals for fast phase ambiguity resolution for short as well as long baselines. First we speculate on the optimum choice of the third frequency with respect to accurate and reliable ambiguity resolution. Second, we compare the result with the proposed frequency.

2 Ambiguity estimation for dual frequency data

Consider the following dual frequency phase and code observation equations for receiver-to-satellite ranges (Sjöberg 1996, 1997, 1998)

$$\left.\begin{aligned} \tilde{l}_1 &= \tilde{\phi}_1\lambda_1 = u + \lambda_1 N_1 - \tfrac{\mu}{f_1^2} + \epsilon_{11} \\ \tilde{l}_2 &= \tilde{\phi}_2\lambda_2 = u + \lambda_2 N_2 - \tfrac{\mu}{f_2^2} + \epsilon_{12} \\ \tilde{l}_3 &= \tilde{R}_1 = u + \tfrac{\mu}{f_1^2} + \epsilon_{21} \\ \tilde{l}_4 &= \tilde{R}_2 = u + \tfrac{\mu}{f_2^2} + \epsilon_{22}, \end{aligned}\right\} \tag{1}$$

where $\tilde{l}_1$ and $\tilde{l}_2$ (with phase $\tilde{\phi}_1$ and $\tilde{\phi}_2$) are the phase observables scaled by their wavelengths λ_1 and λ_2, and $\tilde{l}_3$ and $\tilde{l}_4$ are the code observables. ϵ_{11}, ϵ_{12}, ϵ_{21} and ϵ_{22} are random observation errors. The unknowns are $u = \rho + c\Delta\delta$, which is the sum of the satellite-to-receiver range (ρ) and the product of velocity of light (c) and receiver and satellite clock bias difference ($\Delta\delta$). Furthermore μ is the ionosphere bias and N_1 and N_2 are the phase ambiguities on L_1 and L_2, respectively, with the known frequencies f_1 and f_2. Usually we will consider these equations for double differenced data, i.e. for pairs of receivers and satellites. Obviously eqn. (1) contains 4 independent equations and 4 unknowns, and, at least in principle, it can be directly solved for N_1 and N_2. The problem is, however, that these estimates are too poor to be useful (Sjöberg ibid.). On the contrary the widelane ambiguity can be accurately determined by

$$\hat{N}_w = \hat{N}_1 - \hat{N}_2 = \frac{\tilde{l}_1}{\lambda_1} - \frac{\tilde{l}_2}{\lambda_2} - \frac{f_1 - f_2}{f_1 + f_2}\left(\frac{\tilde{l}_3}{\lambda_1} + \frac{\tilde{l}_4}{\lambda_2}\right) \tag{2}$$

Subsequently $\hat{N}_w$ is independent of baseline length, ionosphere bias (and the time variable satellite-to-receiver range), and in most cases it can quickly be fixed to its correct integer value.

Having fixed N_w we may form an observation equation for the base ambiguity N_1:

$$N_1 = N_w + \hat{N}_2 - \epsilon_2, \tag{3}$$

where $\hat{N}_2$ is the primary estimate of N_2 from eqn. (1), and ϵ_2 is its random error. Another equation of N_1 is given by

$$N_1 = \hat{N}_1 - \epsilon_1, \tag{4}$$

i.e. by its primary estimate $\hat{N}_1$ from eqn. (1) with error ϵ_1. The errors ϵ_1 and ϵ_2 are very significantly correlated. Denoting the covariance matrix between the above two equations Q the least squares solution for N_1 becomes

$$\hat{\hat{N}}_1 = \left(e^T Q^{-1} e\right)^{-1} e^T Q^{-1} \begin{pmatrix} \hat{N}_1 \\ \hat{N}_2 + N_w \end{pmatrix}, \tag{5}$$

where

$$e^T = (1, 1).$$

It turns out (Sjöberg 1996 and 1998) that the base ambiguity is well determined by this method for short baselines, where the ionosphere bias (μ) can be omitted from the model (1). This conclusion is confirmed by the numerical analyses of Almgren (1998). On the other hand the result is very pessimistic for long baselines, due to the fact that the ionosphere bias prevent the nice error reduction of the joint solution as was the case for short baselines.

The above solution (5) could also be obtained more directly from the original equations (1) after the substitution of the second equation by

$$\tilde{l}_2' = \tilde{l}_2 + \lambda_2 N_w = u + \lambda_2 N_1 - \frac{\mu}{f_2^2} + \epsilon_{12}. \tag{6}$$

This means that we have reduced the number of unknowns to three by using the known widelane ambiguity. The least squares solution of the revised eqn. (1) is the same as in (5).

3 Ambiguity estimation for triple frequency data

Assuming that there are three independent GPS signals L_1, L_2 and L_3 with carrier wavelengths λ_1, λ_2 and λ_3 and frequencies f_1, f_2 and f_3, we can form 6 independent observation equations similar to the system (1):

$$\left.\begin{array}{l} \tilde{l}_1 = u + \lambda_1 N_1 - \frac{\mu}{f_1^2} + \epsilon_{11} \\ \tilde{l}_2 = u + \lambda_2 N_2 - \frac{\mu}{f_2^2} + \epsilon_{12} \\ \tilde{l}_3 = u + \lambda_3 N_3 - \frac{\mu}{f_3^2} + \epsilon_{13} \\ \tilde{l}_4 = u + \frac{\mu}{f_1^2} + \epsilon_{21} \\ \tilde{l}_5 = u + \frac{\mu}{f_2^2} + \epsilon_{22} \\ \tilde{l}_6 = u + \frac{\mu}{f_3^2} + \epsilon_{23}. \end{array}\right\} \tag{7}$$

From these observables we may estimate the following widelane ambiguities

$$N_{w12} = N_1 - N_2 = \frac{\tilde{l}_1}{\lambda_1} - \frac{\tilde{l}_2}{\lambda_2} - \frac{f_1 - f_2}{f_1 + f_2}\left(\frac{\tilde{l}_4}{\lambda_1} + \frac{\tilde{l}_5}{\lambda_2}\right) \tag{8}$$

and

$$N_{w13} = N_1 - N_3 = \frac{\tilde{l}_1}{\lambda_1} - \frac{\tilde{l}_3}{\lambda_3} - \frac{f_1 - f_3}{f_1 + f_3}\left(\frac{\tilde{l}_4}{\lambda_1} + \frac{\tilde{l}_6}{\lambda_3}\right). \tag{9}$$

For f_3 chosen rather close to f_1 the small factors $(f_1 - f_2)/(f_1 + f_2)$ and $(f_1 - f_3)/(f_1 + f_3)$ efficiently reduce the noise of the code observables $\tilde{l}_4$, $\tilde{l}_5$ and $\tilde{l}_6$. This explains the resulting low noise in the widelane ambiguity.

Let us now insert eqs. (8) and (9) into the second and third equations of (7). This yields the revised system of equations (in matrix form and altered order).

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & \nu^2 & 0 \\ 1 & \alpha^2 & 0 \\ 1 & -1 & 1 \\ 1 & -\nu^2 & \nu \\ 1 & -\alpha^2 & \alpha \end{pmatrix} \begin{pmatrix} u \\ \mu/f_1^2 \\ \lambda_1 N_1 \end{pmatrix} = \begin{pmatrix} \tilde{l}_4 - \epsilon_{21} \\ \tilde{l}_5 - \epsilon_{22} \\ \tilde{l}_6 - \epsilon_{23} \\ \tilde{l}_1 - \epsilon_{11} \\ \tilde{l}_2 + \lambda_2 N_{w12} - \epsilon_{12} \\ \tilde{l}_3 + \lambda_3 N_{w13} - \epsilon_{13} \end{pmatrix}, \tag{10}$$

where

$$\nu = f_1/f_2 = \lambda_2/\lambda_1$$

and

$$\alpha = f_1/f_3 = \lambda_3/\lambda_1.$$

The system (10) is over determined with 3 redundancies. However, the first 3 observations, from the pseudoranges, have much lower accuracy than the last three ones of phase observables. (The ratio of the standard errors is of the order 100/1.) Subsequently, the code observables adds very little to the least squares solution of the system (10). Neglecting these equations we are left with the system

$$AX = L - \epsilon,$$

where

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & -\nu^2 & \nu \\ 1 & -\alpha^2 & \alpha \end{pmatrix} \qquad X = \begin{pmatrix} u \\ \mu/f_1^2 \\ \lambda_1 N_1 \end{pmatrix} \tag{11}$$

and $L - \epsilon$ is the vector of the last three equations of (10). The system (10) corresponds to the normal equations

$$A^T A X = A^T L \tag{12}$$

with the unique solution

$$\hat{X} = \left(A^T A\right)^{-1} A^T L \tag{13}$$

and the covariance matrix of $\hat{X}$

$$Q_{\hat{X}\hat{X}} = \sigma_0^2 \left(A^T A\right)^{-1}, \tag{14}$$

where σ_0^2 is the variance of unit weight. In this study we are particularly interested in the standard error of the estimated base ambiguity N_1, which is included in (14). More precisely it reads

$$\sigma_{\hat{N}_1} = \frac{\sigma_0}{\lambda_1} \left(A^T A\right)_{33}^{-\frac{1}{2}}. \tag{15}$$

¿From the matrix A given by (11) one easily obtains

$$\sigma_{\hat{N}_1} = \frac{\sigma_0}{\lambda_1} \sqrt{\frac{c}{d}}, \tag{16}$$

where

$$c = 2\left(1 + \nu^4 + \alpha^4 - \alpha^2 - \nu^2 - \alpha^2\nu^2\right)$$

and

$$\begin{aligned} d = {} & 3\left(1 + \nu^4 + \alpha^4\right)\left(1 + \alpha^2 + \nu^2\right) + 2\left(1 + \nu + \alpha\right)\left(1 + \alpha^2 + \nu^2\right)\left(1 + \alpha^3 + \nu^3\right) \\ & - \left(1 + \nu + \alpha\right)^2\left(1 + \alpha^4 + \nu^4\right) - 3\left(1 + \alpha^3 + \nu^3\right)^2 - \left(1 + \alpha^2 + \nu^2\right)^3. \end{aligned}$$

For modern geodetic GPS receivers the carrier phase observable noise (σ_0) can be set to 3 mm. The L_1 carrier wavelength (λ_1) is 19.0 cm and $\nu = f_1/f_2 = \lambda_2/\lambda_1 = 24.4/19.0 = 1.28421$. Also $f_1 = 1575.42$ MHz and $f_2 = 1227.60$ MHz. For these constants $\sigma_{\hat{N}_1}$ is given as a function of $\alpha (= f_1/f_3 = \lambda_3/\lambda_1)$ in Fig 1 (dashed curve). It shows that the standard error increases dramatically for $\alpha = 1 (f_1 = f_3)$ and $\alpha = \nu (f_2 = f_3)$. For, say, $\alpha < 0.8$ and $\alpha > 1.5\sigma_{\hat{N}_1}$ is well below unity. The figure shows also (solid line) the standard error of the widelane ambiguity determined from the L_1 and L_3 signals given by the (approximate) formula

$$\sigma_{\hat{N}_{w13}} = \frac{\sigma_R}{\lambda_1} \frac{|1 - \alpha|}{1 + \alpha} \left(1 + \frac{1}{\alpha}\right) \tag{17}$$

with α_R set to 30.0 cm. This curve has a minimum (0) for $\alpha = 1$. For α close to 0.75 and 1.58 both $\sigma_{\hat{N}_1}$ and $\sigma_{\hat{N}_{w13}}$ are close to 0.6 as presented in Table 1. Obviously, the optimum choice of α can be found for these values.

The L_3 signal proposed by DoD of 1146.45 MHz (λ_3=25.44 cm) to some extent fulfills the above demands. For this frequency α becomes 1.339, yielding the standard errors of N_{w13} and N_1 less than 0.5 and about 2.5, respectively. (Cf. Fig. 1.)

Table 1: Optimum choices of L_3 and corresponding standard errors of $\hat{N}_1$ and N_{w13}

α	λ_3(cm)	f_3	$\sigma_{\hat{N}_1}$	$\sigma_{\hat{N}_{w13}}$
0.75	14.3	2100.6 MHz	0.56	0.53
1.58	30.0	997.1 MHz	0.59	0.58

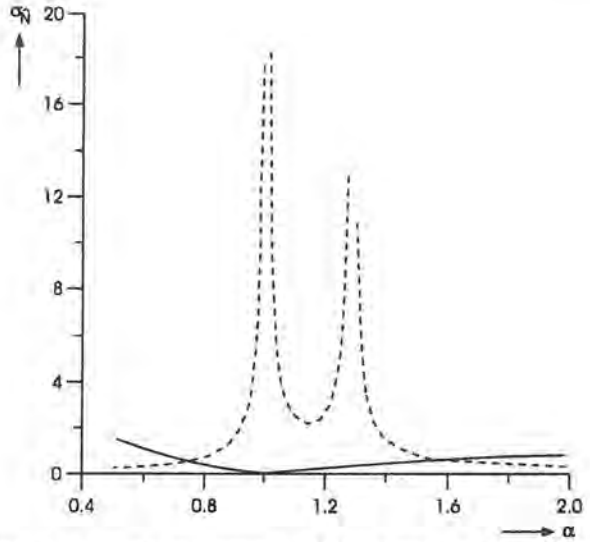


Figure 1: The figure shows the standard errors of the base ambiguity N_1 (dashed line) and the widelane ambiguity N_{w13} (solid line) as functions of α. Constants: $\sigma_R = 30.0$ cm, $\sigma_0 = 3$ mm, $\lambda_1 = 19.0$ cm.

4 Conclusions

We conclude that our method to resolve GPS phase ambiguities from double difference phase and code data with three GPS signals will be optimal for a third signal of frequency of about 997 MHz or 2101 MHz, at which frequencies both the widelane and base phase ambiguities are quickly resolved. Obviously this method solves the problem caused by the ionosphere bias for long baselines. The L_3 frequency (1146.5 MHz) proposed by DoD will be a good tool in solving this problem.

Acknowledgements. The numerical calculations displayed in Fig 1 by E. Asenjo are cordially acknowledged.

References

Almgren, K.: A new method for GPS phase ambiguity resolution on-the-fly, Division of Geodesy Report No. 1047, Royal Institute of Technology, Stockholm, 1998. (Doctoral Dissertation.)

Sjöberg, L. E.: Application of GPS in detailed surveying, ZfV, Vol. 121, No. 10, pp. 485-491, 1996.

Sjöberg, L. E.: On optimality and reliability for GPS base ambiguity resolution by combined phase and code observables. ZfV, Vol. 122, no 6, pp. 270-275, 1997.

Sjöberg, L. E.: A new method for GPS base ambiguity resolution by combined phase and code observables. Survey Review, 34, 268, 363-372, 1998.

Sjöberg, L. E.: Unbiased vs biased estimation of GPS phase ambiguities from dual frequency phase and code observables. Journal of Geodesy (in print), 1999.

Map projections and boundary problems

S. Leif Svensson

Abstract
This discussion of the transformation of the spherical or elliptical boundary value problems of physical geodesy into essentially plane problems was inspired by Grafarend and Krumm [1996]. By transforming the Laplacian on the sphere or the ellipsoid under a conformal map projection, transformations may be found for the classical boundary problems. This leads to the idea of performing local geoid computations by variational methods, using perturbation methods for constructing suitable trial functions. As examples stereographic projections and projections of Mercator type are discussed.

1 Introduction

Among the many fields covered in the works of Erik W. Grafarend one is that of the boundary value problems of physical geodesy - some recent contributions are found in Grafarend and Krumm [1996] and in Grafarend and Martinec [1997]. Another field is that of map projections e.g. Grafarend [1995], Grafarend and Syffus [1997], [1998]. However disparate they seem to be, there are connections between the two fields as may be inferred by the ideas of Grafarend and Krumm [1996].
Conventionally, boundary value problems like the Stokes' problem on the sphere are solved in terms of geodetic coordinates (ϕ, λ), which are, in effect, nothing but the $(x, y)-$ coordinates of a plate carré map projection. In terms of this the Laplacian on the sphere of radius 1 may be expressed as

$$\Delta_S = \partial^2/\partial\phi^2 - \tan\phi\,\partial/\partial\phi + (\cos\phi)^{-2}\partial^2/\partial\lambda^2.$$

By way of this representation eigenvalues and eigenfunctions (surface harmonics) of Δ_S are computed and they may be used also for solving problems

$$q(\Delta_S)u = v$$

of pseudodifferential equation type as for instance the Stokes problem

$$R^{-1}\{(-\Delta_S + 1/4)^{1/2} - 3/2\}u = v,$$

where v is gravity anomaly and u geoidal height.
Then the question arises: If other map projections than the plate carré projection are used, how does Δ_S transform and is the result in any way useful for dealing with the problems of physical geodesy?
Concerning the first part of the question the answer, if we concentrate to conformal transformations, is

$$\Delta_S = s^2\Delta_P,$$

where s^2 is the local area scale, and where Δ_P is the Laplacian in the map plane. This formula is derived in section 2. The formula may be used to transform the classical formulae of physical geodesy to expressions in terms of the (x, y) map coordinates.

Of course the fact that s^2 is not constant for conformal transformations complicates the situation but still there might be some areas - e.g. working with heterogeneous data locally or regionally - where conformal map projections might be useful. The applicability is discussed in section 3. Sections 4-5 are devoted, in turn, to the stereographic projection and to Mercator type projections. Conformal conical projections and the Mercator projection of an ellipsoidal of rotation could be dealt with in the same way.

2 Transformation of the problems

Consider a map projection π mapping points on the Earth sphere S_R of radius R into the map plane P. We shall compute the transformations of pseudodifferential operators $q(\Delta_S)$ on the sphere to operators in the plane. It is natural to consider only conformal transformations π, since the invariance of the operators under transformations preserving the Riemann geometry should be recognized.
We start by recollecting some facts and introducing some notations relating to the conformal map projection π. The projection induces a linear mapping π_* from the tangent space T_a to $T_{\pi a}$ for any point $a \in S$. For T_a there is an orthonormal system $\boldsymbol{e}_\phi = R^{-1}\partial/\partial\phi$ $\boldsymbol{e}_\lambda = R^{-1}(\cos\phi)^{-1}\partial/\partial\lambda$, where, for azimuthal or conical projections, ϕ is angular distance to the centre point on the sphere and λ azimuth from that point. For the Mercator projection ϕ is instead latitude and λ longitude. When dealing with transversal Mercator mappings, ϕ will be angular distance *to* and angle *along* the central meredian, respectively. For $T_{\pi a}$, which may be identified with the plane P, we use the standard orthonormal system $\boldsymbol{e}_x = \partial/\partial x$ and $\boldsymbol{e}_y = \partial/\partial y$. In terms of the ortonormal basis systems $\boldsymbol{e}_\phi, \boldsymbol{e}_\lambda$ and $\boldsymbol{e}_x, \boldsymbol{e}_y$ the linear mapping π_* is given by the matrix

$$A = R^{-1}\begin{bmatrix} \partial x/\partial\phi & \cos^{-1}\phi\partial x/\partial\lambda \\ \partial y/\partial\phi & \cos^{-1}\phi\partial y/\partial\lambda. \end{bmatrix}$$

Since we assume the map projection to be conformal we have

$$A = sU,$$

where s is the local (length) scale of the map and U is an orthogonal matrix. The area scale is, consequently, s^2.
Now, turning to the problem of transforming the Laplacian Δ_S on the sphere, we shall relate it to the Laplacian Δ_P in the plane. For that purpose it is convenient to use exterior algebra (see e.g. Flanders [1963]). For a general Riemann manifold M - we shall indeed restrict ourselves to an orientable manifold of dimension 2 or, even more specific, to a sphere or an ellipsoid of rotation - the Laplacian of a function f is defined by

$$\Delta f = *d * df,$$

where d is exterior differentiation, and $*$ the Hodge star operator. For a conformal map projection $(x, y) = \boldsymbol{x} = \boldsymbol{x}(\phi, \lambda)$ of a sphere or an ellipsoid of rotation into the plane

$$\begin{aligned} \boldsymbol{e}_\phi &= \frac{1}{s}\frac{\partial \boldsymbol{x}}{\partial\phi} \\ \boldsymbol{e}_\lambda &= \frac{1}{s\cos\phi}\frac{\partial \boldsymbol{x}}{\partial\lambda} \end{aligned}$$

form (cf. the matrix $A = sU$) an orthogonal moving frame for the plane. Hence

$$d\boldsymbol{x} = sd\phi\boldsymbol{e}_\phi + s\cos\phi d\lambda\boldsymbol{e}_\lambda$$

and we find an orthogonal frame for one-forms in the plane

$$\begin{aligned}\sigma_1 &= s d\phi \\ \sigma_2 &= s\cos\phi d\lambda\end{aligned}$$

with proper orientation so that

$$\begin{aligned}*\sigma_1 &= \sigma_2 \\ *\sigma_2 &= -\sigma_1 \\ *(\sigma_1\sigma_2) &= 1.\end{aligned}$$

Now we can compute the Laplacian

$$\begin{aligned}\Delta_P f &= *d*(s^{-1}f'_\phi\sigma_1 + (s\cos\phi)^{-1}f'_\lambda\sigma_2) \\ &= *d(s^{-1}f'_\phi\sigma_2 - (s\cos\phi)^{-1}f'_\lambda\sigma_1) \\ &= *d(\cos\phi f'_\phi d\lambda - (\cos\phi)^{-1}f'_\lambda d\phi) \\ &= *((\cos\phi f'_\phi)'_\phi + (\cos\phi)^{-1}f''_{\lambda\lambda})d\phi d\lambda) \\ &= *(s^{-2}(\Delta_S f)s^2\cos\phi\, d\phi d\lambda) \\ &= *((s^{-2}\Delta_S f\, dxdy) \\ &= s^{-2}\Delta_S f\end{aligned}$$

We formulate this as a theorem.

Theorem 1 *If π is a conformal map projection of the sphere S_R into the plane P, the corresponding relation between the Laplacians is*

$$\Delta_S = s^2\Delta_P,$$

where s is the length scale of the projection.

Now it is easy to see how invariant pseudodifferential operators on the sphere transform.

Corollary 1 *If $p = q(\Delta_S)$ is an invariant pseudodifferential operator on the sphere S, it is transformed, by a conformal map projection π with local area scale s^2, to the operator $p = q(s^2\Delta_P)$ in the plane.*

3 Application to physical geodesy

The Stokes' problem on the sphere may be written - see Svensson [1983] - as

$$R^{-1}\{(-\Delta_S + 1/4)^{1/2} - 3/2\}u = v$$

and hence, according to Theorem 1 in terms of the map coordinates

$$R^{-1}\{(-s^2\Delta_P + 1/4)^{1/2} - 3/2\}u = v.$$

For the fixed boundary value problem the corresponding equations are

$$R^{-1}\{(-\Delta_S + 1/4)^{1/2} + 1/2\}u = v \tag{1}$$

and

$$R^{-1}\{(-s^2\Delta_P + 1/4)^{1/2} + 1/2\}u = v.$$

Globally there does not seem to be much of a point in using map coordinates, since obviously we will get the eigenvalues and eigenfunctions just by taking them from the sphere and transforming

them by the projection. However, locally or regionally, the situation may be different. We are going to discuss mainly local problems, which may be solved by a perturbation technique. Hence, consider the problem

$$R^{-1}\{(-s^2\Delta_P + 1/4)^{1/2} - 3/2\}u = v \text{ in } \Omega \quad (2)$$

$$u = 0 \text{ on } \delta\Omega \quad (3)$$

Here $H_{1/2}$ denotes the Sobolev-Slobodeckiĭ space of all u such that $(-\Delta_P + 1)^{1/4}u$ is square integrable, Ω is a domain in the map plane, and $H^\circ_{1/2}(\Omega)$ is the subspace of u vanishing outside Ω

One approach is to work with eigenvalues and eigenfunctions for Dirichlet problems

$$pu = -s^2\Delta_P u = \lambda u \text{ in } \Omega$$
$$u = 0 \text{ on } \delta\Omega.$$

Such eigenfunctions u satisfy

$$q(s^2\Delta_p)u = q(\lambda)u$$
$$u \in H^\circ_{1/2}(\Omega).$$

For the approximate case s =constant, eigenfunctions for many important classes of domains are very well known. Hence, for a circular disc of radius ρ the eigenvalues are $\lambda = (\alpha_{nk}/\rho)^2$, where α_{nk} is the k:th zero of the Bessel function J_n and the eigenfunctions are $u = J_0(r/\rho)$ if $n = 0$ and $u = J_n(r/\rho)\sin(n\theta)$ or $u = J_n(r/\rho)\cos(n\theta)$ if $n > 0$.For a rectangle with sides parallell to the $x, y-$ directions and of lengths a, b the eigenvalues are $\lambda_{jk} = ((j/a)^2 + (k/b)^2)^{1/2}$, $j, k \geq 0$ with eigenfunctions $(\sin j(x - x_0)\pi/a) \cdot (\sin k(y - y_0)\pi/b)$, (x_0, y_0) being the lower left corner of the rectangle.

In order to compute corresponding eigenvalues and eigenfunctions for the perturbed system we recall briefly the technique in the simplest case. Hence, let p be a formally selfadjoint operator, densely defined on a Hilbert space H and assume that λ is an eigenvalue with a single eigenfunction u. Let $p + \delta p$ also be selfadjoint and assume that δp is, in some sense small. Then one might hope that there is an eigenvalue $\lambda + \delta\lambda$ and an eigenfunction $u + \delta u$ in some sense close to λ and u respectively. The computation is iterative. First we formulate the equation

$$(p + \delta p)(u + \delta u) = (\lambda + \delta\lambda)(u + \delta u)$$

and rewrite it as

$$(p - \lambda)\delta u = -(\delta p - \delta\lambda)u + \delta\lambda\delta u - \delta p\delta u$$

or, with $\delta u_0 = 0$, $\delta\lambda_0 = 0$,

$$(p - \lambda)\delta u_{n+1} = -(\delta p - \delta\lambda_{n+1})u + \delta\lambda_n\delta u_n - \delta p\delta u_n \quad (4)$$

the condition for solvability of the equation is the orthogonality condition

$$((\delta p - \delta\lambda_{n+1})u + \delta\lambda_n\delta u_n - \delta p\delta u_n, u) = 0,$$

which yields

$$\delta\lambda_{n+1} = (\delta p u + \delta\lambda_n\delta u_n - \delta p\delta u_n, u)/(u, u)$$

and $\delta u_{n+1}s$ as the solution (preferrably chosen orthogonal to u in order to get uniqueness) of equation (4).

The mixed problem (2) may be solved by variational methods: Minimize $(pu - v, u)$, $u \in M_0$, where the standard scalar product in the plane is used and M_0 is a trial function space of finite dimension. Here we must assume that p is a positive operator on $H^\circ_{1/2}(\Omega)$. This is the case if

the diameter of Ω on the sphere is less than 149° - see Svensson [1988]. For the fixed boundary value problem (1) p is positive for any Ω.
Assume that $\{u_j\}_{j=1}^m$ is a basis for M_0 and that each u_j is an eigenfunction for p i.e. $pu_j = \lambda_j u_j$. Then the variational problem ends up with the problem of solving the system

$$\sum_{k=1}^{m}(pu_j, u_k)x_k = (v, u_j),\ j = 1, 2, \ldots m$$

or

$$\sum_{k=1}^{m}\lambda_j(u_j, u_k)x_k = (v, u_j),\ j = 1, 2, \ldots, m$$

and putting

$$u = \sum_{j=1}^{m} x_k u_k.$$

Hence, eigenfunctions computed by the perturbation technique may be used favourably as trial functions. One may use for example rectangular or triangular grids and eigenfunctions vanishing outside individual rectangles or triangles.
Another approach is an iterative minimization, where the trial functions used are eigenfunctions but now of the approximate operator

$$\tilde{p} = q(s_0^2\Delta),$$

where s_0 is an approximate value for s in the region. We put $u_0 = 0$ and minimize for $n = 0, 1, 2, \ldots$

$$(\tilde{p}u_{n+1} + (p - \tilde{p})u_n - v, u_{n+1}).$$

The question of the convergence of the iteration schemes, wether in the direct or indirect eigenvalue approach, is rather technical and we leave it for the time being.

4 Stereographic projection

Among the azimuthal projections, the stereographic is unique in that it is perfectly conformal. The projection is given by

$$(x, y) = (c\sin\phi/(1 + \cos\phi)(\cos\lambda, \sin\lambda).$$

where c/R is the length scale of the map at the centre, and

$$s = c/(R(1 + \cos\phi))$$

the local length scale varying with the distance to the centre. Expressing this in terms of the map coordinates we get

$$s = \frac{c}{2R}(1 + (\frac{r}{c})^2),$$

where $r = \sqrt{x^2 + y^2}$.

5 The Mercator type projections

Among the cylindrical map projections we only consider the one which is conformal, the Mercator projection. It is given by

$$(x, y) = c(\ln(\tan(\pi/4 + \phi/2), \lambda)$$

with

$$s = c/(R\cos\phi).$$

Here ϕ is angular distance to the central meredian and λ arc length measured along the central meredian. Expressed in map coordinates we get

$$s = (c/R)(\cosh x/c)^{-1}$$

Hence

$$\Delta_S = (c/R)^2(\cosh x/c)^{-2}(\partial^2/\partial x^2 + \partial^2/\partial y^2)$$

It should be observed that, since the scale factor here depends upon x only, computation of eigenfunctions, vanishing outside rectangles, is faciliated by simplifying the problem essentially to a problem in one dimension. The eigenfunctions then have the form

$$u = w(x) \cdot \sin k\pi(y - y_0)/b,$$

where

$$-s^2(x)(w''(x) - (k\pi/b)^2) = \lambda w.$$

References

[1963] Flanders, H., Differential Forms, Academic Press, New York, London, (1963).

[1995] Grafarend, E, W., The optimal universal transverse Mercator projection, Manuscripta Geodaetica, 20, 421-468 (1995).

[1996] Grafarend, E. W., Krumm, F., The Stokes and Vening-Meinesz functionals in moving tangent space, J. Geodesy, 70, 419-430 (1996).

[1997] Grafarend, E. W., Martinec, Z., Solution to the Stokes boundary value problem on an ellipsoid of revolution, Studia Geophys.& Geod. 41, 103-129 (1997)

[1997] Grafarend, E. W., Syffus, R., Mixed cylindric map projections of the ellipsoid of revolution, J. Geodesy, 71, 685-694 (1997).

[1998] Grafarend, E. W., Syffus, R., The solution of the Korn-Lichtenstein equations of conformal mapping: the direct generation of ellipsoidal Gauss-Krueger conformal coordinates or the transversal Mercator projection, J. Geodesy, 72, 282-293 (1998).

[1983] Svensson, S. L., Solution of the altimetry-gravimetry problem, Bull. Géod., Vol. 8 (1983)

[1984] Svensson, S. L., Finite Elements on the Sphere, J. Appr. Th., Vol. 40 (1984)

[1988] Svensson, S. L., Some remarks on the altimetry-gravimetry problem, Manuscr. Geod., Vol. 13 (1988)

GPS, INTEGERS, ADJUSTMENT AND PROBABILITY

Peter J.G. Teunissen

ABSTRACT

In this contribution we give a brief review of the theory of integer estimation as it has been developed for use with the Global Positioning System (GPS). First we consider the GPS observation equations (*GPS and Integers*). These observation equations are not of the usual type, since some of the parameters are known to be integer. This implies that an extention of 'classical' adjustment theory is needed. Such an extention is presented by showing ways of how to solve integer adjustment problems (*Integers and Adjustment*). Different integer estimators are given and a class of admissible integer estimators is defined. Next we consider some qualitative aspects of an integer adjustment (*Adjustment and Probability*). It is argued that the usual qualitative description by means of second moments or variance-covariance matrices is not sufficient. A direct probabilistic description is needed instead. Such a description is presented by means of the probability mass function of the integer ambiguities. For GPS ambiguity resolution, the probability of correct integer estimation is particularly of interest (*Probability and GPS*). It describes the ambiguity succes rate and shows whether or not the estimated ambiguities may be treated deterministically. Since different integer estimators have different success rates, one is particularly interested in the estimator which maximizes this success rate. The answer is given by the theorem provided.

1 GPS and Integers

As witnessed by the enormous GPS literature available, there exists a great variety in GPS models currently in use. They may range from single- site models used for local monitoring purposes to multi-baseline models used as a tool for studying geodynamic phenomena, or from supershort multi-baseline models used for local attitude determination to wide-area models for transmitting differential corrections. Depending on the application at hand, each one of these models may differ in the way the observed signals are linked, in the way the reference systems and the orbits are treated, or in the way the receiver and propagation delays are modelled. An overview of these and other GPS models, together with their applications in surveying, navigation, geodesy and geophysics, can be found in textbooks such as [*Hofmann-Wellenhof et al.*, 1997], [*Leick*, 1995], [*Parkinson and Spilker*, 1997], [*Strang and Borre*, 1997] and [*Teunissen and Kleusberg*, 1998].

Despite the differences in application of the various GPS models, their observation equations are to a large part the same. For the single baseline case, using two receivers, each tracking the same two satellites, the double differenced (DD) phase and code observation equations for a single epoch i read

$$\begin{aligned} \phi_1(i) &= \rho(i) - \mu_1 I(i) + T(i) + \lambda_1 a_1 \\ \phi_2(i) &= \rho(i) - \mu_2 I(i) + T(i) + \lambda_2 a_2 \\ p_1(i) &= \rho(i) + \mu_1 I(i) + T(i) \\ p_2(i) &= \rho(i) + \mu_2 I(i) + T(i) \end{aligned} \tag{1}$$

where ϕ_1 and ϕ_2 are the DD phase observables on L_1 and L_2; p_1 and p_2 are the DD code

observables on L_1 and L_2; ρ is the DD-form of the unknown receiver-satellite range; I and T are the DD-forms of resp. the unknown ionospheric and troposheric delay, and a_1 and a_2 are the unknown integer ambiguities. The known wavelengths are denoted as λ_1, λ_2. Since the ionospheric delay is to a first order inversely proportional to the square of the frequency, we have to the same degree of approximation $\mu_1 = \frac{\lambda_1}{\lambda_2} = \frac{60}{77}$, $\mu_2 = \frac{\lambda_2}{\lambda_1} = \frac{77}{60}$.
Depending on the application the atmospheric delays may be assumed present or absent. For sufficiently short baselines, the ionospheric delays can usually be assumed absent. The same holds true for the troposheric delays in case the differences in height are not too large. When one is forced to assume the delays to be present, a strenghtening of the model is sometimes still possible. This can be done by either using an *a priori* weighting of the delays or by using a further parametrization in fewer parameters. For instance, since the ionosphere decorrelates spatially, the ionospheric delays are often weighted a priori as function of the baseline length. And in case of the troposphere, a parametrization in fewer parameters is often done using the so-called mapping functions.
Although the above equations are already useful for many applications, they are not yet useful for positioning purposes. For that to be the case, a further parametrization of $\rho(i)$ into the unknown baseline vector between the two receivers is needed. Such models are referred to as *geometry-based*, because of the explicit presence in the model of the relative receiver-satellite geometry. Without such a parametrization the model is referred to as *geometry-free*.
In all applications where the above equations form the backbone of the particular GPS model used, the unknown carrier phase ambiguities a_i, a_2 are known to be *integer*. Hence, despite the differences in application of the various GPS models, there should in principle be no difference in how these integer parameters are dealt with. Their ambiguity resolution problems should intrinsically be the same. This implies that it should be possible to develop a single theoretical framework that applies to every GPS model for which ambiguity resolution would make sense. Such a framework did not exist in the pre-GPS era, simply because 'classical' adjustment theory is not equiped to deal with integer parameters. Even in the GPS era it took quite some time before the theoretical framework, as we know it today, took its shape. It started with the development of methods of integer estimation. Although some of the methods proposed in the beginning were ad hoc, inefficient and sometimes even simply wrong, the current state of affairs is that one can indeed speak of a reasonable mature theory of integer estimation. However, this can not yet be said of the necessary probabilistic theory. It is amazing to see that in almost all contributions the probabilistic part is either ignored completely or thought to be solved by 'classical' means such as those of 'the linear model'. It is only recently that integer adjustment has been complemented to some extent with some of the necessary probabilistic theory. In this contribution we will confine ourselves to this probabilistic part and briefly review some of the results as obtained in [*Teunissen*, 1997, 1998, 1999].

2 Integers and Adjustment

In principle all GPS models that are based on (1) can be cast in the following conceptual frame of linear(ized) observation equations

$$y = Aa + Bb + e \tag{2}$$

where y is the given GPS data vector of order m, a and b are the unknown parameter vectors respectively of order n and o, and where e is the noise vector. The matrices A and B are the corresponding design matrices. They are assumed to be of full rank. The data vector y will usually consist of the 'observed minus computed' single- or dual- frequency double-difference (DD) phase and/or pseudorange (code) observations accumulated over all observation epochs. The entries of vector a are then the DD carrier phase ambiguities, expressed in units of cycles rather than range. They are known to be *integers*, $a \in Z^n$. The entries of the vector b will

consist of the remaining unknown parameters, such as for instance receiver-satellite ranges in case of the geometry-free model or baseline coordinates in case of the geometry-based model, and possibly atmospheric delay parameters as needed for the troposphere and/or ionosphere. All the parameters collected in b are real-valued.
The procedure which is usually followed for solving the GPS model (2), can be divided into three steps. In the *first* step one simply disregards the integer constraints $a \in Z^n$ on the ambiguities and performs a standard adjustment. As a result one obtains the (real-valued) estimates of a and b, together with their variance-covariance (vc-) matrix

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}, \begin{bmatrix} Q_{\hat{a}} & Q_{\hat{a}\hat{b}} \\ Q_{\hat{b}\hat{a}} & Q_{\hat{b}} \end{bmatrix} \tag{3}$$

In GPS terminology this solution is referred to as the 'float' solution. In the *second* step the 'float' ambiguity estimate $\hat{a}$ is used to compute the corresponding integer ambiguity estimate $\check{a}$. This implies that a mapping $F : R^n \mapsto Z^n$ is introduced, from the n-dimensional space of reals to the n-dimensional space of integers, such that

$$\check{a} = F(\hat{a}) \tag{4}$$

Once these integer ambiguities are computed, they are used in the *third* step to finally correct the 'float' estimate of b. As a result one obtains the 'fixed' solution $\check{b} = \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1}(\hat{a} - \check{a})$.
It will be clear that the actual integer adjustment is confined to the second step of above, in particular to the choice of the map F. Due to the discrete nature of Z^n, the map F will not be one-to-one, but instead a many-to-one map. This implies that different real-valued ambiguity vectors will be mapped to the same integer vector. One can therefore assign a subset $S_z \subset R^n$ to each integer vector $z \in Z^n$:

$$S_z = \{x \in R^n \mid z = F(x)\}, \quad z \in Z^n \tag{5}$$

The subset S_z contains all real-valued ambiguity vectors that will be mapped by F to the same integer vector $z \in Z^n$. This subset is referred to as the *pull-in-region* of z. It is the region in which all ambiguity 'float' solutions are pulled to the same 'fixed' ambiguity vector z.
Since the pull-in-regions define the integer estimator completely, one can define classes of integer estimators by imposing various conditions on the pull-in-regions. Such a class was introduced by the author and referred to as the class of admissible integer estimators. These integer estimators are defined as follows:

Definition
An integer estimator is said to be *admissible* if

$$\begin{array}{ll} (i) & \bigcup_{z \in Z^n} S_z = R^n \\ (ii) & S_{z_1} \bigcap S_{z_2} = \{0\}, \ \forall z_1, z_2 \in Z^n, z_1 \neq z_2 \\ (iii) & S_z = z + S_0, \ \forall z \in Z^n \end{array} \tag{6}$$

This definition is motivated as follows. Each one of the above three conditions describe a property of which it seems reasonable that it is possessed by an arbitrary integer ambiguity estimator. The first condition states that the pull-in-regions should not leave any gaps and the second that they should not overlap. The absence of gaps is needed in order to be able to map any 'float' solution $\hat{a} \in R^n$ to Z^n, while the absence of overlaps is needed to guarantee that the 'float' solution is mapped to just one integer vector. Note that we allow the pull-in-regions to have common boundaries. This is permitted if we assume to have zero probability that $\hat{a}$ lies on one of the boundaries. This will be the case when the probability density function (pdf) of $\hat{a}$ is continuous.

The third and last condition follows from the requirement that $F(x+z) = F(x)+z, \forall x \in R^n, z \in Z^n$. Also this condition is a reasonable one to ask for. It states that when the 'float' solution is perturbed by $z \in Z^n$, the corresponding integer solution is perturbed by the same amount. This property allows one to apply the *integer remove-restore* technique: $F(\hat{a} - z) + z = F(\hat{a})$. It therefore allows one to work with the fractional parts of the entries of $\hat{a}$, instead of with its complete entries.

There exist many admissible integer estimators. In fact, one can 'invent' one's own admissible integer estimator by simply specifying pull-in-regions that satisfy the above definition. Here we will give as an example the pull-in-regions of three more commonly used admissible estimators. The simplest integer map is the one that corresponds to an integer rounding. In this case the integer vector is obtained from a rounding of each of the entries of $\hat{a}$ to its nearest integer. Since componentwise rounding implies that each real-valued ambiguity estimate $\hat{a}_i, i = 1, \ldots, n$, is mapped to its nearest integer, the absolute value of the difference between the two is at most $\frac{1}{2}$. The subsets $S_{R,z}$ that belong to this integer estimator are therefore given as

$$S_{R,z} = \cap_{i=1}^{n} \{\hat{a} \in R^n \mid \mid \hat{a}_i - z_i \mid \leq \frac{1}{2}\}\ ,\ \forall z \in Z^n \tag{7}$$

The subset $S_{R,z}$ is an n-dimensional cube, with sides of length 1 and centred at the grid point z.

Another relatively simple integer ambiguity estimator is the bootstrapped estimator. This estimator can be seen as a generalization of the previous one. It still makes use of integer rounding, but it also takes some of the correlation between the ambiguities into account. The bootstrapped estimator results from a sequential conditional least- squares adjustment and is computed as follows. If n ambiguities are available, one starts with the first ambiguity $\hat{a}_1$, and rounds its value to the nearest integer. Having obtained the integer value of this first ambiguity, the real-valued estimates of all remaining ambiguities are then corrected on the basis of their correlation with the first ambiguity. Subsequently the second, but now corrected, real-valued ambiguity estimate is rounded to its nearest integer. Having obtained the integer value of the second ambiguity, the real-valued estimates of all remaining $n-2$ ambiguities are again corrected, but now on the basis of their correlation with the second ambiguity. This process of rounding and correcting is continued until all ambiguities are taken care of.

With c_i denoting the ith canonical unit vector having a 1 as its ith entry, the pull-in-regions $S_{B,z}$ that belong to the bootstrapped estimator can be shown to be given as

$$S_{B,z} = \cap_{i=1}^{n} \{\hat{a} \in R^n \mid \mid c_i^T L^{-1}(\hat{a} - z) \mid \leq \frac{1}{2}\}\ ,\ \forall z \in Z^n \tag{8}$$

with matrix L being the lower triangular unit matrix that follows from applying a triangular decomposition to the variance-covariance matrix of $\hat{a}$. Note that these pull-in-regions reduce to the ones of (7) when L becomes diagonal. This is the case when the ambiguity variance-covariance matrix is diagonal. In that case the two integer estimators $\check{a}_R$ and $\check{a}_B$ are identical.

The third admissible estimator of which the pull-in-region will be given is the integer least-squares estimator. By again using the LDU-decomposition of $Q_{\hat{a}}$ the least-squares' pull-in-region reads

$$S_{LS,z} = \cap_{c_i \in L^{-1}(Z^n)} \{\hat{a} \in R^n \mid \mid c_i^T D^{-1} L^{-1}(\hat{a} - z) \mid \leq \frac{1}{2} c_i^T D^{-1} c_i\} \tag{9}$$

Note that (9) and (8) become identical when the matrix entries of L^{-1} are all integer. This is the case when L is an admissible ambiguity transformation.

3 Adjustment and Probability

In the previous section it was shown how an integer adjustment can be performed. One first needs to define the type of integer estimator. This is done by choosing the integer map $F : R^n \mapsto Z^n$

or by choosing the corresponding pull-in-regions $S_z \subset R^n$. Once the type of integer estimator is chosen, the actual adjustment can be executed. This is done by searching for the pull-in-region in which the data vector lies. The integer solution follows then once this pull-in-region is identified.

It will be clear that a mere adjustment (or parameter estimation) is not enough. One also needs to be able to describe the quality of the adjustment result. After all one can always perform an adjustment whether the data are of good quality or not. In 'classical' adjustment theory a special place is reserved for the variance-covariance (vc-) matrix. It is often the vc-matrix of the estimated parameters which is used to decribe the quality. In 'classical' adjustment theory, the vc-matrix is even used as criterion for deriving optimal estimators. The least-squares estimator for instance, is known to be optimal in the sense that its variance is the smallest of all linear unbiased estimators.

Although the vc-matrix is often used in practice, the reason for its usage as quality measure is not always stated explicitly and unequivocally. Why is the vc-matrix used and what does it tell us? The vc-matrix of a random vector is defined as the second (central) moment of the vector's distribution. It describes how large the squared differences between sample values and the mean (the first moment) will be *on average*. Thus the vc-matrix is a measure of the expected spread around the random vector's mean. In practice the term 'precision' is used for this notion. But although such precision information is of importance in its own right, it is not the only reason for the popular usage of the vc-matrix. In practice, the popularity of the vc-matrix also stems from the fact that often the vc-matrix is used as a tool for constructing confidence regions. Standard ellipses for instance, are often interpreted as such confidence regions. This particular usage of the vc-matrix implies however that one assumes the random vector to be normally distributed. Since the normal distribution is completely specified once its first two moments are known, knowledge of the vc-matrix is sufficient for determining confidence regions and for determining the probability that the outcome of the estimator stays within a certain limit from its mean. In general however, the information content of the vc-matrix is not enough to determine confidence regions. For making such probability statements, the complete probability distribution is needed and not only its second (central) moment. After all, different distributions can have identical vc-matrices. The definition of the second (central) moment is not even restricted to continuous, unimodel distributions, but applies to discrete and multimodal distributions as well.

The above makes clear that in our case of an integer adjustment, the concept of the vc-matrix is much less useful. Although the vc-matrix still describes the 'expected spread', it can not be used anymore to determine probabilities and confidence regions. After all, in case of an integer adjustment the distribution of the estimator will be of the discrete type, thus not continuous and certainly not normal or Gaussian. The distribution of the integer estimator $\check{a}$ will be a probability *mass* function (pmf), which we shall denote as $P(\check{a} = z)$, with $z \in Z^n$. Thus $P(\check{a} = z)$ denotes the probability that the ambiguity vector $\check{a}$ equals the integer vector $z \in Z^n$. In order to determine this distribution, we first need the probability density function (pdf) of $\hat{a}$. The pdf of $\hat{a}$ will be denoted as $p_a(x)$, with $x \in R^n$. The subindex is used to show that the pdf still depends on the unknown parameter vector $a \in Z^n$. In case the pdf of $\hat{a} \in R^n$ is *elliptically contoured* it is of the form

$$p_a(x) = \sqrt{\det(Q_{\hat{a}}^{-1})}\, G(\| x - a \|^2_{Q_{\hat{a}}}) \tag{10}$$

where $G : R \mapsto [0, \infty)$ is decreasing and $Q_{\hat{a}}$ is positive-definite. Several important distributions belong to this family. The multivariate normal distribution can be shown to be a member of this family by choosing $G(x) = (2\pi)^{-\frac{n}{2}} \exp -\frac{1}{2}x$, $x \in R$. Another member is the multivariate t-distribution.

The pmf of $\check{a}$ can now be obtained as follows. Using the concept of the pull-in-region, the integer estimator is defined as $\check{a} = z \;\Leftrightarrow\; \hat{a} \in S_z$. This shows that $P(\check{a} = z) = P(\hat{a} \in S_z)$. The probability that $\check{a}$ coincides with z is therefore given by the integral of the pdf $p_a(x)$ over the

pull-in-region $S_z \subset R^n$. Hence, the pmf of $\check{a}$ follows as

$$P(\check{a} = z) = \int_{S_z} p_a(x)dx \ , \ \forall z \in Z^n \tag{11}$$

It is this function which gives a complete description of the random characteristics of the integer ambiguity estimator $\check{a}$. Hence, it is this function which should be used when studying the qualitative aspects of the integer estimator.

4 Probability and GPS

The quality of the integer ambiguity estimator is particularly of interest in case of GPS. In case of GPS one usually treats the estimated integer ambiguities as if they were deterministic variates. Theoretically this is not correct as the previous section has shown. Neglecting the random nature of the estimated integer ambiguities when applying the error propagation law to

$$\check{b} = \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1} (\hat{a} - \check{a}) \tag{12}$$

implies that a too optimistic quality description is obtained for the so-called 'fixed' estimator $\check{b}$. Whether this is acceptable or not depends of course on the approximation involved. One should therefore have a diagnostic tool available on the basis of which one can decide whether the approximation is acceptable or not. This diagnostic tool is provided by the *ambiguity success rate*, which is defined as the probability of correct integer estimation

$$\text{ambiguity success rate} = P(\check{a} = a) \tag{13}$$

One should therefore first compute the ambiguity success rate and check whether it is sufficiently close to one, before deciding that a deterministic treatment of $\check{a}$ is acceptable.
Note that the pmf (11) as well as the success rate (13) still depend on the type of pull-in-region and thus on the type of integer estimator chosen. Changing the geometry of the pull-in-region will change both the pmf and the ambiguity success rate. It is therefore not only of theoretical interest, but also of practical interest, to know which integer estimator maximizes the ambiguity success rate. The answer is given by the following theorem:

Theorem (*Teunissen*)
Let the pdf of $\hat{a}$ be elliptically contoured and the integer least-squares estimator be given as

$$\check{a}_{LS} = \arg \min_{z \in Z^n} \| \hat{a} - z \|^2_{Q_{\hat{a}}}$$

Then

$$P(\check{a}_{LS} = a) \geq P(\check{a} = a) \tag{14}$$

for any admissible estimator $\check{a}$.

This theorem gives a probabilistic justification for using the *integer* least-squares estimator. As a probabilistic justification it may be considered to replace the theorem of Gauss which states that the *real-valued* least-squares estimator has smallest variance of all linear unbiased estimators. The theorem particularly applies to GPS ambiguity resolution, for which often the multivariate normal distribution is assumed to hold true. For GPS ambiguity resolution one is thus better off using the integer least-squares estimator than any other admissible integer estimator, such as, for instance, the 'rounding' estimator or 'bootstrapped' estimator.

5 References

[1] Hofmann-Wellenhof, B., H. Lichtenegger, J. Collins (1997): *Global Positioning System: Theory and Practice.* 4th edition. Springer Verlag.

[2] Leick, A. (1995): *GPS Satellite Surveying*, 2nd edition, John Wiley and Sons, New York.

[3] Parkinson, B., J.J. Spilker (eds) (1996): *GPS: Theory and Applications*, Vols 1 and 2, AIAA, Washington DC.

[4] Strang, G. and K. Borre (1997): *Linear Algebra, Geodesy, and GPS*, Wellesley-Cambridge Press.

[5] Teunissen, P.J.G. (1997): Some remarks on GPS ambiguity resolution. *Artificial Satellites*, Vol. 32, No. 2, pp. 119-130.

[6] Teunissen, P.J.G. (1998): On the integer normal distribution of the GPS ambiguities. *Artificial Satellites*, Vol. 33, No. 2, pp. 49-64.

[7] Teunissen, P.J.G. (1999): A theorem on maximizing the probability of correct integer estimation. *Artificial Satellites*, Vol. 34, No. 2.

[8] Teunissen, P.J.G., A. Kleusberg (eds) (1998): *GPS for Geodesy*, 2nd enlarged edition, Springer Verlag.

GEOPHYSICAL GEODESY BEYOND 2000

Péter Varga

Abstract

In this paper some of the problems of the physical geodesy, not solved till the end of the second millennium are discussed. They are the problems connected to the Newtonian law of gravity first of all. The gravitational constant is the most inaccurately determined constant of the nature and the questions of the so-called non-Newtonian gravimetry are also not answered. An another problem which needs further investigations is connected to the Mac Cullagh theorem: the time derivative of the second zonal geopotential coefficient obtained from satellite orbit determinations is of about 50 times bigger than those obtained from the spindown value. The study of the tidal friction attracted a lot of interest in the course of the XXth century. So great names as G. Darwin, A. Einstein, H.Jeffreys expressed their interest in this field. In spite of this there are many important questions unsolved. Among them should be mentioned the problem of the rotational history of our planet, the temporal variation of the figure, of the normal gravity field of the Earth and the problem of the changes in its inner structure.

1. Introduction

Our knowledge on the Earth increased dramatically in the course of last two, our three decades. The progress in Earth sciences and among them in geophysical geodesy was reached first of all due to the development in the technology. The new possibilities in computing techniques, in physics of very low temperatures, in satellite sinence allows to produce equipment's the realization which was possible only in the world of scifies earlier. The methods of space geodesy, as Very Long Base Interferometry (VLBI), Satellite and Lunar Laser Ranging (SLR and LLR), Global Positioning System (GPS), the new absolute (ballistic) and relative (criogenic) gravimeters allows to determine the figure and the gravity field together with their temporal variations. The close cooperation of researchers, working in different branches of geo-sciences, allows the progress in interpretation of observed geodynamical phenomena. The use of results of seismology, of plate tectonics, of reology, of earth magnetism studies among others render possible better understanding the figure and the gravity field of the Earth and time dependent phenomena influencing the development of our planet.

To illustrate the progress the comparison of earth tidal research targets of sixties can be compared with those of the present. Thirty years ago the tidal potential catalogue consists 377 elementary waves and the accuracy of the tidal development was ~ 1 nm $\cdot$ s^{-2}. Today we have 12935 waves and the accuracy of this serie is ~700 times higher as earlier. In the course of sixties it was supposed that the best places for tidal studies are the central part of continents because at that places the indirect effect of oceanic tides not modifies the earth tidal data. Today we know that it is not true, the oceanic tidal influence is significant everywhere on the mainland's, but this effect can be removed with the use of the most modern cotidal maps. Thirty years ago nobody beleived that the microgal level can be reached in the future with recording gravimeters. At the end of XXth century we are near to talk about nanogal level in gravimetry. Earlier it was supposed that from the global tidal response significant new information's can be obtained - beyond the information's derived from seismology - on physical properties and on inner structure of the Earth. The optimism of "earth-tidalists" was connected first of

all with possible detection of lateral in homogeneities within the Earth from their observations. This hope was not confirmed by the development in the science, but there are another fields, where tidal research is able to provide knowledge which seismology can not. This is first of all the study of the nearly diurnal resonance of the Earth. At the same time the modern criogenic recording gravimeters are able monitor beside the tides the polar motion and environmental phenomena, too.

2. Problems related to Newtonian law of gravity

First of all the scale dependence of the Newtonian gravitational constant G will be discussed in this section.

At astronomical distance no G values can be provided due to the unknown masses, but the Newton's law can be tested with high accuracy. (Hubler, Cornaz, Kündig, 1995). In the range of 10^3 m - 10^7 m the inverse - square law has been confirmed satisfactory from a comparison of the Earth-surface data with orbital parameters of the LAGEOS satellite. At geophysical distances (10^2m-10^4m) many experiments ware carried out. G_∞ obtained from this determinations differs significantly from the results of G_0 values obtained in laboratories at distances 10^{-2} - 10^0 m (Stacey et at.,1987) . A theoretical basis for this deviation in principle can be given by introduction of the so called Yukawa type non-Newtonian gravity potential (Stacey et al., 1987):

$$V = -\frac{G_\infty m_1 m_2}{r}(1+\alpha \cdot e^{-\beta}) = V_N + V_Y \qquad (1)$$

If $\alpha = 0, V = V_N$ where V_N is the Newtonian potential. If $\alpha \neq 0$ V_Y appears additionally which is called the Yukawa term.

If $\beta \langle\langle 1$ we get

$$G_0 = G_\infty(1+\alpha) \qquad (2)$$

On the basis of experiments with scale (10^{-2}m - 10^{-4}m) - Airy type experiments - Stacey et al. (1987) got for G_∞ a value which is bigger than the G_0 value obtained in laboratories (10^{-2} - 10^0 m distances). The difference G_∞ - $G_0 \sim 0.01$ what leads to $\alpha \sim 0.0075 \pm 0.0036$.

For the investigation of the reliability of $G_0 \neq G_\infty$ a special experiment was carried out at the Geodynamical Observatory in Budapest deeply under the surface of the Earth. An underground calibration line was set up which consists of 14 stations with a range of 1400±1 μ Gal. Gravity differences, separation and the elevation difference between neighbouring stations are 100 μ Gal, 2-5 m, less than 2 cm respectively.

The gravity values for this horizontal line were determined with a computer regulated Eötvös torsion balance. The instrumental constant of the torsion balance was obtained by the measurement of sensor masses, the length of the arm of the balance and the torsion of the wire. This means: the gravity values of the underground calibration line were obtained without the use of the gravimeters.
On the other hand a new gravimeter calibration device was proposed and designed by Varga (1989) and it also was installed is the underground laboratory of the Geodynamical Observatory Budapest. The principle of this instrument is simple. The artificially induced gravity changes are generated by a suspended cylindrical ring with an inner diameter somewhat bigger than the width of the gravimeter (usually LCR instruments) to be calibrated. The ring is raised and lowered vertically and moved over the gravimeter equipped by a distant reading device and installed on a column of suitable height.

There are many advantages of this calibration procedure:

- The homogeneity of the generated gravity field is very high at the extrema;
- The moved vertically ring does not load the ground around the instrument;
- The gravimeter is stationary during the procedure what is necessary for a small instrumental drift;
- The experiment is symmetrical with respect to the gravimeter and owing to technical reasons the gravity change brought about by the ring is greater than that caused by another geometrically regular body.

Due to this positive features similar device was used in Italy (Achilli et al., 1995) and in the United States (Schwartz, J.P. et al., 1998). All technical problems and the results of the calibrations are described in Varga et al., (1991) and in Varga et al., (1995). With this device absolute calibrations with accuracy of 0.1-0.2% can be carried out.

The difference of the calibration factors obtained for the same gravimeter along the calibration line (i.e. by means of gravitational effect of 10^2-10^4 m scale, air type experiment and derived from measurements with the heavy cylindrical mass (10^{-2}-10^0 m distances)) is of the order of 10^{-3}, what means that the difference between the G_∞ and G_0 is also at most 10^{-3} and not 10^{-2} as was supposed earlier (Stacey et al., 1987).

Another problem connected to the gravitational constant G is its, supposed by many authors, temporal variation. The need of $\frac{dG}{\partial t} \neq 0$ follows from the cosmological considerations. Dirac's expanding Universe model proposed in 1937 naturally leads to a decreasing constant of gravitation and to the theory of the expanding Earth of course. Using Dirac's theory Jordan concluded (1966) that the Earth radius increases with a speed $da/dt = 0.5$ mm· y^{-1}. Similar value for the expansion was derived by Egyed (1997) $da/dt = 0.7$ mm·y^{-1}, who supposed that originally the surface of our planet was as big as the areas of all recent continents together. The most recent and complete description of these theories can be found in the book by Carey (1988).

The critical review of da/dt and consequently of dG/dt can be carried out on the basis of the study of the influence of tides on the long-term variations of the angular speed. Study of this type are usually based on the principle of conversation of angular momentum and it is supposed that the Earth-Moon system is isolated. For the sake of simplicity it can be supposed that the Moon revolves around the Earth on a circular orbit in the plane of the terrestrial equator. In this case Euler's equation can be written as

$$\frac{\partial(C\omega)}{\partial t} = L$$

$$L = \frac{1}{3}\frac{MM_m}{M+M_m} R_m^2 \frac{\partial n_m}{\partial t} \tag{3}$$

In (3) M, C, ω are the mass, the polar moment of inertia and the angular speed of the Earth respectively. M_m, R_m and n_m stands for the mass of the Moon, for the Earth-Moon distance and for the orbital speed of the Moon. Kepler's law can be written as

$$n_m^2 R_m^3 = G(M + M_m)$$

and its time derivative is

$$2n_m R_m^3 \frac{\partial n_m}{\partial t} + 3n_m^2 R_m^2 \frac{\partial R_m}{\partial t} = \frac{\partial G}{\partial t}(M + M_m) + G\frac{\partial (M + M_m)}{\partial t}$$

In r.h.s. of above equation it can be evidently supposed that the time derivative of G is not time dependent $(\partial G / \partial t = C)$ while the second term is zero. This way

$$\frac{\partial n_m}{\partial t} = -\frac{3}{2}\frac{n_m}{R_m}\frac{\partial R_m}{\partial t} + \frac{\partial G}{\partial t}\frac{M + M_m}{2n_m R_m^3} = \frac{3}{2}\frac{n_m}{R_m}\frac{\partial R_m}{\partial t} + C^*$$

C^* is of course a constant value. Introducing $\partial n_m / \partial t$ into (3)

$$\frac{\partial (C\omega)}{\partial t} = -\frac{1}{2}\frac{MM_m}{M + M_m} n_m R_m \frac{\partial R_m}{\partial t} + \frac{1}{6}\frac{MM_m}{n_m R_m}\frac{\partial G}{\partial t} = L + \frac{1}{3}\frac{MM_m}{M + M} R_m^2 C^* \qquad (4)$$

From astronomical data

$$\frac{\partial (C\omega)}{\partial t} \approx -4.1 \cdot 10^{16} N_m$$

The total tidal torque is composed by the atmospheric (L_{AT}), the earth (L_{ET}) and the oceanic (L_{OT}) tidal torque's:

$$L = L_{AT} + L_{ET} + L_{OT} = 5 \cdot 10^{15} N_m - \left(5 \cdot 10^{15} + 5 \cdot 10^{16}\right) N_m = -5 \cdot 10^{16} N_m$$

Consequently in (4) $\frac{\partial G}{\partial t} \geq 0$ what is in contradiction with the theories on the expanding Universe and Earth, because an increasing gravitational constant requires compression.

The third problem which will be discussed in this study in connection of the Newtonian law is the problem of the numerical value and the accuracy of the gravitational constant. The value of the gravitational constant G is known with much less accuracy than other fundamental constants of physics. Authors of the best determinations of this universal constant claim to their results an accuracy of 10^{-4}, but the following list of the G values obtained by different scientists shows that the disagreement between the individual results is of the order of 10^{-3}.

AUTHORS	YEAR	$G \cdot 10^{11} Nm^2 kg^{-2}$
Rose et al.	1969	6.6699±0.0014
Facy & Poinkis	1970, 1971	6.6714±0.0006
Renner	1974	6.668±0.0002
Sagitov et al.	1978	6.6745±0.0008
Luther & Towler	1982	6.6726±0.0005
De Boer	1987	6.6670±0.0007
Michaelis et al.	1996	6.7154
Schwarz et al.	1998	6.6873±0.0094

Moreover it can be concluded that G is the least known constant of fundamental physics. The following compilation shows the relative errors of basic physics constants:

Avogadro constant	$5.2 \cdot 10^{-10}$
Boltzman's constant	$1.2 \cdot 10^{-4}$
Elementary charge	$2.8 \cdot 10^{-6}$
Faraday constant	$2.8 \cdot 10^{-6}$
Gravitational constant*	$8.5 \cdot 10^{-4}$
Mass of the neutron	$5.1 \cdot 10^{-6}$
Planck's constant	$5.5 \cdot 10^{-6}$
Rydberg's constant	$8.3 \cdot 10^{-6}$
Spead of the light	$4.0 \cdot 10^{-4}$

* The error value of G is the value given by CODATA (Cohen &Taylor, 1986)

There are several explanations why *G* is known with a low accuracy. First of all should be mentioned the weakness of gravitational attraction in scales used in laboratories. For example: a force interaction of two masses of 1g at the distance of 1 cm is 10^{-12} Newton while the pressure of the light of the Sun is 10^{-10} Newton or the forces acting between a proton and a neutron are 10^{-8} Newton. Additionally there is a metrological difficulty: G is defined by the fundamental quantities time, length and mass in absolute scale, what leads of course to experimental difficulties. And finally there is a "psychological problem" too: at this time there are no big research problems in the science, which would urgently need a more accurate value of *G*.

The scatter of the G data listed above suggests that there can be systematic error in gravitational constant values determined in different experiments. The hearth of them - expect the two last ones in the table – is a torsion balance which was used in the beginning in static way and later on – after the successful attempt of Eötvös at the very and of XIX century – dynamically. It was discovered however that the torsion force is dependent on the frequency with which the torsion bar is oscillating. The variability of the elastic constants is particularly significant at low frequencies used in laboratory experiments. According to Maddox (1995) the frequency dependence of the elastic parameters of the materials used in torsion balances is the main source of the systematic and big differences between the laboratory *G* determinations.

In spite of the considerable difficulties it is important to try to increase the accuracy of *G* determinations. It seems that one way can be in this direction the use of the laboratory calibration device developed by us. This experimental tool has a clear geometry and the used quantities (mass of the ring, its position etc.) are already or can be obtained with an accuracy necessary to get *G* with a relative error of 1 part in 10^4 (or even a ten times 10^5). (Varga et al., 1995). To reach this level in our knowledge about the value of the gravitational constant some development of the calibration device in needed.

1. The influence of microseismic noise must be reduced significantly. The systematic beating with a period of some minutes caused by the microseisms characterised with a period of some minutes caused by the microseisms characterised with periods between 5 and 10 s produces gravity variation of about 10 μGal.
2. Because of the need of very accurate determination of the extreme it is necessary to introduce adjustment calculations. This can be the least square method (the L_2 norm) in case of Gaussian error (noise) distribution. In the observations of the gravity during the calibration procedure a number of outliers - possibly due to the long-periodic beating of microseisms – were detected which can be handled with the robust estimations (Somogyi & Závoti, 1993).
3. If the construction of new superconducting gravimeters allows, an effective way to increase the gravity effect is the reduction of the inner diameter of the ring from 30 cm to 20 or to 15. The

corresponding gravity effect generated by the cylindrical ring of mass 3200 kg will be 178 or 236 μGal, instead of 112 μGal.

Of course to get uncertainties of 10^{-4} or even better the spring gravimeters - used until now - must be replaced by new, transportable superconducting gravimeters with reduced diameter which are under development recently at GWR company.
The superconducting gravimeters should be calibrated first along the calibration lines, measured with absolute gravimeters. The accuracy of these calibration lines is 10^{-5} (Atzbacher & Gerstenecker, 1993). Afterwards with the calibrated gravimeters the gravity effect generated by a ring moved up and down must be measured. The gravity effect caused by the ring is known with an accuracy of $\sim 10^{-5}$. If the G value is suitable the measured and the generated gravity values should coincide. With other words: the comparison of these two gravity values allows to determine the G value.

3. Tidal friction, paleogeodesy and the development of the dynamical properties of the Earth in geological time scale

Fossils and tidal deposits as well as the possibility to compute values of the lunisolar tidal torque for different geological epochs allow us to model the variations in time the angular speed, the despinning rate and the time variations of the Earth's figure, assuming that the latter remains, on global scale, close to a hydrostatic equilibrium figure. On this basis we were able to infer the most important kinetic parameters over much of the geological past.
Lambeck (1980) performed a linear regression of paleontological data of Phanerozoic (last (5-6) 10^8 years of geological history) and obtained a constant despinning rate of $-5.4 \cdot 10^{-22}$ rad $\cdot s^{-2}$ similar but somewhat higher rates follows from SLR and LLR ($-5.98 \cdot 10^{-22}$ rad $\cdot s^{-2}$ and $-6.07 \cdot 10^{-22}$ rad $\cdot s^{-2}$). These last data differ to a certain extent obtained from astronomical observations ($-5.6 \cdot 10\text{-}^{22}$ $\cdot$rad s^{-2}) (see e.g. Varga et al., 1998). It was found on the basis of more complete paleontological and sedimentological data sets, that the mean despinning rate was smaller in the Proterozoic than in the Phanerozoic. The linear trend in the variation of length of day (l.o.d.) in the Phanerozoic ca be modelled as

$$LOD = 24.00 - 4.98\tau \tag{5}$$

On the other hand for the linear trend in the Proterozoic it can be suggested tentatively

$$LOD = 21.43 - 0.97\tau \tag{6}$$

Where τ is the time before present (BP) in 10^9 years.

It is clear from (5) and (6) that during the Poterozoic (Ptz) $\left((2.5-0.5)\cdot 10^9 \text{ years BP}\right)$ the despinning rate was five time smaller than during the Phanerozoic (Pz).

The result concerning the low despinning rate in the Proterozoic solves the problem of the Moon having been ever too near to the Earth. But on the other hand significant difference in the despinning reflected in (5) and (6) between Pz and Ptz needs explanation. At least two mechanisms may be invoked, but both of them are liable to be critized . The first involves the idea that the world ocean was less deep two or three billion years ago than it is now, and the shelf lines were shorter in global scale. The second idea is that the formation of the core was not completed entirely soon after the Earth it self was formed intensively up to $5 \cdot 10^8$ years BP. On the basis of tidal friction data it makes sence to estimate the paleogeodetic and geodynamical parameters of the Earth. The next table shows the Earth angular speed (ω), length of day (l.o.d.), geometric flattening (f), dynamic shape factor (J_2) and precession constant (H) in the course of geological history:

Time BP	ω	*l.o.d.*	$10^3 f$	$10^3 J_2$	$10^3 H$
(in 10^6 years)	°/*hr*	Hours			
0	15.00	24.00	3.35	1.08	3.27
50	15.00	23.68	3.44	1.11	3.36
100	15.31	23.50	3.49	1.13	3.41
200	15.33	23.52	3.56	1.13	3.42
300	15.73	22.58	3.69	1.19	3.60
400	16.60	21.69	4.11	1.33	4.00
500	17.25	20.87	4.43	1.43	4.31
800	17.44	20.64	4.53	1.46	4.41
1000	17.84	20.18	4.74	1.53	4.61
1400	18.32	19.65	5.01	1.61	4.87
1600	18.53	19.43	5.12	1.65	4.98
2000	18.74	19.21	5.23	1.69	5.08

These data set renders possible the study of the geodetic – geodynamical development of our planet during its history.

The author would like to express his thanks at this place to Erik W.Grafarend who called his attention to the MacCullagh theorem and to its role in understanding geodynamical phenomena.
In case of hydrostatic equilibrium the external potential of the Earth can be written as

$$V = \frac{GM}{r}\left[1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^{2n-2} J_{2n-2}^{\circ} P_{2n-2}^{\circ}(\sin\Phi)\right]$$

(a is the semimajor axis of the Earth r is the distance, and Φ the latitude) In case of hydrostatic equilibrium J_n° decrease with the increase of n as $f^n \sim \frac{1}{300}$. J_n° values are of the order of 10^{-6} except $J_2^{\circ} = 1.08 \cdot 10^{-3}$. The mentioned above MacCullagh formula based on the first two terms of the r.h.s. of above equations is

$$V = -\frac{GM}{r}\left(1 - \frac{a^2}{r^2} J_2^{\circ}\right) = -\frac{GM}{r}\left(1 - \frac{C - A}{Mr^2}\right) \tag{7}$$

Equation (7) – in which C and A are the polar and equatorial moments of inertia – is one of the important starting points of the study of the dynamics and structure of the Earth.

Because in the scientific literature the MacCullagh formula (7) is, as a rule, without a reference to its author it seems necessary to give some basic information on its discoverer James MacCullagh (1809-1847). He was an Irish mathematician and physicist, had a brilliant carrier at the Trinity College in Dublin and was an elected fellow of the Royal Irish Academy. He held at first the chair of mathematics (1832-43) and made a mathematical center from his university. From 1843 he worked at the chair of natural philosophy. His main field of interest was geometry and optics, published also different remarkable studies in gravimetry and on rotation solid bodies. With the use of (7) the external gravity potential of a rotating body *(U)* for n=0,2 is

$$U = V + W = -\frac{GM}{r} + GMa^2 J_2^{\circ}(\sin\Phi) - \frac{1}{2}\omega^2 r^2 \cos^2\Phi \tag{8}$$

The last term of the r.h.s. is the potential of the centrifugal force, which generates variations in gravity if ω is time dependent due to the despinning of the Earth axial rotation for example.
The second term in r.h.s. of (8) can be expressed as

$$-W = \frac{1}{2}\omega^2 r^2 \cos^2 \Phi = \frac{1}{3}\omega^2 r^2 \left(1 - P_2^\circ(\sin\Phi)\right) \tag{9}$$

Here the last term contributes to the dynamics of the Earth similarly to the second term in the MacCullagh equation

$$V^* = GM\frac{a^2}{r^3} J_2 P_2^\circ(\sin\Phi) \tag{10}$$

The second term of r.h.s. of (9) would be equal to the r.h.s. of (10) if a coefficient of proportionality k is introduced which involves the integrated mechanical properties of the Earth.
Therefore with the use of the r.h.s. of (9)

$$W^* = \frac{k}{3}\omega^2 r^2 P_2^\circ(\sin\Phi) \tag{11}$$

and introducing the Helmert's geodynamical constant

$$m = \frac{\omega^2 a^3}{GM} \tag{12}$$

ω^2 in (11) can be replaced

$$W^* = \frac{km}{3} GM \frac{r^2}{a^3} P_2^\circ(\sin\Phi)$$

If – as it was assumed above $-V^* = W^*$

$$J_2^\circ = \frac{k}{3} m \frac{r^5}{a^5} \tag{13}$$

At the surface of the Earth (13) gives

$$J_2^\circ = \frac{k}{3} m \tag{14}$$

The time derivative of the second zonal geopotential coefficient J_2° can be obtained from (14) with the use of (12) as

$$\frac{dJ_2^\circ}{dt} = \frac{2}{3} k \frac{a^3 \omega}{MG} \frac{d\omega}{dt} \tag{15}$$

On the basis of spindown value valid for the present epoch and for the last $0.5 \cdot 10^9$ years (Pz) with the use of (15)

$$\frac{dJ_2^\circ}{dt} = (-5.12 \pm 0.48) \cdot 10^{-13} \text{ year}^{-1} \tag{16}$$

The time derivative of the dynamic shape factor J_2° has undergone significant variations in the course of geological history. As a consequence of (5) and (6) the variations of the dynamics of our planet is different in the Ptz and the Pz:

$$\begin{aligned}
&\frac{dJ_2^\circ}{dt} = -4.5 \cdot 10^{-13} \text{ year}^{-1} && \text{during the last } 10^9 \text{ years} \\
&\frac{dJ_2^\circ}{dt} = -3.1 \cdot 10^{-13} \text{ year}^{-1} && \text{during the last } 2 \cdot 10^9 \text{ years} \\
&\frac{dJ_2^\circ}{dt} = -1.6 \cdot 10^{-13} \text{ year}^{-1} && \text{in the time-interval } (2.0\text{-}1.0) 10^9 \text{ years BP}
\end{aligned} \tag{17}$$

With (17) the time derivatives of the polar and equatorial momentums of the polar and the equatorial momentums of inertia are

$$\begin{aligned}
&\frac{dC}{dt} = \frac{2}{3} Ma^2 \frac{dJ_2^\circ}{dt} = (-4.1 \pm 0.9) \cdot 10^{25} \text{ kgm}^2 \text{ year}^{-1} \\
&\frac{dA}{dt} = -\frac{1}{3} Ma^2 \frac{dJ_2^\circ}{dt} = (2.1 \pm 0.5) \cdot 10^{25} \text{ kgm}^2 \text{ year}^{-1}
\end{aligned} \tag{18}$$

Values obtained in (17) and (18) are tools for the study of geodynamical processes acting long time (say longer than 10^6-10^7 years). They are expressing changes in the inner structure of the Earth. (16) and (17) apparently contradict to results obtained for the secular changes in J_2 obtained with the laser data of geodetic satellites. The mean of these data is (Varga 1998)

$$-2.7 \cdot 10^{-11} \text{ year}^{-1} \tag{19}$$

yields

$$\begin{aligned}
&\frac{dC}{dt} = -4.2 \cdot 10^{27} \text{ kgm}^2 \text{ year}^{-1} \\
&\frac{dA}{dt} 2.1 \cdot 10^{27} \text{ kgm}^2 \text{ year}^{-1}
\end{aligned} \tag{20}$$

what is evidently too high for long lasting (longer than 10^6-10^7 years) geological processes. If, for example, the question is: when A will be – hypothetically - equal C on the basis of present day data the following relation can be derived from (18)

$$\Delta t = \frac{C - A}{\frac{dJ_2}{dt} Ma^2} = (3 - 5) \cdot 10^7 \text{ years}$$

What means that A will be equal to C in case of (16) within $2.1 \cdot 10^9$ years. Studies of the present glacial discharges show that dJ_2/dt deduced from satellite data can be explained by this phenomenon. As it was shown by Vermeersen et al.(1997), time derivative of J_2 allows us a study of the viscosity profile of the Earth's mantle and the dependence of dJ_2/dt on mantle viscosity. The secular variation of the second degree zonal harmonic has its maximum when the viscosity is about 10^{20} Pa·s in the

upper and 10^{21} Pa·s in the lower mantle. In case of decreasing viscosity, the magnitude of dJ_2/dt gets significantly reduced. This circumstance can be important for the explanation of the difference between satellite (18) and geological (16) values for the time derivative of the second degree component of the geopotential.

4. Conclusion

The above described unsolved problems of geodynamics are subjectively selected. There were not mentioned many still not solved questions. For example the excitation mechanisms of the Chandler wobble not understood yet, the frequency of the core notations is different from observations and from theory possibly due to the use of simplified theoretical model of the Earth. The scientist of XXI century shall solve these questions of course together with many another ones. One of the varantiy for this is the excellent school of theoretical geodetic research founded and led by Professor Erik W. Grafarend at the Stuttgart University.

References

Achilli,V., Baldi, P., Casula, G.., Errani, M., Focari,S:, Guerzoni, M., Palmonari, F., Raguni, G., 1995: A calibration system for superconducting gravimeters. Bulletin Geodetique, 69, 73-80.

Atzbacher, K., Gerstenecker, K., 1993: Secular gravity variations: recent crustal movements or scale factor changes? J.Geodynamics, 18, 1-4,107-121.

Carey, S.W., 1988: Theories of the Earth and Universe. Stanford University Press.

Cohen,E.H., Taylor,B.N., 1986: Adjustment of the physical constants. Codata Bulletin, 62.

Egyed, L., 1957: A new dynamic conception of the internal constitution of the Earth. Geological Rundschau, 46, 101-121.

Hubler, B., Cornaz, A., Kündig, W., 1995: Determination of the gravitational constant with a lake experiment: new constraints for non-Newtonian gravity. Phys.Rev. D., 51, 4005-4016.

Jordan, P., 1966: Die Expansion der Erde. Verlag Viehweg & Sohn.

Lambeck,K., 1980: The Earth's variable rotation (Geophysical Causes and Consequences). Cambridge University Press.

Maddox, J., 1995: Systematic errors in "Big G"? Nature, 377,573.

Schwartz, J.P., Robertson. D.S., Niebauer, T.M., Faller, J.E., 1998: A free – fall determination of the Newtonian constant of Gravity, Science, 282, 5397, 2230-2234.

Somogyi, J., Závoti, J., 1993: Robust estimation with interactively reweighted least square method. Acta Geod. Geoph. Mont. Hung., 28, 2-4, 465-490.

Stacey, F.D., Tuck, G.J., Moore, G.I., Holding, S.C., Goodwiz, G.D., Zhow, R.K., 1987: Geophysics and the law of gravity. Rev. of modern Physics, 59, 1, 157-174.

Varga, P., 1989: Laboratory calibration of gravimeters. Österreichische Beiträge zur Meteorologie und Geophysik, 2, 111-121.

Varga, P., Csapó G., Becker, M., Groten, E., 1991: Laboratory calibration of LCR type gravimeters. XXth General Assembly, IUGG, Vienna.

Varga, P., Hajósy, A., Csapó, G., 1995: Laboratory calibration of LaCoste – Romberg type gravimeters by using a heavy cylindrical ring. Geohys. J.Int., 120, 745-752.

Varga, P., Denis, C., Varga, T., 1998: Tidal friction and its consequence in paleogeodesy, in the gravity field variations and in tectonics, J. Geodynamics, 25, 61-84.

Varga, P., 1998: Temporal variation of second degree geopotential. Reports of the Finnish Geodetic Institute, 4, 125-128.

Vermeersen. L.L.A., Fournier, A., Sabadini, R., 1997: Changes in rotation induced by Pleistocene ice masses with stratified analytical Earth model. J.Geophys. Res., 102, 27689-27702.

Geodetic science and the tools of the trade

Martin Vermeer

Abstract:

When asked to contribute to this Festschrift for my doctoral opponent and old friend Erik W. Grafarend, I ended up writing once again a philosophical little text, in which I consider the roles of mathematics, and of information technology, in geodesy and by extension in science in general.
I believe that science, to remain true to itself, should adhere to *conceptual transparency*, i.e., in everything the scientist does, he/she should strive to understand, and if necessary control, every step of the process that is relevant to producing the final scientific result. One such essential tool is mathematics; and happily, with his physics background, Prof. Grafarend has always been sharply aware of its importance and has wielded the tool with great flair.
In using mathematics as well as information technology, one great danger is "garbage-in, garbage-out". Throughout his professional career Prof. Grafarend has stressed the importance of *understanding what is going on*; posing the "*why?*" question, conceptual clarity, not automated garbage processing. The present contribution is written in this spirit.

1 Science and understanding

One of the things that distinguishes science from other fields of human endeavour, is an almost pathological preoccupation with the question "*why?*", and with ways to answer it. Sometimes this inclination borders on the ridiculous; where for the vast majority of people it is quite sufficient to have a common sense confirmation that something is so or "just works", without any deep understanding of what makes this so, this will never do for a scientist.
The remarkable thing is, that one can actually lead a full and rewarding life without ever asking the *why* question. Nevertheless, the fact that this is so is to a large extent the achievement of a fairly small number of exceptional individuals that knew to ask this question, at the right time, in the right context. All of our science and technology, and our prosperity and good life, is built upon this.
As experience has shown, knowing what you're doing, understanding what you're doing and why, can be worth gold and is a hard requirement if you want to do something properly. Industry and government have learned this lesson; that's why in all developed countries, research and development are rightly considered activities vital to society.
As scientists, we have a duty to remain loyal to this core element of our tradition. Understanding matters.
In science, we can distinguish three stages of activity:

1. Observation, measurement, data collection
2. Data processing, analysis, theory building
3. Dissemination of results, processed data, publication.

The requirement of knowing what you are doing, of *conceptual transparency*, should be applied to all three of these stages. A scientist should have a fundamental understanding and working knowledge of all the tools of his or her trade.

2 The role of mathematics

One important tool in the exact sciences, of course, is mathematics; mathematically formulating a scientific finding or argument gives it an authoritativeness that is much envied by practitioners of those fields of scientific study where it is inappropriate, or not feasible, to use it. But where it is both appropriate and feasible, the use of the language of mathematics conveys a tractability to the argument which allows anyone with a command of this tool to absorb the argument on a level of detail and exactness that cannot be achieved without it. And surprisingly but fortunately, the physical universe seems to be very friendly to mathematical description.

2.1 Scaling laws in biology

As an illustration of the use of mathematics in science, I want to refer to a presentation given by Geoffrey West at the recent 22nd Conference on Mathematical Geophysics in Cambridge UK [6]. He presented his ideas on the scaling laws of river systems, starting from the scaling laws that he had found in biology, e.g., that the metabolic rate is proportional to body weight to the power 3/4 for a very wide range of organisms. According to his studies, the physical mechanism producing this law is simply the supply of some vital resource (food, oxygen, river water) to/from all the nooks and crannies of the system sustained, dissipating minimum energy in transport.

2.2 Scaling laws for GPS network densifications

I decided to try if it would be possible to similarly model the propagation of the resource "co-ordinate precision" through a cascade of geodetic GPS network densifications to the end points established, using minimal total station occupation time.
We cover an area S with GPS stations in a hierarchical arrangement; the hierarchy level we call ω, $\omega = 1, 2, \ldots, \Omega$. Then, when the "branching number" for a given level ω is n_ω, starting from a single level-zero station, the total number of stations is:

$$N = \prod_{\omega=1}^{\Omega} n_\omega .$$

The mean distance between stations up to a level ω, assuming, e.g., a rectangular arrangement of stations and leaving off any irrelevant constant factors:

$$l_\omega = \sqrt{S} \left[\prod_{\omega'=1}^{\omega} n_{\omega'} \right]^{-\frac{1}{2}} .$$

We postulate the measurement variance to depend on both baseline length and occupation time in the following way:

$$\sigma_\omega^2 = C\, l_\omega^p\, t_\omega^q ,$$

where t_ω is occupation time and C a proportionality constant. For the assumption of temporally uncorrelated noise, $q = -1$. For comparison, is known that for levelling networks, $p = +1$, which might also apply to GPS networks (Bernese rule-of-thumb, [3]).

We compute the total measurement variance (assuming uncorrelated hierarchy levels ω) as

$$\sigma^2 = \sum_{\omega=1}^{\Omega} \sigma_\omega^2 = \sum_{\omega=1}^{\Omega} C\, l_\omega^p\, t_\omega^q =$$
$$= \sum_{\omega=1}^{\Omega} CS^{\frac{p}{2}} \left[\prod_{\omega'=1}^{\omega} n_{\omega'}\right]^{-\frac{p}{2}} t_\omega^q . \tag{1}$$

Now, defining

$$K \equiv CS^{\frac{p}{2}}$$

and

$$N_\omega \equiv \prod_{\omega'=1}^{\omega} n_{\omega'} ,$$

we may write for the total measurement receiver time:

$$\tau = \sum_{\omega=1}^{\Omega} \left[\prod_{\omega'=1}^{\omega} n_{\omega'}\right] t_\omega = \sum_{\omega=1}^{\Omega} \tau_\omega , \tag{2}$$

with

$$\tau_\omega \equiv \left[\prod_{\omega'=1}^{\omega} n_{\omega'}\right] t_\omega = N_\omega\, t_\omega .$$

Substitute this into Eq. (1):

$$\sigma^2 = \sum_{\omega=1}^{\Omega} K\, N_\omega^{-\frac{p}{2}} \cdot N_\omega^{-q} \tau_\omega^q =$$
$$= \sum_{\omega=1}^{\Omega} K N_\omega^{-\frac{p}{2}-q} \cdot \tau_\omega^q ,$$

Now *optimize*, i.e., minimize, σ^2, given τ by Eq. (2). Lagrange multipliers: find the point where the gradient of $\sigma^2\,(\tau_1, \tau_2, \ldots, \tau_\Omega)$ stands perpendicular to the surface $\sum \tau_\omega = \tau$. I.e., demand:

$$\frac{\partial\,(\sigma^2)}{\partial \tau_\omega} = qK N_\omega^{-\left(\frac{p}{2}+q\right)} \tau_\omega^{q-1} = \lambda, \quad \omega = 1, 2, \ldots, \Omega.$$

It follows that

$$\tau_\omega^{q-1} = \frac{\lambda}{qK} N_\omega^{\left(\frac{p}{2}+q\right)} \quad \Rightarrow \quad \tau_\omega = \kappa N_\omega^{\left(\frac{\frac{p}{2}+q}{q-1}\right)} ,$$

where

$$\kappa \equiv \left[\frac{\lambda}{qK}\right]^{\frac{1}{q-1}}$$

is another constant. So:

$$\tau_\omega = \kappa N_\omega^{\frac{\frac{p}{2}+q}{q-1}} , \tag{3}$$

with the above definition of N_ω, and

$$t_\omega = N_\omega^{-1} \tau_\omega = \kappa N_\omega^{\frac{\frac{p}{2}+q}{q-1}-1} . \tag{4}$$

τ_ω is the *total instrument-time* of a certain stage ω in the network densification. t_ω is the duration of *one measurement* at stage ω. N_ω is the total number of stations up to stage ω.

2.3 Example

Let $p = +1$, $q = -1$, $n_\omega = 16$, so $N_\omega = 16^\omega$. Then, using Eqs. (3) and (4):

$$\frac{\frac{p}{2}+q}{q-1} = \frac{-\frac{1}{2}}{-2} = \frac{1}{4}$$

$$\frac{\frac{p}{2}+q}{q-1} - 1 = \frac{-\frac{1}{2}}{-2} - 1 = -\frac{3}{4}.$$

And we find:

ω	$\tau_\omega = \kappa \cdot 16^{\omega/4}$	$t_\omega = \kappa \cdot 16^{-3\omega/4}$
1	2κ	$\kappa/8$
2	4κ	$\kappa/64$
3	8κ	$\kappa/512$
4	16κ	$\kappa/4096$
$\tau =$	30κ	

Applying this to the Finnish permanent GPS network FinnRef^{TM}, and the Finnish EUREF densification campaigns, yields:

Level	ω	N_ω	n_ω	$\tau_\omega^{\text{theor}}$	t_ω^{theor}	t_ω^{pract}	l_ω
Permanent network	1	12	12	1.8κ	0.15κ	"∞"	200 km
EUREF densification	2	100	8	3.0κ	0.03κ	48^{h}	70 km
"Project 400"	3	400	4	4.5κ	0.01κ	4^{h}	35 km
Total receiver-time				9.3κ			

We see that in practice, the time spent on the second densification phase (Project 400) was quite a bit shorter than theory requires. This was for practical reasons: limited resources available and the time constraints of moving from point to point within a working day.
Of course also, the above assumed p and q values are probably not quite realistic. More likely, p lies between $+1$ and $+2$, while q is almost certainly larger than -1 in the occupation time range considered.

2.4 Discussion

The scaling law derived above under severely simplifying assumptions of course formulates in an unconventional way existing, conventional wisdom that geodesists have always been aware of, occasionally formulated explicitly, and the validity of which which the GPS age has in no way diminished:

> A network should be built up hierarchically, moving from the large to the small scale — the "from large to small" paradigm.

This has relevance for the discussion on the possibility of short-circuiting the hierarchical process by providing precise geodetic positions without the intermediation of network layers, even in real time by the RTK (real time kinematic) positioning technique. In some countries, nation wide services are being set up based on the existing geodetic permanent GPS infrastructures.
It would appear from the above that the logistics of such a set-up may be unnecessarily heavy and better – both optimally accurate and affordable – results are almost always guaranteed using the simplest, most cost effective approach of setting up your own local reference receiver and radio link within the area of study.
Our Finnish EUREF densification project will provide a good starting point for this sensible practice.

3 The role of information technology

And now for something entirely different! We recall our division of the scientific process into three phases, cf. Section 1. Modern information technology has relevance to all three of these phases. In the following, we will shortly consider the issue mainly from the point of view of GPS geodesy, with which the author is most familiar.

3.1 Data collection

Data is collected by GPS receivers in the field, permanently mounted in fixed locations, and transmitted to a data collection centre at regular intervals. In Finland, the Finnish Geodetic Institute operates twelve such receivers, which send their collected data at 24 hour intervals by modem and dial-up phone line to the data centre in Masala, close to Helsinki. This permanent network is multi-purpose, serving, e.g., the study of Earth crustal motions, and providing a basis for the establishment and maintenance of a precise national mapping datum.
Many other countries around the world have such networks, which only differ in the technical details of their operation.
These data collection systems, which operate largely autonomously, have of course to be *reliable*. Running this kind of service reliably is a seriously nontrivial operation. From our own experience, it is highly desirable to be in control of all the elements of the system. The fact that in our current system most of the software running it, including the operating system (!) of the download server, is closed commercial software of which we do not have the source code – a shortage of *conceptual transparency* –, has on occasion complicated our task.

3.2 Analysis

For scientific analysis of GPS data, in case one is interested in the highest precision and in results that hold water scientifically, one should preferably use GPS processing software that is documented by its source code. There are several such packages circulating in the GPS geodetic community; we use the Bernese software package, produced by the Astronomical Institute of the University of Berne. This kind of software is typically written by users who are themselves scientists and publish in the scientific literature.
Having the source is important even if one does not have the time or the inclination to read it. Some people do, and any discrepancy between documented and actual behaviour does not remain invisible for long. As an illustration, one of our researchers (Matti Ollikainen, personal comm.) found a bug in the tropospheric mapping function code.
Of course using commercial processing software, which hides the true complexity of GPS processing from the user, takes its own, presumably intelligent, policy decisions and offers superior ease of use, is a good solution in many production situations. But it isn't necessarily the best if you want your scientific results to meet certain standards of tractability.

3.3 Visualization, presentation and publication

It should not be forgotten that also publishing results, for colleagues and possibly the public at large to read, is an important part of the scientific process. Valid results are significantly more useful if published well, and may even go unread if published badly.
Besides the issues of producing well written language and of publishing in the proper place, there is the important issue of the format of publication.
One thing information technology has done is make publication in electronic form not just possible, but easy. As a minimum one can post abstracts of paper publications on the World Wide Web. This circumstance rewards the use of authoring systems that allow for export of documents alternatively in Web or paper print form, or for easy conversion between formats. Besides

the up-and-coming XML format, the traditionally used scientific authoring tools fortunately do these things well, better in fact than currently popular word processors.

3.3.1 Web publication

Information technology enables publishing on the World Wide Web using hypertext mark-up language. As readers familiar with the www know, there is a vast amount of poor format practice going on there. Web sites that are chaotically organized, cluttered, loud, containing hundreds of graphics where text would do fine, not offering textual alternative representations, using tables for formatting non-tabular material, and, last but not least, non-standard, browser-dependent features. Such sites are a nightmare to access with a plain text browser (Lynx) or, e.g., an auditory browser used by people of poor eyesight (cf. [2]).
The principles for a good scientific Web publication are no different from those for a good general Web information page: *don't do* any of the above. Much of the messiness of many www pages appears to be due to the tools used to create them, usually visual mark-up editors that give you full control of what the page will look like (well, in Browser X anyway), but *not of what goes into it.* While visually interactive (WYSIWYG) editors have their place in website design, they all too often illustrate the saying "What You See Is *All* You Get".
It is wise to learn a little HTML syntax, look at the pages produced by your favourite package in a text editor, and clean them up if necessary. And keep it simple. And in the near future, consider XML, the Extensible Mark-up Language, which Gecko, Netscape's new browser engine, will render in a standards-compliant way. Gecko exists also for Internet Explorer, as a plug-in. XML and its document type definitions are a superset of HTML, offering much improved facilities for producing content-structured documents.

3.3.2 Paper publication

Information technology has, perhaps unexpectedly, great relevance for paper publication as well. Excellent tools existed already a decade before personal computers became ubiquitous and word processing software useable by untrained personnel permanently lowered our quality expectations from computer typesetting, cf. [5]. A good looking paper publication is a joy to the eye and easy to read and the likelihood of the message coming across is greatest if the visual structure supports the logic of discourse. And of course, errors tend to distract and irritate the reader, planting doubts as to the reliability of the writer as well.
To produce a good looking paper, one uses a professional typesetting software system with well designed fonts. To easily produce a well structured, error free document having unity of form and function in the visual layout, one uses a professional document processing system. Both feature sets are offered by the LaTeX software [4]. True, it requires the discipline of learning its use, but – we scientists are supposed to be good at learning new things, it comes with the territory. And nowadays there are excellent, easy to use visual word processing tools available for LaTeX [1], removing the learning treshold and facilitating interoperation with non-scientists.
To produce a correct reference listing, one uses BibTeX, which extracts references used from a bibliographic database. There is no excuse for scientists not to use BibTeX; if the database is error free, the reference list will be too, guaranteed. And again, excellent visual tools exist for managing this database.

3.4 Real time systems

One more effect of IT on the scientific process may be the "telescoping" of the three phases listed in Section 1 into one continuous process. Data is collected, analysed in real time, and the results of the analysis visualised and presented to the end user also in real time.
Such systems are already in use in some places; GPS data collected and analysed in real time allows the presentation of animations of, e.g. crustal motion, ionospheric total electron contents,

or troposheric total water vapour contents on a publically accessible web page. We must expect such practices to become more widespread, as the underlying technology becomes more affordable, reliable and easy to use. But it is important to realize that the requirement of knowing what you're doing, of conceptual transparency, continues to apply to all the elements of such a system. There is no royal road to scientific knowledge, and "data" is not "understanding".

4 Conclusions

We leave those as an exercise for the reader.

References

[1] (Anon.). LyX, The Document Processor. Web page, accessed May 1999. URL: `http://www.lyx.org`.

[2] (Anon.). Web Accessibility Initiative. Web page, World Wide Web Consortium, accessed May 1999. URL: `http://www.w3.org/WAI`.

[3] G. Beutler, I. Bauersima, S. Botton, C. Boucher, W. Gurtner, M. Rothacher, and T. Schildknecht. Accuracy and biases in the geodetic application of the global positioning system. *Manuscripta geodaetica*, 14(1):28–35, 1989.

[4] Leslie Lamport. *LaTeX, a document preparation system.* Addison Wesley Longman, Reading, Massachusetts, 1994. Second edition.

[5] J. Parsons. Creating an Archaeological Book with LaTeX. Web page, accessed May 1999. URL:

`http://www.ocf.berkeley.edu/~jparsons/hysplex/hx_howto.html`.

[6] Geoffrey B. West. The Origin of Scaling Laws in River Networks. In Herbert Huppert et al., editors, *22nd International Conference on Mathematical Geophysics*, volume 3:1 of *Journal of Conference Abstracts*, page 67. Cambridge Publications, Cambridge UK, July 1998.

Fortgeschrittene geodätische Deformationsanalyse

Walter Welsch

Hintergrund

Nach klassischem Verständnis ist der Gegenstand geodätischer Überwachungsmessungen die Erfassung geometrischer Veränderungen eines Meßobjekts. Unter geometrischen Veränderungen werden Bewegungen und Verformungen des Überwachungsobjektes verstanden. Bedenkt man jedoch, daß zu den Zielsetzungen geodätischer Überwachungsmessungen u.a. auch die Möglichkeit der Prognose des mutmaßlichen Objektverhaltens in der näheren Zukunft und des Verhaltens unter bestimmten Lastfällen gehört, greift die klassische Definition zu kurz. Die Prognose eines noch nicht eingetretenen Verhaltens setzt voraus, daß gewisse Gesetzmäßigkeiten der Objektreaktion in den Analyseprozeß miteinbezogen werden müssen. Diese Gesetzmäßigkeiten betreffen zum einen die Einflußgrößen, die auf das Objekt wirken, und zum anderen die Auswirkung der Einflußgrößen, die sich infolge der Struktur des Objekts auf seine geometrischen Eigenschaften ergibt. Folgt man dieser erweiterten Definition der geodätischen Deformationsanalyse, so bedeutet diese die Beobachtung und die quantitative, zumindest aber qualitative Bewertung des gesamten Deformationsprozesses. Man kann auch sagen, geodätische Deformationsanalyse bedeutet die geodätische Analyse eines Deformations- oder allgemeiner eines dynamischen Prozesses.

Hierzu sind besondere Modelle und Analysemethoden bereitzustellen. Sie werden im folgenden in der gebotenen Kürze besprochen.

Klassische Deformationsanalyse

Deformation:
Bewegung und Verformung eines Objektes

Abb. 1: Konventionelle Deformationsanalyse im „Deformationsraum"

Die klassische Deformationsanalyse untersucht die Bewegung und Verformung eines Objekts in Raum und Zeit. Ihre Bestandteile sind die Ermittlung der Deformation durch Beobachtung und die Auswertung der Beobachtungen.

An das Meßkonzept werden räumliche und zeitliche Forderungen gestellt. Das Kontinuum des zu untersuchenden Objekts ist in Einzelpunkte aufzulösen (zu diskretisieren) so, daß die Bewegungen der Punkte die Bewegungen und Verformungen des Objektes zutreffend (repräsentativ) beschreiben (charakteristische Punkte). Die Beobachtung oder Abtastung dieser Punkte ist so zu terminieren, daß auch zeitlich nicht-lineare Verformungen erfaßt werden können. Beide Forderungen an das Meßkonzept verlangen zumindest qualitative à priori Informationen über das Objektverhalten.

Die Auswertung der Beobachtungen führt zu rein deskriptiven Aussagen über die Punktbewegungen in Raum und Zeit auch dann, wenn die Bewegungen mathematisch-statistisch analysiert und modelliert werden, um das Verformungsverhalten des Objektes als Ganzes aufzuzeigen. Die Analyse in der Zeit schließt auch das Verhalten im Frequenzbereich ein.

Die klassische Deformationsanalyse verläßt bei all ihren Untersuchungen nicht den "Deformationsraum" des untersuchten Objekts, den sie phänomenologisch beschreibt (WELSCH 1981).

Möglichkeiten und Ziele zeitgemäßer geodätischer Überwachungsmessungen

Dem heutigen Selbstverständnis der Geodäsie genügt es jedoch nicht mehr, die Deformation eines Körpers lediglich zu beschreiben. Die Ergebnisse der Deformationsuntersuchungen sollen vielmehr in einen größeren Zusammenhang eingebettet werden, da sie nicht Selbstzweck, sondern ein Werkzeug zur Beobachtung und Analyse natur- und ingenieurwissenschaftlicher Phänomene sind (WELSCH et al. 1999).

Die Deformation eines Körpers ist das Ergebnis eines Prozesses, eines „Spiels der Kräfte". Es gilt, dies Spiel der Kräfte, die Dynamik des Prozesses, zu analysieren. Grundlage der Untersuchungen sind neben den ermittelten Verformungen die beeinflussenden Kräfte und bekannte oder zu ermittelnde Eigenschaften und Gesetzmäßigkeiten des untersuchten Objekts. Die Geodäsie, vornehmlich die Ingenieurvermessung, ist von der zeitlich-räumlichen Beschreibung der Bewegungen und Verformungen eines Objekts zur Erforschung der Prozesse, denen der Körper unterworfen ist, fortgeschritten (CHRZANOWSKI und CHEN 1986, 1990). Die Aufgabe erfordert eine fachübergreifende Integration von Methoden, die geeignet sind, „dynamische Systeme" zu beschreiben.

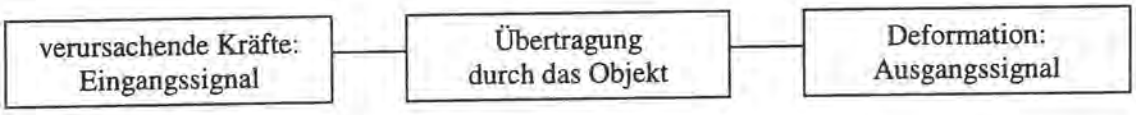


Abb. 2: Deformation als ein Element eines dynamischen Prozesses

Recherchiert man in diesem Sinne die einschlägige Literatur der vergangenen Jahre, so kristallisieren sich vier Gruppen von Analysemodellen heraus, die geeignet sind, in fortschreitender Komplexität die Aufgaben geodätischer Deformationsuntersuchungen zu lösen:

Deformations-modell	Kongruenz-modell	kinematisches Modell	statisches Modell	dynamisches Modell
Faktor Zeit	keine Modellierung	Deformationen als Funktion der Zeit	keine Modellierung	Deformationen als Funktion der Zeit und der Belastung
verursachende Kräfte	keine Modellierung	keine Modellierung	Deformation als Funktion der Belastung	
Zustand des Objekts	hinreichend in Ruhe	permanent in Bewegung	unter Belastung hinreichend in Ruhe	permanent in Bewegung

Abb. 3: Klassifizierung von Deformationsmodellen

Das Ziel der Überwachung ingenieur- und geotechnischer Phänomene ist also nicht nur die Beantwortung der Frage nach den geometrischen Veränderungen, die die untersuchten Objekte erfahren haben, sondern auch die Klärung der Ursachen, die dazu geführt haben, und schließlich - aufbauend auf der Kenntnis des Übertragungsverhaltens - die Prognosebildung.

Charakterisierung der Modelle

Die Charakterisierung der Auswertemodelle von Überwachungsmessungen und ihre Systematisierung sind in der jüngeren Vergangenheit wiederholt Gegenstand von Untersuchungen gewesen (PFEUFER 1993; HEUNECKE 1995; WELSCH 1996; JAEGER et al. 1997; HEUNECKE, PELZER 1998; HEUNE-

CKE et al. 1998; MILEV, PAPO 1998; LEVENHAGEN 1998). Da zu überwachende Meßobjekte als dynamische Systeme betrachtet werden können, bietet es sich an, sich an der Vorgehensweise der Systemtheorie zu orientieren. Hauptaufgabe der Systemtheorie ist es gerade, für die in der Realität existierenden Systeme Modelle bereitzustellen, die die Eigenschaften des Systems entsprechend den Zielsetzungen der Betrachtung idealisieren, repräsentieren und analysieren, um so zu Erkenntnissen über das Systemverhalten zu gelangen und quantitative Ergebnisse zu erzielen (Systemidentifikation).

Dynamische Systeme im allgemeinen Sinne der Systemtheorie sind Systeme (Objekte), die Energie speichern und zeitverzögert abgeben können, wobei es bei der Auswertung nur möglich ist, die Vergangenheit (das Gedächtnis) bis zu einem festzulegenden Zeitpunkt zu berücksichtigen. Ein Spezialfall liegt vor, wenn das System (im Modell) verzögerungsfrei (gedächtnislos) in einen neuen Gleichgewichtszustand übergeht. Die Ausgangsgröße ist dann nur von Eingangsgrößen zum selben Zeitpunkt, eventuell unter vorhergehender Abspaltung einer Verzögerungszeit, abhängig. Gemeinhin wird dann von einem statischen oder gedächtnislosen System gesprochen.

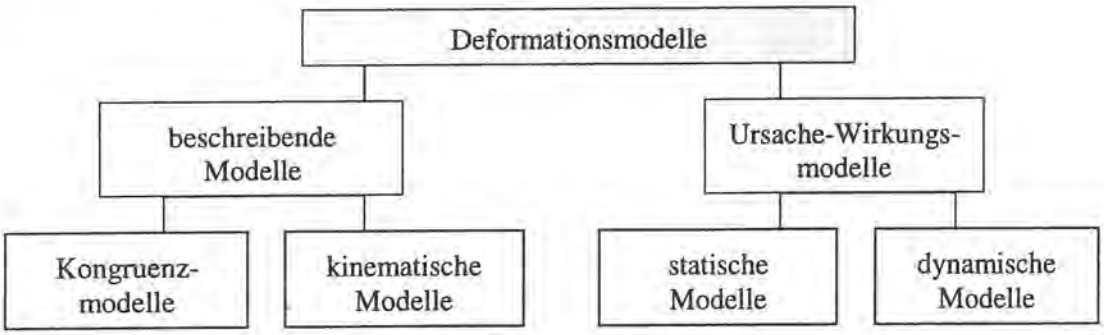


Abb. 4: Hierarchie der Modelle zur geodätischen Deformationsanalyse

Beide, dynamische und statische Systeme, verlangen das Modellieren der deterministisch wirkenden Eingangsgrößen, der sog. Stellgrößen. Systeme, die ohne Betrachtung der auf sie wirkenden Eingangsgrößen untersucht werden sollen, sind weitere Spezialfälle. Systeme, deren Reaktion als reine Funktion der Zeit dargestellt werden, sind kinematische Systeme. Ein System, für dessen Größen kein Bewegungsgesetz aufgestellt werden kann, wird als „random walk"-System bezeichnet; in der geodätischen Deformationsanalyse wird es im Kongruenzmodell beschrieben.

In dieser Kategorisierung der Abbildung 4 kommt die Dreiteilung von Bewegungszuständen in der Mechanik zum Ausdruck:

- Kinematik: Beschreibung von Bewegungsvorgängen ohne Berücksichtigung ihrer Entstehung
- Statik: Lehre von den Gleichgewichtsbedingungen, unter denen sich die unter Einwirkung von Kräften stehenden Körper in Ruhe befinden
- Dynamik: Betrachtung von Bewegungsvorgängen in Zusammenhang mit den sie verursachenden Kräften.

Die wesentlichen Inhalte und Aufgaben der Modelle für die Auswertung von Überwachungsmessungen lassen sich demnach wie folgt zusammenstellen:

- Kongruenzmodelle: Gegenstand der Betrachtung ist der rein geometrische Vergleich des Zustandes eines Objekts (im Extremfall einer einzelnen Koordinate) zu einem Zeitpunkt mit demjenigen eines anderen Zeitpunktes. Neben der Untersuchung der Kongruenz oder Identität der Geometrie sind darüber hinaus zusätzliche Betrachtungen zur Generalisierung aufgetretener Verformungen durch affinen Abbildung des Meßobjektes (Strainanalyse) oder auch durch Polynomansätze häufig Ge-

genstand dieser Modellklasse. Die Generalisierung leitet von einer zunächst punktweisen Untersuchung zur einer kontinuierlichen Betrachtung über.

- Kinematische Modelle: Zweck ist die rein zeitabhängige Beschreibung des Verhaltens von Objektpunkten, insbesondere durch Polynomansätze oder trigonometrische Funktionen, wobei deren Anwendung bereits die Kenntnis gewisser theoretischer Zusammenhänge bedingt. Ziel ist es letztlich, von den Messungen zu bestimmten diskreten Zeitpunkten auf die Objektbewegung und ihre Parameter im allgemeinen zu schließen (Zeitreihenanalyse; KUHLMANN 1996). Ein (quantitativer) Zusammenhang mit ursächlichen Belastungen wird nicht hergestellt, die Analyse des Bewegungsverhaltens erfolgt deskriptiv. Auch hier stellt sich das Generalisierungsproblem, wenn ausgehend von den zunächst einzeln betrachteten Objektpunkten das Meßobjekt als Ganzes über der Zeitachse zu betrachten ist.
- Statische Modelle: Ein statisches Auswertemodell beschreibt den funktionalen Zusammenhang zwischen der Beanspruchung eines Meßobjektes und seiner Reaktion (KERSTING 1992). Zum Zeitpunkt der Messungen muß sich das Objekt hinreichend in Ruhe befinden, so daß die Zeit im Modell nicht explizit zu berücksichtigen ist.
- Dynamische Modelle: Gegenstand des dynamischen Auswertemodells ist die Betrachtung von Objektreaktionen als Funktion der Zeit und der Beanspruchung.

Die zur Verfügung stehenden Methoden und Modelle zur Auswertung von Überwachungsmessungen umfassen ein weites Spektrum. Bei der Bearbeitung einer konkreten Aufgabenstellung kommt es darauf an, die geeigneten „Werkzeuge" auszuwählen und einzusetzen. In Kombination mit der der Problemstellung adäquaten Meßtechnik ist die Auswahl bereits ein wesentlicher Bestandteil der Analyse. Dabei sollte der Grundsatz gelten, einfache Fragen mit einfachen Vorgehensweisen zu lösen.

Strukturmodelle und Verhaltensmodelle

Bei der Analyse der Kausalkette eines dynamischen Systems sind zwei Vorgehensweisen zu unterscheiden, die auf die Begriffe der Struktur- und Verhaltensmodelle beziehungsweise der parametrischen und nichtparametrischen Modelle führen. Der Begriff Struktur beziehungsweise Strukturmodell steht hier für den geordneten, räumlich und physikalisch definierten inneren Aufbau eines Systems. Statische Modelle sind immer Strukturmodelle.

Während bei Strukturmodellen das Übertragungsverhalten in physikalisch interpretierbarer Form vorliegt, wird ein Zusammenhang zwischen einer Eingangsgröße und einer Ausgangsgröße in Verhaltensmodellen rein mathematisch über Regressions- oder Korrelationsbetrachtungen hergestellt. Aussagekräftiger sind zweifelsfrei Strukturmodelle, also Modelle, die im Sinne eines „model approach" von theoretischen Vorstellungen über das System ausgehen und diese über Messungen verifizieren.

Wird bei der parametrischen Systemidentifikation lediglich die Zeitabhängigkeit des Prozesses, dem das System unterliegt, bedacht, nicht jedoch dessen Ortsabhängigkeit, wird das System durch "konzentrierte" Parameter beschrieben. Es genügen gewöhnliche Differentialgleichungen.

Die Differentialgleichung der linearen Elastodynamik ist die Grundgleichung eines durch konzentrierte Parameter zu beschreibenden dynamischen Modells ("white box" - Modell):

$$\left| \boldsymbol{K} \quad \boldsymbol{D} \quad \boldsymbol{M} \right| \left| \begin{matrix} \boldsymbol{x}\ (t) \\ \dot{\boldsymbol{x}}\ (t) \\ \ddot{\boldsymbol{x}}\ (t) \end{matrix} \right| = \boldsymbol{y}\ (t)$$

$\boldsymbol{y}(t)$ ist hierin der Systemeingang, $\boldsymbol{x}(t)$ mit seinen Ableitungen ist der (geodätisch zu beobachtende) Systemausgang; in den Matrizen $\boldsymbol{K}$, $\boldsymbol{D}$ und $\boldsymbol{M}$ sind z.B. im Falle einer Problemstellung aus dem Gebiet

der Mechanik Material- bzw. Entwurfsparameter für Steifigkeit, Dämpfung und Masse eines Bauwerks enthalten. Je nach Aufgabenstellung können einzelne Parameter- und Beobachtungsgruppen entfallen. Bei der Untersuchung von Eigenschwingungen entfällt etwa die Dämpfungsmatrix, bei langsamen Verformungen kann die Masse außer acht gelassen werden.

Der Spezialfall

$$\boldsymbol{K}\, \boldsymbol{x}\,(t) = \boldsymbol{y}\,(t)$$

ist erheblich, da er die grundlegende Beziehung eines statischen Systems beschreibt, das nach Aufbringen einer Last in einen neuen Gleichgewichtszustand mit $\boldsymbol{y}(t)$ = const. übergeht.

Mit

$$\boldsymbol{x}\,(t) = \text{const.}$$

ist als Trivialform auch das Identitätsmodell - etwa zur Überprüfung der Kongruenz eines geodätischen Netzes - im allgemeinen Ansatz enthalten.

Im Zusammenhang mit Strukturmodellen ist es von wesentlicher Bedeutung, daß zur Beschreibung von Deformationsvorgängen im Regelfall Koordinatensysteme als Bezugssysteme eingeführt werden. In den meisten Fällen dienen Koordinaten als Zwischengrößen der Auswertung. Zwischengrößen dieser Art werden in der Systemtheorie als Zustandsgrößen bezeichnet, der Unbekannten- oder Parametervektor demzufolge als Zustandsvektor. Außer Koordinaten können bei der Auswertung von Überwachungsmessungen auch weitere Zustandsgrößen auftreten. Die Zustandsgrößen dienen im systemtheoretischen Sinne der Beschreibung der „inneren Zusammenhänge des Systemverhaltens". Sie sind geometrisch und physikalisch interpretierbare Größen und bilden den Zustandsraum. Die Analyse eines Strukturmodells setzt folglich auf der Zustandsraummethodik auf.

Wird bei der parametrischen Systemidentifikation neben der Zeitabhängigkeit auch eine Ortsabhängigkeit des Prozesses betrachtet, so ist das System durch verteilte Parameter zu beschreiben. Dies führt zu partiellen Differentialgleichungen. Durch Verfahren der Ortsdiskretisierung, die auf bereichsweisen Ansätzen aufbaut, wird es möglich, die partiellen durch Differentialgleichungen, die nur in einem beschränkten Definitionsbereich gültig sind, auszudrücken. Die für die einzelnen Definitionsbereiche gefundenen Lösungen sind unter Beachtung von Randbedingungen aneinanderzufügen und ergeben die genäherte Lösung der ursprünglichen Differentialgleichungen. Ein numerisches Verfahren für bereichsweise Ansätze, das sich für die Untersuchung ebener und räumlich ausgedehnter Objekte anbietet, ist die Methode der finiten Elemente (HEUNECKE 1996).

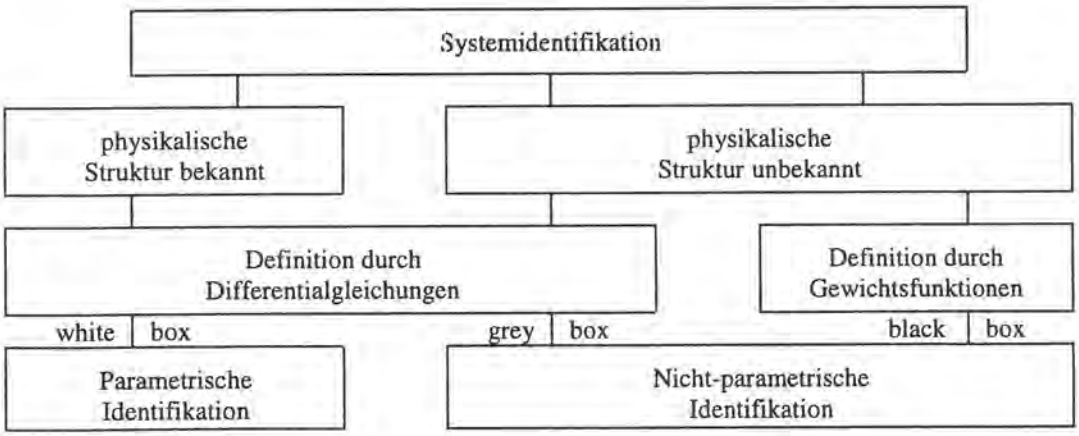


Abb. 5: Methoden der Systemidentifikation

Für viele praktische Problemstellungen ist festzuhalten, daß eine Formulierung physikalischer Zusammenhänge entweder nicht oder nur mit verhältnismäßig großem Aufwand möglich ist. In diesen Fällen sind Verhaltensmodelle eine häufig eingesetzte Alternative, bei denen im Sinne eines „operational approach" allein aus der Beobachtung von Zeitreihen der Ein- und Ausgangsgrößen Erkenntnisse über das System abgeleitet werden sollen. An Stelle von Differentialgleichungen wird im linearen kontinuierlichen Fall auf eine Integralbeziehung in Form des eines Faltungsintegrals zurückgegriffen („black box"-Modell). Für den eindimensionalen Fall kann geschrieben werden (STROBEL 1975):

$$x(t) = \int_0^\infty g(\tau)\, y(t-\tau)\, d\tau .$$

Im Faltungsintegral ist $g(\tau)$ die sog. Gewichtsfunktion, deren zu ermittelnde Koeffizienten beschreiben, mit welchem Gewicht sich die Eingangsgrößen $y(t)$ auf die Ausgangsgröße $x(t)$ auswirken. Im nicht-linearen Fall kann das sog. VOLTERRA-Modell (WERNSTEDT 1989; PFEUFER 1990, 1993) verwendet werden:

$$x(t) = \int_0^\infty g_1(\tau_1)\, y(t-\tau_1)\, d\tau_1 + \int_0^\infty \int_0^\infty g_2(\tau_1\tau_2)\, y(t-\tau_1)\, y(t-\tau_2)\, d\tau_1\, d\tau_2 + \text{Glieder höherer Ordnung.}$$

Zur Herstellung einer Beziehung zwischen einer Eingangs- und einer Ausgangsgröße kann auch ein mathematisches Modell der allgemeinen Form

$$a_q \frac{d^q x}{dt^q} + a_{q-1} \frac{d^{q-1} x}{dt^{q-1}} + \ldots + a_1 \frac{dx}{dt} + a_0 x = b_p \frac{d^p y}{dt^p} + b_{p-1} \frac{d^{p-1} y}{dt^{p-1}} + \ldots + b_1 \frac{dy}{dt} + b_0 y$$

oder die Differenzengleichung

$$x_k = a_1 x_{k-1} + a_2 x_{k-2} + \ldots + a_q x_{k-q} + b_0 y_k + b_1 y_{k-1} + \ldots + b_p y_{k-p}$$

verwendet werden.
Es handelt sich um das sog. ARMA (autoregressive moving average)-Modell. Typisch ist, daß mit $q>3$ und $p>0$ die physikalische Interpretation nicht mehr unmittelbar gegeben ist. Für $q \leq 3$ und $p = 0$ ist jedoch eine physikalische Deutung der Parameter möglich. Man spricht dann von „grey box"-Modellen.
Das ARMA-Modell besteht aus einem rekursiven und einem nicht-rekursiven Anteil:

$$x_k = \sum_{i=1}^{q} a_i\, x_{k-i} + \sum_{j=0}^{p} b_j\, y_{k-j} .$$

Für $p = 0$ ist das Modell autoregressiv: Die aktuelle Beobachtung wird als Linearkombination der zurückliegenden Ausgangssignale und des gegenwärtigen Eingangssignals angesehen. Für $q = 0$ wird das Modell nicht-rekursiv: Der aktuelle Systemausgang ist eine Linearkombination des gegenwärtigen und der zurückliegenden Systemeingänge (ELLMER 1987). Die Koeffizienten b_j können als Koeffizienten einer Regressionsanalyse betrachtet werden.

Verhaltensmodelle sind die nicht-parametrische Möglichkeit der Angabe eines Zusammenhanges zwischen den Eingangs- und den Ausgangsgrößen eines dynamischen Systems. Verhaltensmodelle sind universell und flexibel einsetzbar, da eine theoretische Systemanalyse weitgehend vermieden wird.

Kürzlich hat man gänzlich neue Analysetechniken der Steuer- und Regelungstechnik für geodätische Anwendungen übernommen: Neurale Netze und fuzzy-regelbasierte Modelle wurden für die Identifikation von Ein-Ausgangsmodellen getestet und nutzbar gemacht (HEINE, 1999).

Möglichkeiten und Ziele dynamischer Modellierungen

Zur Analyse technischer und natürlicher Phänomene ist aus vielen Gründen die Entwicklung und Anwendung dynamischer Modelle von großer Bedeutung; sie bieten die weitreichendsten Möglichkeiten bei der Analyse und Interpretation von Deformationsvorgängen. Es werden deshalb Auswerteverfahren angestrebt, die vorrangig folgende Aufgaben verfolgen (PFEUFER 1990):

- Auswertung großer Datenmengen verschiedenartiger (hybrider) Meßsysteme zur Erfassung der Ein- und Ausgangssignale eines dynamischen Systems
- Identifizierung des Systemverhaltens in Modellen, die dem überwachten Prozeß adäquat sind
- Bearbeitung von Störeinflüssen durch Filterung
- Vorhersage der Systemreaktion beim Vorliegen „normaler" Einflüsse
- Abschätzung des Objektverhaltens bei Einwirkung extremer Einflüsse
- Ermittlung der Anteile der einzelnen Einflußgrößen an der Gesamtreaktion (-deformation)
- Ermittlung der Haupteinflußgrößen
- Möglichkeiten zur Steuerung des Deformationsprozesses über beeinflußbare Einflußgrößen
- Optimierung des Meßaufwandes anhand der gewonnenen Erkenntnisse
- umfassende Interpretation der Überwachungsergebnisse durch überschaubare und physikalisch begründete Parameter des analysierten Prozesses.

Da vielfach nur geringe Kenntnisse über die inneren und äußeren Zusammenhänge eines ablaufenden Deformationsprozesses vorliegen, ist ein allgemeingültiges Schema zur Analyse der verschiedensten Untersuchungsobjekte durch ein universelles Modell mit bestimmten, wiederkehrend zu benutzenden Methoden nicht denkbar. Die Modellklassen der Struktur- und der Verhaltensmodelle bieten jedoch die Möglichkeit, eine große Vielfalt von Prozessen sachgerecht bearbeiten zu können (NATKE 1983). Die Entwicklungen auf diesem Gebiet sind in den letzten Jahren erheblich vorangetrieben worden; dies wird sich in den kommenden Jahren fortsetzen.

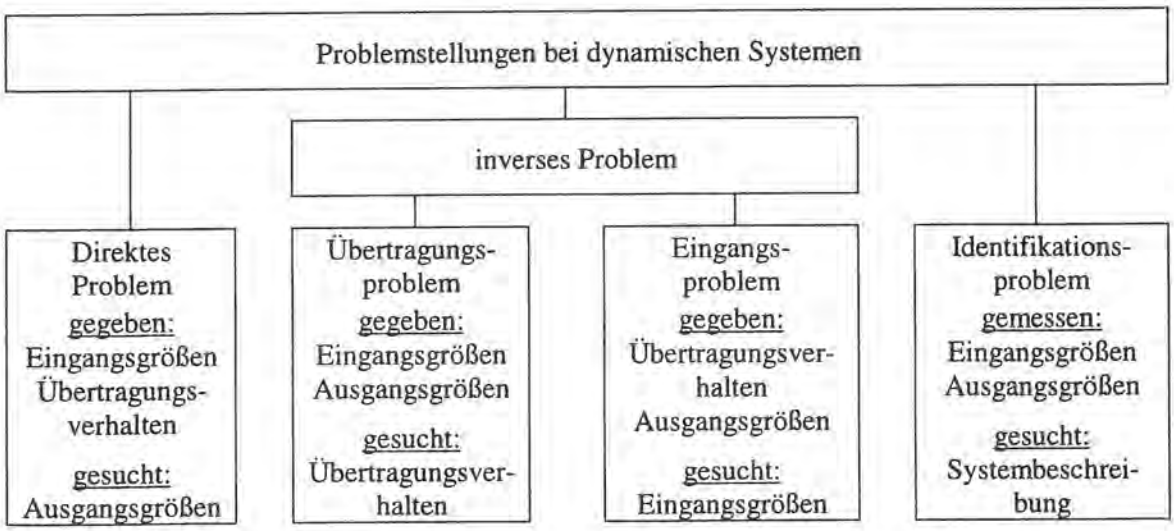


Abb. 6: Problemstellungen bei dynamischen Systemen

Wichtige Impulse sind in der letzten Zeit vor allem aus der Systemtheorie erwachsen. Überträgt man die dort entwickelten Gedankengänge, sieht man sich bei Überwachungsaufgaben mit den in Abb. 6 umrissenen grundlegenden Problemstellungen konfrontiert. Deren Umsetzung ist eine der Herausforderungen, die sich gegenwärtig stellen. Aus geodätischer Sicht ist das Identifikationsproblem der Regelfall, bei dem aus gemessenen Ein- und Ausgangsgrößen das Systemverhalten, das bei der direkten Problemstellung als bekannt vorausgesetzt wird, zu verifizieren ist. Invers wird die Aufgabenstellung, wenn entweder das Übertragungsverhalten oder die wirkenden Eingangsgrößen unbekannt sind. Die geodätisch zu erfassende Ausgangsseite liefert in diesen Fällen mitunter die einzigen Informationen für eine quantifizierbare Lösung des Problems (KERSTING 1992). Dieses Vorgehen wird auch als „reverse engineering" bezeichnet.

Systemidentifikation wird durchgeführt, um den physikalischen Zustand eines deformierbaren oder deformierten Körpers, die Spannungen, denen er unterliegt und – ganz allgemein - die Beziehung zwischen Belastung und Verformung zu bestimmen. Wenn diese Beziehung einmal ermittelt ist, kann das Ergebnis als sog. Systemgleichung als Grundlage für ein Prädiktionsmodell dienen. Durch den Vergleich der prädizierten Deformationen mit dem Ergebnis der geodätisch-geometrischen Analyse (Meßgleichung) wird ein besseres Verständnis des dynamischen Prozesses erreicht (Innovation, KALMAN-Filterung; HEUNECKE 1995, 1996). Auf diese Weise kann der Vermessungsingenieur entscheidend zu einer realistischen Interpretation des beobachteten Prozesses beitragen. Andererseits enthalten die von anderen Ingenieurdisziplinen entwickelten Systemmodelle wichtige Informationen für die Planung und Durchführung von Überwachungsmessungen und deren Auswertung. Unglücklicherweise ist dies Szenarium einer wirklich interdisziplinären Zusammenarbeit bei Entwurf und Durchführung einer Überwachungsaufgabe in der Praxis nicht sehr häufig. Die Gründe liegen darin, daß einerseits der Vermessungsingenieur die Methoden der Systemidentifikation nur unvollkommen beherrscht und daß andererseits auch die anderen Spezialisten mit den fortgeschrittenen geodätischen Analysemethoden nicht sehr vertraut sind (CHRZANOWSKI 1992). In den letzten Jahren kann man jedoch eine erfreuliche Entwicklung in der richtigen Richtung beobachten, wie häufige Beiträge zu Symposien und Kongressen und tiefschürfende und immer zahlreichere Doktorarbeiten beweisen.

Literaturhinweise

CHRZANOWSKI, A., CHEN, Y.Q. (1986): An Overview of the Physical Interpretation of Deformation Measurements. Deformation Measurements Workshop, Modern Methodology in Precise Engineering and Deformation Surveys. MIT, Massachusetts, pp. 207-220

CHRZANOWSKI, A., CHEN, Y.Q. (1990): Deformation Monitoring, Analysis and Prediction - Status Report. XIX. FIG Congress, Helsinki, Commission 6, Paper No. 604.1

CHRZANOWSKI, A., CHEN, Y.Q. (1992): Design, Monitoring, and Analysis of Deformation Surveys: Problems and Solutions. VI. Int. Symposium on Deformation Measurements, Hannover, pp. 145-160

ELLMER, W. (1987): Untersuchung temperaturinduzierter Höhenänderungen eines Großturbinentisches. Schriftenreihe des Studiengangs Vermessungswesen, Universität der Bundeswehr München, Nr. 26, Neubiberg

HEINE, K. (1999): Zur Analyse und Identifikation dynamischer Deformationsprozesse durch Ein-Ausgangs-Modelle. Dissertation, Technische Universität Braunschweig, DGK Reihe C, in print

JAEGER, R., HAAS, U., WEBER, A. (1997): Ein ISO 9000 Handbuch für Überwachungsmessungen. Schriftenreihe des Deutschen Vereins für Vermessungswesen Nr. 27, S. 415-427, K. Wittwer

HEUNECKE, O. (1995): Zur Identifikation und Verifikation von Deformationsprozessen mittels adaptiver Kalman-Filterung (Hannoversches Filter). Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Nr. 208, Hannover

HEUNECKE, O. (1996): Einige Gedanken zur fachübergreifenden Untersuchung von Deformationsvorgängen, dargestellt am Beispiel der Filterung einer Biegelinie eines Pylons. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Nr. 209, S. 5-92, Hannover

HEUNECKE, O., PELZER, H. (1998): A new Terminology for Deformation Analysis Models based on System Theory. In: KAHMEN, H. / BRÜCKL, E. / WUNDERLICH, T. (eds.): Geodesy for Geotechnical and Structural Engineering, Eisenstadt, pp. 285-292

HEUNECKE, O., PELZER, H., WELSCH, W. (1998): On the Classification of Deformation Models and Identification Methods in Engineering Surveying. XXI. FIG Congress, Commission 6, Brighton, pp. 230-245

KERSTING, N. (1992): Zur Analyse rezenter Krustenbewegungen bei Vorliegen seismotektonischer Dislokationen. Schriftenreihe des Studiengangs Vermessungswesen, Universität der Bundeswehr München, Nr. 42, Neubiberg

LEVENHAGEN, J. (1998): Identifikation synthetischer dynamischer Übertragungsmodelle und ihre Störungen im Zeitbereich. Mitteilungen aus den Geodätischen Instituten der Rheinischen Friedrich-Wilhelms-Universität, Nr. 83, Bonn

MILEV, G., PAPO, H. (1998): Classification of methods and models for determination, analysis and interpretation of deformations. XXI. FIG Congress, Commission 6, Brighton, pp. 193-201

NATKE, H.G. (1983): Einführung in Theorie und Praxis der Zeitreihen- und Modalanalyse. Verlag Friedrich Vieweg u. Sohn, Braunschweig - Wiesbaden

PFEUFER, A. (1990): Beitrag zur Identifikation und Modellierung dynamischer Deformationsprozesse. Vermessungstechnik 38, S. 19-22

PFEUFER, A. (1993): Analyse und Interpretation von Überwachungsmessungen - Terminologie und Klassifikation. Zeitschrift für Vermessungswesen 118, S. 470-476

STROBEL, H. (1975): Experimentelle Systemanalyse. Akademie-Verlag, Berlin

WELSCH, W. (1981): Gegenwärtiger Stand der geodätischen Analyse und Interpretation geometrischer Deformationen. Allgemeine Vermessungs-Nachrichten 88, S. 41-51

WELSCH, W. (1996): Geodetic analysis of dynamic processes: classification and terminology. 8th FIG International Symposium on Deformation Measurements, Hong Kong, pp. 147-156

WELSCH, W., HEUNECKE, O., KUHLMANN, H. (1999): Auswertung geodätischer Überwachungsmessungen. Wichmann Verlag, Heidelberg

WERNSTEDT, J. (1989): Experimentelle Prozeßanalyse. R. Oldenbourg Verlag, München-Wien

Zur geodätischen Beschreibung von Deformationsprozessen im 3D-Bereich

Rüdiger Wittenburg

1. Einleitung

Strainanalyse, eine Spezifikation der Zerlegung von Vektorfeldern im kinematischen Anwendungsbereich, ist in der modernen Geodäsie zu einem festen Begriff geworden. Bereits im Jahre 1967 wurde sie von E. GRAFAREND zur Beschreibung gebirgskinematischer Deformationen angewandt (GRAFAREND 1967). In der Folgezeit sind im geodätischen Anwendungsbereich eine Vielzahl von Anwendungen bekannt geworden, die sich jedoch auf die 2D-Strainanalyse beschränkten. Vergleichbare Entwicklungen einer 3D-Strainanalyse waren nicht zu beobachten, obgleich auch in dieser Richtung durch E.GRAFAREND eine vorzügliche Vordenker-Arbeit geleistet wurde (GRAFAREND 1985; GRAFAREND 1986). Offensichtlich ist der Grund hierfür in der Spezifik der geodätischen Methode zu suchen:
Die Standardmethode der geodätischen Deformationsanalyse basiert in überwiegender Mehrheit der Anwendungsfälle auf wiederholter Positionsbestimmung bzw. Koordinatenvergleich von Punkten, welche das deformierte Objekt allein auf seiner Randfläche (z.B. Erdoberfläche) repräsentieren. Methodisch gesehen, gestattet also der somit auf die Beschreibung von Objektoberflächen eingeschränkte Anwendungsbereich nur ***externe*** Deformationsbeschreibungen (bzw. externes Monitoring).
Die Zielstellung einer 3D-Beschreibung von Deformationsprozessen ist darum zunächst mit der Frage konfrontiert:
Ist mit geodätischen Standardmessungen *(Messung von Richtungen, Winkeln, Strecken und Höhenunterschieden)* ***allein, überhaupt eine 3D-Beschreibung von Objektdeformationen möglich ?***

2. Informationslücken der geodätischen Deformationsmessung im 3D-Bereich

Werden die Koordinatenunterschiede einer geodätischen Wiederholungsmessung

$$\mathbf{d} = \mathbf{X}' - \mathbf{X} \tag{1}$$

mit $\mathbf{d} = |\Delta X, \Delta Y, \Delta Z|^T$ = Vektor der Koordinatenunterschiede

$\mathbf{X} = |X, Y, Z|^T$ = Koordinatenvektor der Referenz-Messung

$\mathbf{X}' = |X', Y', Z'|^T$ = Koordinatenvektor einer Folge-Messung

als Repräsentanten eines Vektorfeldes (Geschwindigkeitsfeld) interpretiert, so kann dieses im Sinne von HELMHOLTZ (1858) zerlegt werden:

- in die Parameter der Starrkörperbewegung (Translation und Rotation), welche die Bewegung gegenüber einem Koordinatensystem beschreiben und
- in die Parameter der Deformation (Dilatation und Scherung), welche die „innere" Kinematik unabhängig vom Koordinatensystem beschreiben:

Starrkörper-bewegung:	**Deformation:**
Translation und Rotation	**Dilatation +Verformung**
	(Größen – und Formänderung)

Die entsprechende Analyse, bzw. die Zerlegung des Vektorfeldes $\mathbf{d} = \mathbf{d}(\mathbf{X}, \mathbf{X}')$ in seine Komponenten basiert auf einer Modellanpassung $\mathbf{d} = \mathbf{d}(\mathbf{X})$, für welche in geodätischen Anwendungsfällen vielfach die Taylorentwicklung

$$\mathbf{d} = \mathbf{d}_0(\mathbf{X}_0) + \left(\frac{\partial \mathbf{d}}{\partial \mathbf{X}}\right)_0 (\mathbf{X} - \mathbf{X}_0) + \frac{1}{2}\left(\frac{\partial^2 \mathbf{d}}{\partial \mathbf{X}^2}\right)_0 (\mathbf{X} - \mathbf{X}_0)^2 + \ldots\ldots \tag{2}$$

den Modellhintergrund liefert. Als Entwicklungspunkt $\mathbf{X}_0$ wird sinnvoll der Schwerpunkt der Objekt-Koordinaten angenommen, so daß im Ergebnis mit zentrierten Netzvergleichen bzw mit zentrierten Koordinaten (Schwerpunktkoordinaten)

$$\mathbf{x} = \mathbf{X} - \mathbf{X}_0 \tag{3}$$

wobei $\mathbf{x} = |x, y, z|^T$

und zentrierten Verschiebungen

$$\mathbf{u} = \mathbf{d} - \mathbf{d}_0 \tag{4}$$

wobei $\mathbf{u} = |u, v, w|^T$

und $\mathbf{d}_0 = |t_x, t_y, t_z|^T$ = Translationsvektor

gearbeitet werden kann.
Unter Vernachlässigung von Gliedern höherer Ordnung kann in hinreichend kleinen Homogenbereichen (insbesondere bei der Beschreibungen finiter Elemente) das Modell (2) durch das einfache Modell einer homogenen Deformation

$$\mathbf{d} = \mathbf{d}_0 + \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\right)_0 \mathbf{x}$$

$$\text{bzw.} \quad \mathbf{d} = \mathbf{d}_0 + (\text{grad}\,\mathbf{u})_0 \mathbf{x} \tag{5}$$

angenähert werden. In diesem Modell entspricht dem Gradienten $grad\mathbf{u}$ eine Matrix zentrierter Differenzenquotienten

$$d\mathbf{F} = \begin{vmatrix} \frac{u}{x} & \frac{u}{y} & \frac{u}{z} \\ \frac{v}{x} & \frac{v}{y} & \frac{v}{z} \\ \frac{w}{x} & \frac{w}{y} & \frac{w}{z} \end{vmatrix} = \begin{vmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{vmatrix} \tag{6}$$

Mit

$$\mathbf{F} = \mathbf{I} + d\mathbf{F} \tag{7}$$

worin $\mathbf{I}$ = Einheitsmatrix

ergibt sich das Modell der Affin-Transformation (WOLFRUM, O.1978; WELSCH,W. 1982):

$$\mathbf{X'} = \mathbf{F}\mathbf{x} + \mathbf{d}_0 \tag{8}$$

Methodisch ist die Bestimmung der Matrix $\mathbf{F}$ dem Geodäten wohl vertraut, doch liefern die geodätischen Standard-Meßprogramme im 3D-Anwendungsfall nur sechs der in $d\mathbf{F}$ benötigten neun Differenzenquotienten (WITTENBURG 1998). Aus Lagemessungen folgen vier-, ausHöhenmessungen folgen zwei Parameter:

Verschiebungsgradient aus Lagemessungen:

$$d\mathbf{F}_{H1} = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} \tag{9}$$

Verschiebungsgradient aus Höhenmessungen:

$$d\mathbf{F}_{H2} = \begin{vmatrix} w_x \\ w_y \end{vmatrix} \tag{10}$$

Unbestimmter Vertikal-Gradient:

$$d\mathbf{F}_V = \begin{vmatrix} u_z & v_z & w_z \end{vmatrix}^T \tag{11}$$

Information über die Parameter des Vertikalgradienten $d\mathbf{F}_V$ werden durch geodätische Standard-Meßprogramme nicht geliefert, so daß eine vollständige 3D-Beschreibung von Deformationsprozessen auf dieser Grundlage nicht möglich ist.

Angesichts dieser Situation steht zunächst die Frage:

Lassen sich mit einer weitergehenden Modellanpassung (als (5)) auch weitergehende Informationen gewinnen ?

2. Grenzen der Modellerweiterung

2.1 Grenzen der Modellerweiterung bei der Bestimmung der 3D-Rotation

Der methodische Werdegang einer Deformationsanalyse beginnt mit der Trennung von äußerer Starrkörperbewegung und innerer Deformation. Nach Abspaltung der Translation ist also noch vom Verschiebungsgradienten (6) die Rotation zu trennen.
In den üblichen Anwendungsfällen (mit Strainraten zwischen 10^{-6} und 10^{-3}) kann der entsprechende strenge Zusammenhang

$$d\mathbf{F} = \mathbf{E} \bullet \mathbf{R} \tag{12}$$

mit E = Straintensor
und R = Rotationstensor

durch die Approximation

$$d\mathbf{F} = \frac{1}{2}(d\mathbf{F} + d\mathbf{F}^T) + \frac{1}{2}(d\mathbf{F} - d\mathbf{F}^T) \tag{13}$$

bzw.

$$d\mathbf{F} = \mathbf{E}_0 + d\mathbf{R} \tag{14}$$

angenähert werden. Dem entspricht die Vernachlässigung des Gliedes 2. Ordnung $d\mathbf{F}^{\mathrm{T}} d\mathbf{F}$ in der Lagrangeschen Strainmatrix (DERMANIS / GRAFAREND 1992):

$$\mathbf{E} = \frac{1}{2}(d\mathbf{F} + d\mathbf{F}^{\mathrm{T}} + d\mathbf{F}^{\mathrm{T}} d\mathbf{F}) \tag{15}$$

Mit zentrierten Differenzenquotienten gilt somit für die antisymmetrische 3D- Rotationsmatrix

$$d\mathbf{R} = \begin{vmatrix} 0 & \frac{1}{2}(u_y - v_x) & -\frac{1}{2}(u_z - w_x) \\ -\frac{1}{2}(u_y - v_x) & 0 & \frac{1}{2}(v_z - w_y) \\ \frac{1}{2}(u_z - w_x) & -\frac{1}{2}(v_z - w_y) & 0 \end{vmatrix} \tag{16}$$

Da jedoch die Koeffizienten u_z und v_z nicht gegeben sind, ist die Bestimmung der vertikalen Rotationen $\mathbf{R}_{\mathrm{Y}}(\omega_y)$ und $\mathbf{R}_{\mathrm{X}}(\omega_x)$ mit einem Informationsdefizit konfrontiert.
Unter diesen Bedingungen erscheint es opportun, die aus geodätischen Höhenmessungen ableitbaren Neigungen w_x und w_y, die auch unter der Bezeichnung "Schieflagen" bekannt sind, als vertikale Rotationen zu interpretieren:

$$\omega_{\mathrm{X}} = w_Y \tag{17}$$

$$\omega_{\mathrm{Y}} = w_X \tag{18}$$

Diese zunächst plausible Arbeitshypothese impliziert jedoch die Symmetrieannahmen

$$w_Y = -v_Z \qquad \text{und} \qquad w_X = -u_Z \tag{19}$$

Die Konsequenz dieser Annahmen wird durch die symmetrische Strainmatrix

$$\mathbf{E}_0 = \begin{vmatrix} u_X & \frac{1}{2}(u_Y + v_X) & \frac{1}{2}(u_Z + w_X) \\ \frac{1}{2}(v_X + u_Y) & v_Y & \frac{1}{2}(v_Z + w_Y) \\ \frac{1}{2}(w_X + u_Z) & \frac{1}{2}(w_Y + v_Z) & w_Z \end{vmatrix} \tag{20}$$

verdeutlicht, worin mit (19) die Parameter der Scherung γ^E (engineering shear) den Betrag Null annehmen:

$$\begin{aligned} \gamma^E{}_{XZ} &= 2e_{XZ} = u_Z + w_X = 0 \\ \gamma^E{}_{YZ} &= 2e_{YZ} = v_Z + w_Y = 0 \end{aligned} \tag{21}$$

Damit wird die Unzulänglichkeit des homogenen Modells (5) verdeutlicht.

2.2 Grenzen der Modellerweiterung bei der Bestimmung des 3D-Scherstrains

Ähnlich der Beziehung zwischen "Schieflage" und vertikaler Rotation, ist ein geometrischer Zusammenhang zwischen *Scherung und Krümmung* anzunehmen.
Doch bedarf eine differentialgeometrische Beschreibung der Krümmung der 2.Ableitung des Positionsvektors bzw. der 1.Ableitung des Tangentenvektors. Hierfür ist das Modell (5) nicht hinreichend adäquat. Es sollte also im Sinne von (2) um das nächst höhere Glied erweitert werden

$$\mathbf{u} = \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\right)_0 \mathbf{x} + \frac{1}{2}\left(\frac{\partial^2 \mathbf{u}}{\partial \mathbf{x}^2}\right)_0 \mathbf{x}^2 + \ldots \tag{22}$$

Für die Funktion $w = w(x, y, z)$ gilt nunmehr (in Skalarform)

$$w = \frac{\partial w}{\partial x}x + \frac{\partial w}{\partial y}y + \frac{\partial w}{\partial z}z + \frac{1}{2}\left(\frac{\partial^2 w}{\partial x^2}x^2 + \frac{\partial^2 w}{\partial y^2}y^2 + \frac{\partial^2 w}{z^2}z^2\right) + \frac{\partial^2 w}{\partial x \partial y}xy + \frac{\partial^2 w}{\partial x \partial z}xz + \frac{\partial^2 w}{\partial y \partial z}yz \tag{23}$$

Wegen des o.g. Informationsdefizites der geodätischen Standardmessungen kann in (23) jedoch nur die Funktion $w = (x, y)$ bestimmt werden

$$w(x, y) = \frac{\partial w}{\partial x}x + \frac{\partial w}{\partial y}y + \frac{1}{2}\left(\frac{\partial^2 w}{\partial x^2}x^2 + \frac{\partial^2 w}{\partial y^2}y^2\right) + \frac{\partial^2 w}{\partial x \partial y} \tag{24}$$

Diese Funktion liefert über die Parameter

$$\frac{\partial^2 w}{\partial x^2} = \frac{\partial w_X}{\partial x} = K_X \qquad \text{und} \quad \frac{\partial^2 w}{\partial y^2} = \frac{\partial w_Y}{\partial y} = K_Y \tag{25}$$

Information über die Neigungsänderungen bzw. Krümmungen $\{K_X ; K_Y\}$.
Diesen entspricht im differentialgeometrischen Verständnis eine Veränderung des Richtungswinkels der Tangente $(t_2 - t_1)$, die meßtechnisch als Winkeländerung zweier Schieflagen bzw. als Winkelverzerrung

$$\alpha_{XZ}\text{'} - \alpha_{XZ} = e_{XZ}(\cos 2t_2 - \cos 2t_1) + \frac{1}{2}(e_{ZZ} - e_{XX})(\sin 2t_2 - \sin 2t_1) \tag{26}$$

bestimmt wird.
Die Modellerweiterung (22) ermöglicht also die zusätzliche Bestimmung der Parameter

$$e_{XZ}, e_{YZ}, e_{ZZ} - e_{XX}, e_{ZZ} - e_{YY} \tag{27}$$

Bis auf die Größe e_{zz} wären nunmehr im symmetrischen Straintensor $\mathbf{E}$

$$\mathbf{E} = \begin{vmatrix} e_{xx} & e_{xy} & e_{xz} \\ & e_{yy} & e_{yz} \\ & & e_{zz} \end{vmatrix} \tag{28}$$

alle Parameter gegeben. Eine vollständige Darstellung der 3D-Deformation ist also weiterhin ***nicht*** möglich, doch kann mit den Parametern (27) der 3D-Scherstrain

$$\gamma_T^{3D} = \sqrt{(e_1 - e_2)^2 + (e_1 - e_3)^2 + (e_2 - e_3)^2} \tag{29}$$

über die Beziehungen

$$(e_1 - e_2)^2 = (e_{XX} - e_{YY})^2 + 4e_{XY}^2 \tag{30}$$

$$(e_1 - e_3)^2 = (e_{XX} - e_{ZZ})^2 + 4e_{XZ}^2 \tag{31}$$

$$(e_2 - e_3)^2 = (e_{YY} - e_{ZZ})^2 + 4e_{YZ}^2 \tag{32}$$

bestimmt werden.

Grundsätzlich kann also festgestellt werden:
Als ein methodisches Extern - Verfahren informiert die geodätische 3D-Deformationsmessung vollständig über die tangentialen Scherparameter einer Objektdeformation, jedoch nicht über den radialen Parameter e_{ZZ}, da hierfür eine interne Deformationsmessung notwendig ist.

Also kann über die 3D-Dilatation

$$\Delta^{3D} = e_1 + e_2 + e_3 = e_{XX} + e_{YY} + e_{ZZ} \tag{33}$$

keine abschließende Aussage gemacht werden. Damit ist eine Darstellung der 3D-Deformation in Form eines Deformationsellipsoides (als geometrischer Ort des Straintensors, vgl.Bild 1) nicht möglich.

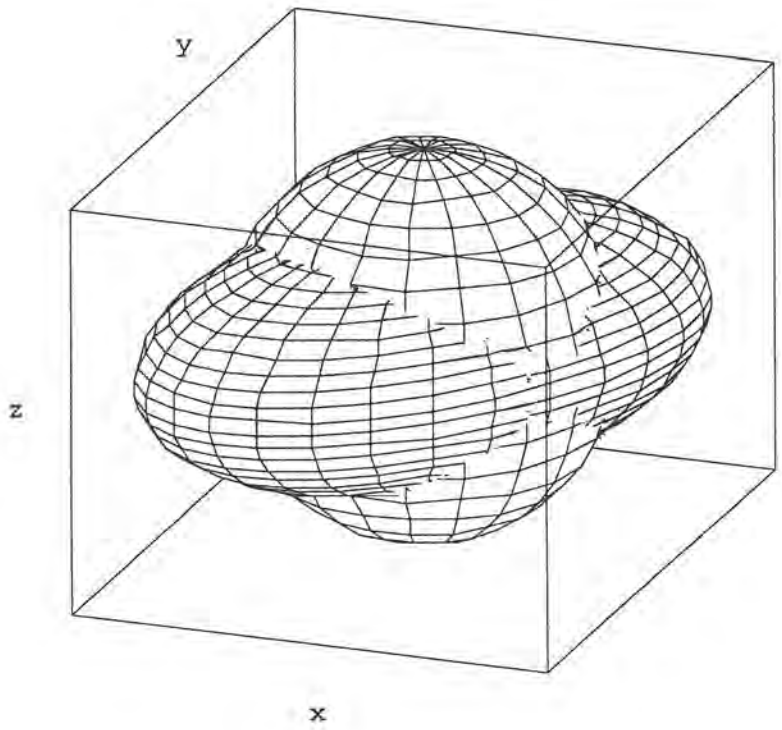


Bild 1: Deformationsellipsoid

Andererseits ist für den 3D-Scherstrain eine vollständige 3D-Darstellung möglich, die über eine 3D-Richtungsfläche (3D-Richtungsrose bzw. Fußpunktkurve der max. Winkelverzerrung) realisiert werden kann (Bild 2):

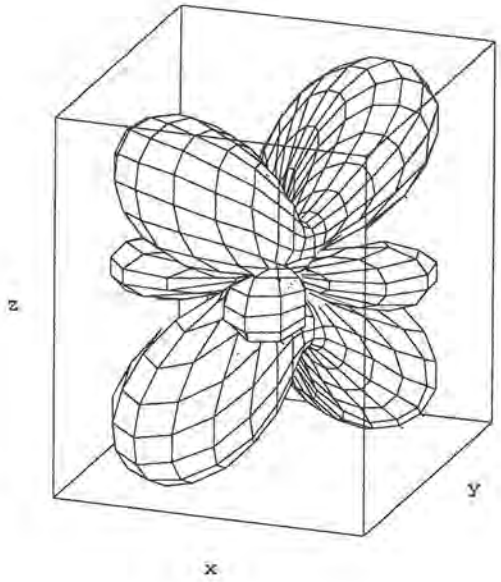


Bild 2: 3D-Richtungsfläche des Scherstrains

3. Schlußfolgerungen

Die unter Pkt.2.1 und Pkt 2.2 abgehandelten Lösungsansätze zur Überwindung der Informationsdefizite der geodätischen Deformationsmessung lösen nicht das bestehende Grundproblem, wonach eine 3D-Deformationsanalyse nur auf der Grundlage von ***externer und interner*** *Meßinformation möglich ist. Auch eine Abschätzung fehlender Parameter durch Rückrechnung (*WITTENBURG u.a. 1998*) kann nicht als Ersatz für eine interne Deformationsmessung gewertet werden.*
Es gilt ein Grundgebot der Naturwissenschaft, welches von GALILLEI formuliert wurde: Meßbar machen, was zu messen ist !

Da das bestehende Informationsdefizit meßtechnisch bedingt ist, sollte primär der Einsatz einer ergänzenden (komplementären) Meßtechnik diskutiert werden (CHEN u.a. 1985):

4. Integration von geodätischen und geotechnischen Messungen bei der Deformationsanalyse

Wie die Fundamental -Beziehung zwischen den Koordinatendifferenzen $\mathbf{d}$ und den kinematischen Ziel-Parametern $\mathbf{p}$

$$\mathbf{d} = \mathbf{H} \bullet \mathbf{p} \tag{37}$$

mit

$$\mathbf{H}=\begin{vmatrix} 1 & 0 & 0 & 0 & z & -y & x & 0 & 0 & y & z & 0 \\ 0 & 1 & 0 & -z & 0 & z & 0 & y & 0 & x & 0 & z \\ 0 & 0 & 1 & y & -x & 0 & 0 & 0 & z & 0 & x & y \end{vmatrix}$$

$$\mathbf{p}^{T}=\left| t_X t_Y t_Z \omega_X \omega_Y \omega_Z e_{XX} e_{YY} e_{ZZ} e_{XY} e_{XZ} e_{YZ} \right|$$

nochmals verdeutlicht, wäre das o.g. geodätische Problem allein schon durch die zusätzliche meßtechnische Erfassung des vertikalen Normalstrains e_{ZZ} zu lösen. Die entsprechende komplementär-geotechnische Informationsbeschaffung wäre also allein schon durch den zusätzlichen Einsatz von Vertikal-Extensometern bzw. von entsprechenden Mehrfach-Extensometern (als adäquate Form einer internen Meßtechnik) zu realisieren.

Beispiele für den Einsatz dieser Meßtechnik sind im zunehmend geotechnisch geprägten Bauwesen hinreichend zu finden. Damit erweist sich die Ingenieurgeodäsie als ein vielversprechendes Anwendungsfeld für weitere Entwicklungen der 3D-Strainanalyse:

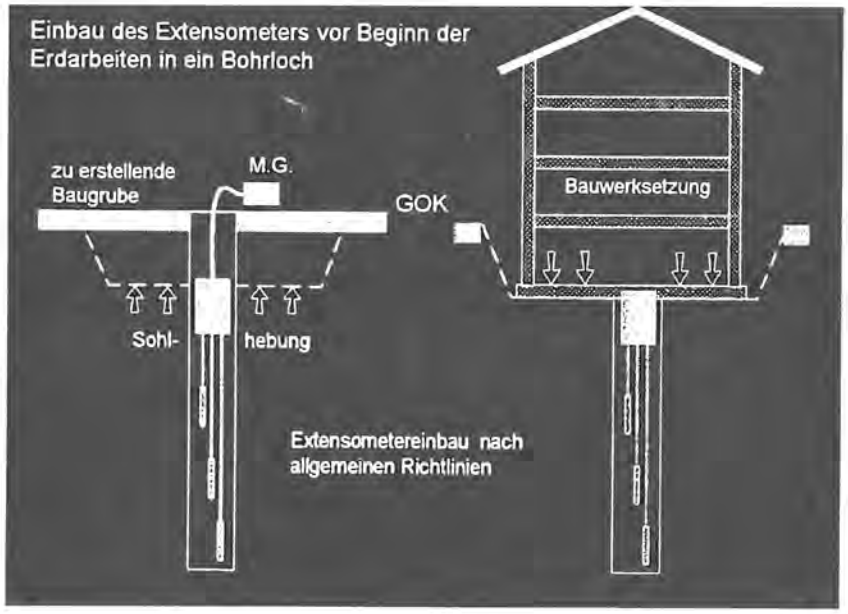


Bild 3: Zur Bestimmung des Parameters e_{zz} mit Mehrfachextensometern bei Sohlhebungs- und Bauwerkssetzungsmessung

5. Literatur

CHEN,Y.Q.;CHRANOWSKY,A.;SECORD,J.M. (1985): Generalised Modelling of Ground Movements by Intergrating Geodetic Surveys with Geotechnical Measurements. In: Proceedings VI. Int. Congress of ISM, Vol. 1, Harrogate - S. 912 - 920

DERMANIS, A. UND E. GRAFAREND (1992): The finite element approach to the geodetic computation of two- und three-dimensional Deformationparameters: a study of fram invariance and parameter estimability. Proc. Int. Conference "Cartography-Geodesy" 24.11.-3.12. in Maracaibo Venezuela.

GRAFAREND, E. W. (1967): Bergbaubedingte Deformationen und ihr Deformationstensor. Bergbau-Wiss. 14, Heft 4, S. 125-132.

GRAFAREND, E. W. (1985): Criterion Matrices for Deforming Network. In: Optimization and Design of Geodetic Networks. Ed. by E.W. Grafarend und F. Sanso, Berlin, Heidelberg, New York, Tokio: Springer Verlag -606 S.

GRAFAREND, E. W. (1986): Three-dimensional Deformation Analysis: Global Vector Spherical Harmonics ans Local Finite Element Representation. Tectonophysics, 130, 337-359.

HELMHOLTZ, H. (1858): Über Integrale der hydrodynamischen Gleichungen, welche den Winkelbewegungen entsprechen. Crelles J. 55 (1858) 25

WELSCH, W. (1982): Zur Beschreibung homogenen Strains oder Einige Betrachtungen zur affinen Transformation. In: ZfV 107 (1982) 5 S.173-182

WITTENBURG,R.;BRUMM,A.;LEHMANN, R.;NEUHAUS,W. (1998): Zur 3D-Deformationsanalyse des Senkungstroges. IN: Geoforschung, Geotechnik und Geoinformatik aus markscheiderisch-geodätischer Sicht. Vorträge zum 49.Berg-und Hüttenmännischen Tag 1998 in Freiberg. Freiberger Forschungshefte A847 S.138-151 TU Freiberg

WITTENBURG, R. (1998): Zur Beschreibung bergbauinduzierter Kinematik. In: Das Markscheidewesen 105 (1998) S.135-140

WOLFRUM, O. (1978): Die Verzerrungseigenschaften der affinen Transformation. In: AVN.-Karlsruhe 85(1978)10,-S.367-374

Random Stress/Strain Tensors and Beyond

Peiliang Xu

Abstract:

Tensors in the Earth Sciences are practically random, since they are either directly measured or indirectly inverted from other types of geo-measurements. Although random tensors have its root in multivariate analysis and nuclear physics, they are now actively investigated more as an independent topic of research; the results of random eigenvalues and eigenvectors from these investigations are mainly of asymptotic nature. In the Earth Sciences, the only result on random tensors was the accuracy of the random spectra of a random stress/strain tensor with a first order approximation. Recently, we have been working on random stress/strain tensors, the results from which are clearly borne in mind for use in the Earth Sciences. The purpose of this paper is to preliminarily review the progress of our recent studies of random second-rank symmetric (SRS) tensors. More specifically, our reviews are mainly limited to: (i) the exact distribution of the random spectra, which is numerically manageable since the dimension of tensors of geo-interest is low; (ii) the biases of the random spectra, which are physically very important but not investigated; and (iii) the accuracy of higher order approximation, which is needed if the ratio of signal to noise in stress/strain measurements is not sufficiently large. Since the eigenvector parameters are as important as the eigenvalues in the Earth Sciences, we have been paying due attention to them. On the other hand, we often encounter constrained tensors (deviatoric stress/strain tensors, pure shear tensors and seismic moment tensors, for example) in the Earth Sciences. Thus we also include the spectral theory of constrained random SRS tensors.

1 Introduction

Stress/strain tensors physically describe the deformation state of a solid deformable body. In the Earth Sciences, they are important to gain knowledge on the deformation of the solid Earth or parts of it (global geology and tectonics, for instance), and could be used to analyze and understand Earth material fracture. Crustal stress/strain tensors can either be directly measured or indirectly inverted. Direct *in situ* stress measurement has been typically based on drilling boreholes. The accuracy of *in situ* stress measurements is generally not better than $10-20$ per cent in magnitude and $10-20^o$ in orientation (Amadei & Stephasson 1997). Currently the worldwide deepest borehole is about $11.6km$. Since drilling a deep borehole is very expensive and technically limited, we can hardly expect to conduct often direct measurement of crustal stresses beyond $10km$ in depth. Inversion methods are thus frequently used to derive crustal stress/strain tensors. Geodetic data have been used to compute crustal strain tensors in lands (see, *e.g.* Frank 1966; Grafarend 1986; Prescott 1981; Savage & Burford 1970). Geometrical data of faults and earthquake focal mechanisms can be inverted for crustal stress tensors in lands and under seas (see, *e.g.* Angelier et al. 1982; Angelier 1984, 1989; Gephart & Forsyth 1984; Horiuchi et al 1995; Lu et al. 1997; McKenzie 1969; Wyss et al. 1992). The significant

advantages of inversion include integration of all types of geophysical, geological, seismological and geodetic data, and improvement of spatial (horizontal and in-depth) resolution of the crustal stress field.

Stress/strain tensors are practically random, since they are either directly measured or indirectly inverted from other geo-measurements. The study of second-rank random tensors started in problems of nuclear physics (see, *e.g.* Mehta 1990) and multivariate statistical analysis (see, *e.g.* Anderson 1958). Unlike deterministic second-rank tensors, second-rank random tensors have to be investigated from the statistical point of view. Given a probability density function (pdf) for a second-rank random tensor, it is profoundly difficult to obtain the exact distribution of the random eigenspectra. Therefore the mathematical interest of second-rank random tensors has been basically focused on approximate and/or limit distributions, for instance, of the products of random matrices and/or the random eigenspectra (see, *e.g.* Anderson 1958; Mehta 1990; Girko 1979, 1985, 1989, 1990a, b, 1992, 1993, 1997; Cohen, Kesten & Newman 1985; Furstenberg & Kesten 1960; Boutet de Monvel, Khorunzhy & Vasilchuk 1996; Oravecz & Petz 1997).

Random tensors have only recently been investigated from the statistical point of view in the Earth Sciences. Since the tensors in the Earth Sciences are physical quantities and their dimensions are generally low (≤ 3 for stress/strain tensors and ≤ 6 for elastic material tensors), mathematically approximate/limit distributions of the random eigenspectra are of limited practical value. In fact, the study of random stress/strain has been focused on the following four aspects: (1) the exact distribution of the random principal stress/strain components, since the dimension of stress/strain tensors is not greater than three and since the number of measurements is always finite; (2) the accuracy of the random eigenspectra. The accuracy is generally not investigated in the mathematical literature of second-rank random tensors. It is however a routine indicator that must be attached to any estimated/derived geo-quantity; (3) the biases of the random eigenspectra. Since the mapping between a stress/strain tensor and its eigenspectra is nonlinear, the random eigenspectra are biased. The biases of the eigenspectra, except for some inequality results on the biases of the random eigenvalues (see *e.g.* Cacoullos 1965), have not been well investigated in the mathematical literature on second-rank random tensors. They can have an important role to play in correctly interpreting the estimated stress/strain field geophysically, however; and (4) the eigendirections. The eigendirections have been almost always treated as nuisance parameters in nuclear physics and multivariate analysis. Geophysically, the eigendirections are very important and thus cannot be ignored.

The first work on random tensors in the Earth Science was to compute the first-order accuracy of the principal eigenvalues of a second-rank symmetric (SRS) random tensor (Angelier et al, 1982 as an appendix and probably independently, Soler & van Gelder 1991). Kagan & Knopoff (1985a, b) studied statistically the first two moments of stochastic three-dimensional (3D) seismic moment tensor invariants, which were then used to explain complex fault geometry (Kagan 1992a). By theoretical arguments and simulations, Kagan (1990, 1992b, 1994) used the Cauchy distribution to study earthquake focal mechanisms and incremental stress distributions. The statistics of random tensors have also contributed to the automatic detection of polarized seismic wave forms (see, *e.g.* Samson 1977; Cichowicz 1993; Dai & MacBeth 1997) and can be important for accurately estimating, interpolating and extrapolating crustal stress orientations (see, *e.g.* Hansen & Mount 1990). Given a probability density function (pdf) for an unconstrained SRS random tensor, Xu & Grafarend (1996a, b) systematically derived the joint and marginal pdfs of the random eigenvalues and eigenvector parameters. By a pdf of a random tensor, we mean in this paper a joint pdf of the random components of a random tensor in a given coordinate system. The pdfs of different random tensor component sets represented in different coordinate systems can be associated with each other using the theorem of transformation of random variables. Xu and Grafarend also computed the biases of the eigenspectral parameters, since the mapping from an SRS random tensor to its random eigenvalues and random eigenvectors is nonlinear. The relationship among the estimated, expected and true strain ellipsis has been

shown and possible physical implications noted in Xu (1996). Taking the effect of nonlinearity into account, Xu & Grafarend (1996a, b) then extended the first order accuracy computation of Angelier et al. (1982) and Soler & van Gelder (1991) to the second order approximation. The effect of nonlinearity on the inference of the relative principal stress components has been illustrated in Xu & Shimada (1997). Recently, Xu (1999) further extended the statistical theory for unconstrained SRS random tensors to the case of constrained SRS random tensors.
The purpose of this paper is basically to preliminarily review the progress on random stress/strain tensors. Although most tensors in geophysics are three-dimensional, the elastic material tensor is obviously six-dimensional. Thus we also consider n-D SRS random tensors whenever elegant results are obtainable. The paper is organized as follows. Section 2 first discusses the representation of 3D SRS stress/strain tensors and then n-D SRS tensors in the notation of rotation parameters. In Section 3, we present the differential forms and Jacobians of SRS tensors. Given a pdf for the original random tensor, we derive the joint and marginal pdfs of the random eigenvalues and eigendirection parameters in Section 4. Finally we discuss the biases and accuracy of the random eigenspectra in Section 5. Asympotic results of random singular values and random eigenvectors are not included, which can be found in a series of papers by Girko (see the references). Statistical correlation of invariants of stochastic 3D tensors is not included in this paper either, for which the reader is referred to Kagan & Knopoff (1985a, b) and Kagan (1992a).

2 Representation of symmetric second-rank tensors

2.1 Representation of 3D stress/strain tensors

The state of stress/strain at a point in the system of 3D rectangular Cartesian coordinates x, y, z can be described by a 3D SRS tensor as follows:

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_x & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \tau_y & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \tau_z \end{bmatrix}, \tag{1}$$

where $\tau_{xy} = \tau_{xy}$, $\tau_{xz} = \tau_{zx}$ and $\tau_{yz} = \tau_{zy}$. τ_x, τ_y and τ_z are the normal stresses/strains, and τ_{xy}, τ_{xz} and τ_{yz} the shear stresses/strains. Obviously a full stress/strain tensor has six independent components. Consider a second system of rectangular Cartesian coordinates x', y', z' with the same origin as the first Cartesian coordinate system x, y, z but different orientations. The relationship between these two Cartesian systems of coordinates is given by the following system of equations:

$$\left.\begin{aligned} x' &= u_{x'x}x + u_{x'y}y + u_{x'z}z \\ y' &= u_{y'x}x + u_{y'y}y + u_{y'z}z \\ z' &= u_{z'x}x + u_{z'y}y + u_{z'z}z \end{aligned}\right\},$$

or in matrix form:

$$\mathbf{x}' = \mathbf{U}^T\mathbf{x}, \tag{2}$$

where $u_{i'j}$ are the direction cosines of the i'-axis with respect to the j-axis, and

$$\mathbf{U} = \begin{bmatrix} u_{x'x} & u_{y'x} & u_{z'x} \\ u_{x'y} & u_{y'y} & u_{z'y} \\ u_{x'z} & u_{y'z} & u_{z'z} \end{bmatrix}, \tag{3}$$

$$\mathbf{x}' = (x',\ y',\ z')^T,$$

$$\mathbf{x} = (x,\ y,\ z)^T.$$

Using the law of transformation for second-rank Cartesian tensors, we have the stress/strain tensor at the point in the new Cartesian coordinate system as follows:

$$\boldsymbol{\tau}' = \mathbf{U}^T \boldsymbol{\tau} \mathbf{U}. \tag{4}$$

Among $\boldsymbol{\tau}'$ in (4), the one without off-diagonal elements is physically very important, whose diagonal elements are called the principal stress/strain components. Since $\mathbf{U}$ is orthogonal and if specially chosen, it is indeed mathematically possible to diagonalize the original stress/strain tensor $\boldsymbol{\tau}$. Denoting the diagonal $\boldsymbol{\tau}'$ by $\mathbf{Y}$ with three diagonal elements y_1, y_2 and y_3, we can rewrite (4) as follows:

$$\boldsymbol{\tau} = \mathbf{U}\mathbf{Y}\mathbf{U}^T, \tag{5}$$

where

$$\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}. \tag{6}$$

The spectral decomposition (5) is not unique. In fact, the change of any column vector of $\mathbf{U}$ in sign will not change the decomposition (5). In order to make the spectral decomposition (5) unique, one has to pre-define the orientations of the three principal stress/strain components through the manipulation of the orthogonal matrix $\mathbf{U}$. Since the number of independent components of $\boldsymbol{\tau}$ is six, the total number of independent parameters that can be used to represent the orientations of the three principal stress/strain axes is equal to three. Practically, we may use one of the following methods to construct the matrix $\mathbf{U}$ of direction cosines in (3): (i) the three Euler angles; (ii) the four Rodrigues quaternion parameters; and (iii) three rotation angles. The three Euler angles have been defined as follows (Altmann 1986): first rotation $\mathbf{U}(\gamma\mathbf{z})$ by γ around the z-axis, second rotation $\mathbf{U}(\beta\mathbf{y})$ by β around the y-axis, and finally third rotation $\mathbf{U}(\alpha\mathbf{z})$ by α around the z-axis again. Thus the matrix $\mathbf{U}$ can be represented using the three Euler angles α, β and γ, as follows:

$$\begin{aligned}\mathbf{U} &= \mathbf{U}(\alpha\mathbf{z})\mathbf{U}(\beta\mathbf{y})\mathbf{U}(\gamma\mathbf{z}) \\ &= \begin{bmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix},\end{aligned} \tag{7}$$

where

$$\mathbf{U}(\alpha\mathbf{z}) = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{U}(\beta\mathbf{y}) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix},$$

and

$$\mathbf{U}(\gamma\mathbf{z}) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since the three Euler angles can represent any point on a unit sphere, the spectral decomposition (5) will not be unique. In order to uniquely define (5), we can trivially confine the definition domains of the three Euler angles to: $-\pi/2 \leq \alpha \leq \pi/2$, $0 \leq \beta \leq \pi$ and $-\pi/2 \leq \gamma \leq \pi/2$.
The second approach to construct the matrix $\mathbf{U}$ is based on the rotation on a unit sphere. The law of motion on a unit sphere is actually the rotation around an axis by some angle (see, *e.g.* Altmann 1986), which can be symbolically written as $\mathbf{U}(\phi\mathbf{n})$, where $\phi \in [0, \pi]$ is the rotation

angle and $\mathbf{n}$ is the unit vector of rotation. The matrix $\mathbf{U}(\phi\mathbf{n})$ can be elegantly represented using the Rodrigues quaternion parameters and is given as follows:

$$\mathbf{U} = \mathbf{U}(\phi\mathbf{n}) = \begin{bmatrix} \mu^2 + m_x^2 - m_y^2 - m_z^2 & 2(m_x m_y - \mu m_z) & 2(m_z m_x + \mu m_y) \\ 2(m_x m_y + \mu m_z) & \mu^2 - m_x^2 + m_y^2 - m_z^2 & 2(m_z m_y - \mu m_x) \\ 2(m_z m_x - \mu m_y) & 2(m_z m_y + \mu m_x) & \mu^2 - m_x^2 - m_y^2 + m_z^2 \end{bmatrix}, \quad (8)$$

(Altmann 1986), where $\mu = \cos(\phi/2)$ and $\mathbf{m} = \sin(\phi/2)\mathbf{n}$. Since a 3D orthogonal matrix has only three independent parameters, the four Rodrigues parameters μ and the unit direction vector $\mathbf{n}$ are not independent but have to satisfy $\mu^2 + ||\mathbf{m}||^2 = 1$. As in the case of the Euler angles, the matrix $\mathbf{U}$ of (8) by the four Rodrigues parameters cannot uniquely determine (5), since the vector of direction cosines $\mathbf{n}$ can point to any point on the unit sphere. A unique decomposition thus requires that $\mathbf{n}$ only point to the space of $z \geq 0$ and $x \geq 0$ (or $y \geq 0$).
The third method, which has been used in the study of random tensors by Xu (1996, 1999) and Xu & Grafarend (1996a, b), is to construct the matrix $\mathbf{U}$ by:

$$\mathbf{U} = \mathbf{U}_{32}\mathbf{U}_{31}\mathbf{U}_{21}, \quad (9)$$

where

$$\mathbf{U}_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{32} & \sin\theta_{32} \\ 0 & -\sin\theta_{32} & \cos\theta_{32} \end{bmatrix},$$

$$\mathbf{U}_{31} = \begin{bmatrix} \cos\theta_{31} & 0 & \sin\theta_{31} \\ 0 & 1 & 0 \\ -\sin\theta_{31} & 0 & \cos\theta_{31} \end{bmatrix},$$

$$\mathbf{U}_{21} = \begin{bmatrix} \cos\theta_{21} & \sin\theta_{21} & 0 \\ -\sin\theta_{21} & \cos\theta_{21} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and the three angles θ_{21}, θ_{31} and θ_{32} are all defined between $-\pi/2$ and $\pi/2$. Thus the spectral decomposition (5) is unique.
Although the matrix $\mathbf{U}$ in (5) can be represented using different sets of rotations, even different ways of parameterization (see, *e.g.* Xu & Grafarend 1996a), all these different representations of $\mathbf{U}$ are mathematically equivalent. The relationship between the three Euler angles and the four Rodrigues quaternion parameters has been given in Altmann (1986). Given the joint pdf of one set of parameters for $\mathbf{U}$, one can also trivially write the joint pdf of the other set of parameters through the Jacobian between these two sets of parameters. However, the difficulty in the computation of Jacobians is different from one set of parameters to the other. On the other hand, it is not convenient to generalize the Euler representation (7) to the n-D case. The same is true for the Rodrigues quaternion representation. Even in the 3D case, it is more difficult to obtain the Jacobian by using the four Rodrigues quaternion parameters than by using our notation (9). Thus in the study of SRS random stress/strain tensors and beyond, we will confine ourselves to the representation (9) in this paper.

2.2 Representation of n-D SRS tensors

For an n-D SRS tensor $\mathbf{\Gamma}$, we can always spectrally decompose it as follows:

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{Y}\mathbf{U}^T, \quad (10)$$

where $\mathbf{Y}$ is the diagonal matrix with the eigenvalues y_1, y_2, ..., y_n in decreasing order, *i.e.* $y_1 \geq y_2 \geq \ldots \geq y_n$, $\mathbf{U}$ is the orthonormal matrix of the eigenvectors, satisfying the following condition:

$$\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}. \quad (11)$$

Because the n-D SRS tensor $\mathbf{\Gamma}$ has $N = n(n+1)/2$ independent components and because there are n independent eigenvalues, the normalized eigenvector matrix in (10) can only have $M = n(n-1)/2$ functionally independent components. If $\mathbf{U}$ is represented with M independent parameters, we can then solve for the n eigenvalues and M eigendirection parameters from the matrix equation (10). The solution $(\mathbf{Y}, \mathbf{U})$ is not unique, however. In order to obtain a unique solution $(\mathbf{U}, \mathbf{Y})$ to (10), one can either impose positive phases to the first elements of all the eigenvectors (Girko 1985; Mehta 1990) or properly select the eigenvector parameters with proper definition domains (Xu & Grafarend 1996a, b).
Following Xu & Grafarend (1996b) and Xu (1999), we define the orthonormal $\mathbf{U}$ as a product of M rotation matrices $\mathbf{U}_{ij}$ $(i > j)$, where $\mathbf{U}_{ij}$ is given by

$$\mathbf{U}_{ij} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cos\theta_{ij} & \mathbf{0} & \sin\theta_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\sin\theta_{ij} & \mathbf{0} & \cos\theta_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix}, \tag{12}$$

where $\mathbf{I}_1$, $\mathbf{I}_2$ and $\mathbf{I}_3$ are the identity matrices of different orders, $-\pi/2 \leq \theta_{ij} \leq \pi/2$, and $\mathbf{0}$ is either a zero matrix or a zero (row or column) vector. Thus the matrix $\mathbf{U}$ can be written as

$$\mathbf{U} = \mathbf{U}_{n(n-1)} \ldots \mathbf{U}_{32} \mathbf{U}_{n1} \ldots \mathbf{U}_{31} \mathbf{U}_{21}. \tag{13}$$

3 Differentials and Jacobians of SRS tensors

3.1 Differentials of SRS tensors

3.1.1 Differentials of SRS tensors without constraints

Differentiating (10) and (11) and then combining these two differentials, we obtain:

$$d\mathbf{\Gamma} = \mathbf{U}(d\mathbf{Y} + \mathbf{U}^T d\mathbf{U}\mathbf{Y} - \mathbf{Y}\mathbf{U}^T d\mathbf{U})\mathbf{U}^T, \tag{14}$$

(see *e.g.* Xu & Grafarend 1996b).
Substituting (13) into (14) yields:

$$d\mathbf{\Gamma} = \mathbf{U}(d\mathbf{Y} + \sum_{k=1}^{M} \mathbf{H}_k d\theta_k)\mathbf{U}^T, \tag{15}$$

where

$$d\theta_k = d\theta_{ij},$$

$$\mathbf{H}_k = \mathbf{U}_{II(k)}^T \mathbf{O}_{ij(k)} \mathbf{U}_{II(k)} \mathbf{Y} - \mathbf{Y}\mathbf{U}_{II(k)}^T \mathbf{O}_{ij(k)} \mathbf{U}_{II(k)},$$

$$\mathbf{O}_{ij(k)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -1 & \mathbf{0} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{U}_{I(k)} = \mathbf{U}_{n(n-1)} \mathbf{U}_{n(n-2)} \ldots \mathbf{U}_{i(j+1)},$$

$$\mathbf{U}_{II(k)} = \mathbf{U}_{i(j-1)} \ldots \mathbf{U}_{32} \mathbf{U}_{n1} \ldots \mathbf{U}_{31} \mathbf{U}_{21}.$$

Since the tensor $\mathbf{\Gamma}$ is symmetric, we vectorize both sides of (15) by eliminating the elements in the upper triangle of each symmetric tensor and obtain:

$$v(d\mathbf{\Gamma}) = \mathbf{D}_n^+ (\mathbf{U} \otimes \mathbf{U}) \mathbf{D}_n \, v\{d\mathbf{Y} + \sum_{k=1}^{M} \mathbf{H}_k d\theta_k\}, \tag{16}$$

where $\mathbf{D}_n$ is the *duplication matrix* and $\otimes$ stands for Kronecker product of matrices (Magnus & Neudecker 1988).
For a full 3D stress/strain tensor, (16) becomes

$$v(d\boldsymbol{\tau}) = \mathbf{D}_3^{+}(\mathbf{U} \otimes \mathbf{U})\mathbf{D}_3 \; v\{d\mathbf{Y} + \mathbf{H}_{21}d\theta_{21} + \mathbf{H}_{31}d\theta_{31} + \mathbf{H}_{32}d\theta_{32}\}, \tag{17}$$

where $\mathbf{U}$ is given by (9), the three non-zero elements of $d\mathbf{Y}$ are dy_1, dy_2 and dy_3, and the three matrices $\mathbf{H}_{21}$, $\mathbf{H}_{31}$ and $\mathbf{H}_{32}$ are respectively given by:

$$\mathbf{H}_{21} = \begin{bmatrix} 0 & y_2 - y_1 & 0 \\ & 0 & 0 \\ symm & & 0 \end{bmatrix},$$

$$\mathbf{H}_{31} = \begin{bmatrix} 0 & 0 & (y_3 - y_1)\cos\theta_{21} \\ & 0 & (y_3 - y_2)\sin\theta_{21} \\ symm & & 0 \end{bmatrix},$$

$$\mathbf{H}_{32} = \begin{bmatrix} 0 & (y_2 - y_1)\sin\theta_{31} & -(y_3 - y_1)\sin\theta_{21}\cos\theta_{31} \\ & 0 & (y_3 - y_2)\cos\theta_{21}\cos\theta_{31} \\ symm & & 0 \end{bmatrix}.$$

3.1.2 Differentials of constrained SRS tensors

Many important SRS tensors in the Earth Sciences are conditionally constrained. A deviatoric 3D stress/strain tensor has only five independent components and is subject to the traceless constraint. In physical geodesy, The Laplace equation of the geopotential field also demands that all gravity tensors be traceless. A double-couple (DC) point source in seismology can be represented by a 3D SRS seismic moment tensor whose components are subject to zero isotropic and zero intermediate principal component constraints (Aki & Richards 1980; Kostrov & Das 1988; Lay & Wallace 1995). Mathematically, the pure shear tensor on a certain oriented plane is of the same form as a DC seismic moment tensor in seismology.
Assume that the components of a 3D SRS tensor $\boldsymbol{\tau}$ are constrained by:

$$\mathbf{h}(\boldsymbol{\tau}) = 0, \tag{18}$$

where $\mathbf{h}(\boldsymbol{\tau})$ is an m_c-dimensional function vector of the components of the tensor $\boldsymbol{\tau}$. The spectral decomposition of a constrained 3D SRS tensor is now equivalent to simultaneously solving (5) and (18) for the eigenvalues and the eigendirection parameters. Since (18) imposes m_c conditions on the components of the 3D SRS tensor $\boldsymbol{\tau}$, the total number of independent equations for the spectral decomposition is equal to $(6 - m_c)$. Thus there can only be $(6 - m_c)$ independent eigenvalues and eigendirection parameters. In the similar manner, in order to derive the differential relationship between a set of $(6 - m_c)$ independent tensor components and $(6 - m_c)$ eigenvalues and eigendirection parameters, we have to first differentiate (18) and then combine it with (17). For more details, we refer the reader to Xu (1999). To summarize, we give the differential relations between constrained 3D SRS tensor components and their eigenvalues and eigendirection parameters as follows:

- **for 3D deviatoric stress/strain tensors:**
  The constraint condition of a 3D deviatoric stress/strain tensor can be expressed either by

  $$\tau_x + \tau_y + \gamma_z = 0,$$

  in terms of the original tensor components, or by

  $$y_1 + y_2 + y_3 = 0,$$

in terms of the eigenvalues ($y_1 \geq y_2 \geq y_3$).

For 3D deviatoric (traceless) stress/strain tensors with (9), (17) becomes:

$$v_{c_1}(d\boldsymbol{\tau}) = \mathbf{D}_{3c_1}^{+}(\mathbf{U} \otimes \mathbf{U})\mathbf{D}_{3c_1} v_{c_1}\{d\mathbf{Y} + \mathbf{H}_{21}d\theta_{21} + \mathbf{H}_{31}d\theta_{31} + \mathbf{H}_{32}d\theta_{32}\}, \qquad (19)$$

where $v_{c_1}(d\boldsymbol{\tau})$ stands for the vectorization operation under the traceless constraint, which is actually equal to $v(d\boldsymbol{\tau})$ without the last element if τ_z and y_3 are taken as nuisance parameters, the matrices $\mathbf{U}$, $\mathbf{H}_{21}$, $\mathbf{H}_{31}$ and $\mathbf{H}_{32}$ are the same as in (17), $\mathbf{D}_{3c_1}$ is the *duplication matrix* in the presence of the traceless constraint and is given by:

$$\mathbf{D}_{3c_1} = \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & -1 & 0 \end{array}\right].$$

- **for the pure shear tensor on an oriented plane:**
  A pure shear tensor on an oriented plane or a 3D DC SRS seismic moment tensor can be represented by four independent tensor components (see, *e.g.* Brekhovskikh & Goncharov 1994; Lay & Wallace 1995). The two constraints for this type of tensors can be written as follows:

$$y_1 + y_3 = 0;$$

$$y_2 = 0,$$

in terms of the eigenvalues, or equivalently,

$$\tau_x + \tau_y + \tau_z = 0;$$

$$det\{\boldsymbol{\tau}\} = 0,$$

in terms of the original tensor components.

The two constraints for the pure shear tensor on an oriented plane will make only four independent equations from (5). Here we choose τ_x, τ_{xy}, τ_{xz} and τ_{yz}, and collect them into a vector,

$$v_{c_2}(d\boldsymbol{\tau}) = (\tau_x, \tau_{xy}, \tau_{xz}, \tau_{yz})^T.$$

Note, however, that the other two components τ_y and τ_z cannot be arbitrary but have to be solved using the two constraints specified in the above. Left-multiplying both sides of (17) by the matrix $\mathbf{D}_{c_2}$:

$$\mathbf{D}_{c_2} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right],$$

we obtain the differential form for the pure shear tensor,

$$v_{c_2}(d\boldsymbol{\tau}) = \mathbf{D}_{c_2}\mathbf{D}_3^{+}(\mathbf{U} \otimes \mathbf{U})\mathbf{D}_3\ v\{d\mathbf{Y} + \mathbf{H}_{21}d\theta_{21} + \mathbf{H}_{31}d\theta_{31} + \mathbf{H}_{32}d\theta_{32}\}. \qquad (20)$$

- **for 3D SRS tensors with only one non-zero eigenvalue:**
A 3D SRS tensor with only one non-zero eigenvalue has three distinct components and physically represents a completely polarized wave field (Samson 1977). Mathematically, it can be rewritten as follows:

$$\boldsymbol{\tau} = \mathbf{U} \begin{bmatrix} y & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^T = y \mathbf{u}_1 \mathbf{u}_1^T, \tag{21}$$

where y is the only non-zero eigenvalue. $\mathbf{U}$ is defined as in (9), and $\mathbf{u}_1$ is the first column vector of $\mathbf{U}$. Obviously the SRS tensor $\boldsymbol{\tau}$ in (21) has only three independent components. Except for the only non-zero eigenvalue y, and since $||\mathbf{u}_1|| = 1$, there can only be two independent components for $\mathbf{U}$. Without loss of generality, we can assume that y is non-negative and represent $\mathbf{U}$ in (9) simply as follows:

$$\mathbf{U} = \mathbf{U}_{31} \mathbf{U}_{21},$$

or elementwise,

$$u_{11} = \cos\theta_{31} \cos\theta_{21},$$
$$u_{21} = -\sin\theta_{21}; \quad u_{31} = -\sin\theta_{31} \cos\theta_{21}.$$

Differentiating both sides of (21), we have

$$d\boldsymbol{\tau} = \mathbf{u}_1 \mathbf{u}_1^T dy + y\, d\mathbf{u}_1 \mathbf{u}_1^T + y\, \mathbf{u}_1 d\mathbf{u}_1^T. \tag{22}$$

Selecting the second column of $\boldsymbol{\tau}$ for (22), we have

$$\frac{\partial(\tau_{xy}, \tau_y, \tau_{yz})^T}{\partial(y, \theta_{21}, \theta_{31})} = \mathbf{M}_\theta \mathbf{Y}_y, \tag{23}$$

where

$$\mathbf{Y}_y = \begin{bmatrix} \sin\theta_{21} & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & y \sin\theta_{21} \cos\theta_{21} \end{bmatrix},$$

$$\mathbf{M}_\theta = \begin{bmatrix} -\cos\theta_{21} \cos\theta_{31} & -\cos 2\theta_{21} \cos\theta_{31} & \sin\theta_{31} \\ \sin\theta_{21} & \sin 2\theta_{21} & 0 \\ \cos\theta_{21} \sin\theta_{31} & \cos 2\theta_{21} \sin\theta_{31} & \cos\theta_{31} \end{bmatrix}.$$

For the n-D case and more details of the derivation, the reader is referred to Xu (1999).

3.2 Jacobians of SRS tensors

For the tensor equation (10) with m_c constraints, if we pre-select a set of $(n(n+1)/2 - m_c)$ independent tensor components and the same number of independent eigenvalues y_i (in non-increasing order) and eigendirection parameters α_i, then the Jacobian of the components of the constrained SRS tensor $\boldsymbol{\Gamma}$ with respect to $\mathbf{Y}$ and $\mathbf{U}$ is defined as follows:

$$J(\mathbf{y}, \boldsymbol{\alpha}) = \left| det \left\{ \frac{\partial(\gamma_1, \gamma_2, ..., \gamma_{(n(n+1)/2 - m_c)})^T}{\partial(y_1, y_2, ..., y_i, \alpha_1, \alpha_2, ..., \alpha_M)} \right\} \right|, \tag{24}$$

where $det\{.\}$ stands for the determinant of a square matrix, γ_j are the functionally independent components of the constrained SRS tensor $\boldsymbol{\Gamma}$ and $M = (n-1)n/2 - m_c - i$. In the case of $m_c = 0$, the SRS tensor $\boldsymbol{\Gamma}$ becomes unconstrained. The Jacobian of an unconstrained SRS tensor with respect to n eigenvalues and an arbitrary set of eigendirection parameters was given in implicit function form of the eigenparameters in Anderson (1958), Mehta (1990) and Girko (1985), which can be summarized as Theorem 1:

Theorem 1 *Let $\mathbf{\Gamma}$ be a real n-D SRS tensor, whose eigenvalues and eigendirection parameters are respectively denoted by $\mathbf{y} = (y_1, y_2, ..., y_n)$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_M)$. Then the Jacobian of $\mathbf{\Gamma}$ with respect to $\mathbf{y}$ and $\boldsymbol{\alpha}$ is given by*

$$J(\mathbf{y}, \boldsymbol{\alpha}) = \left| det \left\{ \frac{\partial v(\mathbf{\Gamma})}{\partial(\mathbf{y}, \boldsymbol{\alpha})} \right\} \right| = \prod_{i<j} (y_i - y_j) s(\boldsymbol{\alpha}), \tag{25}$$

where $s(\boldsymbol{\alpha})$ is an implicit function of $\boldsymbol{\alpha}$ but independent of $\mathbf{y}$.

For the proof of Theorem 1, the reader is referred to Mehta (1990) or Girko (1990b, Chap.3). Since the eigendirection parameters are generally of no interest in multivariate analysis and nuclear physics, the function $s(\boldsymbol{\alpha})$ is not derived explicitly. In mechanics and the Earth Sciences, the eigendirection parameters are as important as (if not more imporant than) the eigenvalues. Thus we have been trying to work out $s(\boldsymbol{\alpha})$, the results of which can be summarized by the following theorem:

Theorem 2 *Let $\mathbf{\Gamma}$ be a real n-D SRS tensor, whose eigenvalues and eigendirection parameters are respectively denoted by $\mathbf{y} = (y_1, y_2, ..., y_n)$ and $\boldsymbol{\theta} = (\theta_{21}, \theta_{31}, ..., \theta_{n1}, \theta_{32}, ..., \theta_{n(n-1)})$, as specified in (13). Then the Jacobian of $\mathbf{\Gamma}$ with respect to $\mathbf{y}$ and $\boldsymbol{\theta}$ is given by*

$$J(\mathbf{y}, \boldsymbol{\theta}) = \left| det \left\{ \frac{\partial v(\mathbf{\Gamma})}{\partial(\mathbf{y}, \boldsymbol{\theta})} \right\} \right| = \prod_{i<j} (y_i - y_j) \prod_{i=1}^{n-2} \prod_{j=i+2}^{n} (\cos\theta_{ji})^{j-i-1}. \tag{26}$$

For the proof of Theorem 2, the reader is referred to Xu (1999). In the 3D case of a full stress/strain tensor, the Jacobian (26) becomes:

$$J(y_1, y_2, y_3, \theta_{21}, \theta_{31}, \theta_{32}) = (y_1 - y_2)(y_2 - y_3)(y_1 - y_3) \cos\theta_{31}. \tag{27}$$

Using the differential results of (19), (20) and (23), we can readily summarize the Jacobian results of deviatoric stress/strain and pure shear tensors and the SRS tensor with only one non-zero eigenvalue in Table 1, where the four elements of $\mathbf{M}_{c_{2S}}$ are given by

$$m_{c_{2S}}^{11} = \cos 2\theta_{31}/3 + (3\cos^2\theta_{21} - 1)(2 - \cos^2\theta_{31})/3,$$

$$m_{c_{2S}}^{12} = \sin 2\theta_{21} \sin\theta_{31} (2 - \cos^2\theta_{31}),$$

$$m_{c_{2S}}^{21} = \sin^2\theta_{21} \sin 2\theta_{32},$$

and

$$m_{c_{2S}}^{22} = \cos 2\theta_{32} - \sin 2\theta_{21} \sin\theta_{31} \sin 2\theta_{32}.$$

4 Probability distributions of random spectra

Let $\mathbf{T}$ be an n-D SRS random tensor whose components are subject to m_c equality constraints. Then in the similar manner to (10), we can decompose $\mathbf{T}$ as:

$$\mathbf{T} = \mathbf{G}\mathbf{\Lambda}\mathbf{G}^T, \tag{28}$$

subject to the m_c constraints, where $\mathbf{\Lambda}$ has i functionally independent random eigenvalues satisfying $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_i$, $\mathbf{G}$ is a random orthonormal matrix represented by $(n(n+1)/2 - m_c - i)$ functionally independent random eigendirection parameters. To represent $\mathbf{G}$, we use our rotation notations but not an arbitrary set of parameters as in Mehta (1990) or Girko (1985), for instance. Thus $\mathbf{G}$ is the same as $\mathbf{U}$ except that θ_{ij} in (10) are replaced with

Table 1: Jacobians of unconstrained and constrained 3D SRS tensors: Unconstrained – unconstrained 3D stress/strain tensors; Isotropic – traceless 3D stress/strain tensors; Pure Shear – pure shear tensors; Single Eigenvalue – 3D SRS tensors with one non-zero eigenvalue; and $g(\theta_{31}) = \cos^3\theta_{31}(1+2\sin^2\theta_{31})$. (modified from Xu (1999))

Models	$y_1, y_2, y_3, \theta_{21}, \theta_{31}, \theta_{32}$
Unconstrained	$(y_1-y_2)(y_1-y_3)(y_2-y_3)\cos\theta_{31}$
Isotropic	$(y_1-y_2)(2y_1+y_2)(y_1+2y_2)g(\theta_{31})$
Pure Shear	$2y_1^3\cos\theta_{31}\lvert det\{\mathbf{M}_{c_{2S}}\}\rvert$
Single Eigenvalue	$y^2\lvert\sin\theta_{21}\rvert^3\cos\theta_{21}$

ϕ_{ij}. Denote the probability density function (pdf) of the random tensor $\mathbf{T}$ by $f_T(\mathbf{\Gamma})$. Very often, if the components of the random SRS tensor can be approximately derived linearly from geodetic measurements, $\mathbf{T}$ has a Gaussian distribution; in multivariate analysis, $\mathbf{T}$ has a Wishart distribution if the observations are normally distributed. In this section, we will first discuss the joint pdf of the random eigenvalues and random rotations of $\mathbf{T}$ without constraints, and then focus on the random eigenvalues and random rotations of 3D random stress/strain tensors.

4.1 Probability distributions of n-D random spectra

For a full n-D SRS random tensor $\mathbf{T}$, we denote the n random eigenvalues and $n(n-1)/2$ random eigenvector parameters by $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$, respectively. Then with the Jacobian (26), we obtain the joint pdf of the random eigenvalues $\boldsymbol{\lambda}$ and the random rotations $\boldsymbol{\Phi}$:

$$\begin{aligned} f_{\lambda\Phi}(\mathbf{y},\boldsymbol{\theta}) &= f_T(\mathbf{UYU}^T)\left|det\left\{\frac{\partial v(\mathbf{\Gamma})}{\partial(\mathbf{y},\boldsymbol{\theta})}\right\}\right| \\ &= \prod_{i<j}(y_i-y_j)\; f_T(\mathbf{UYU}^T)\prod_{i=1}^{n-2}\prod_{j=i+2}^{n}(\cos\theta_{ji})^{j-i-1}, \end{aligned} \tag{29}$$

where $v(\mathbf{\Gamma})$ is the v-operation of $\mathbf{\Gamma}$, $\mathbf{y}$ is the vector of the eigenvalues, and $\boldsymbol{\theta}$ consists of all the rotations used to represent (13).
Hence the marginal (joint) pdf of the distinct random eigenvalues can be obtained by integrating (29) over the definition domain of $\boldsymbol{\theta}$:

$$f_\lambda(\mathbf{y}) = \int_{\Omega_\theta}\prod_{i<j}(y_i-y_j)\; f_T(\mathbf{UYU}^T)\prod_{i=1}^{n-2}\prod_{j=i+2}^{n}(\cos\theta_{ji})^{j-i-1}\,d\Omega_\theta, \tag{30}$$

where the domain Ω_θ of $\boldsymbol{\theta}$ is defined by $-\pi/2 \le \theta_{ij} \le \pi/2$ for all $i < j \le n$. In the similar manner, we can obtain the marginal pdf of the random rotations:

$$f_\Phi(\boldsymbol{\theta}) = \int_{\Omega_y}\prod_{i<j}(y_i-y_j)\; f_T(\mathbf{UYU}^T)\prod_{i=1}^{n-2}\prod_{j=i+2}^{n}(\cos\theta_{ji})^{j-i-1}\,d\Omega_y, \tag{31}$$

where the domain Ω_y is defined by $-\infty < y_n \le \ldots \le y_2 \le y_1 < \infty$. If the n-D SRS random tensor $\mathbf{T}$ is positive definite (the estimated variance-covariance matrix in multivariate analysis, for example), $y_n > 0$.

The marginal pdfs (30) and (31) of the random eigenvalues and random rotations can hardly have a simple analytical expression generally, unless $f_T(\boldsymbol{Gamma})$ takes a certain special form. In this case, the probability of the random eigenvalues or random rotations can be too approximate to use if it is computed by using the formulas in Stroud (1971) if the dimension n is large. However, for a certain class of pdfs of n-D SRS random tensors $\mathbf{T}$, we can have very elegant formulas for the marginal pdfs of the random eigenvalues and random rotations, which is summarized in the following theorem.

Theorem 3 *Let $\mathbf{T}$ be a real n-D SRS random tensor, whose random eigenvalues and random rotation parameters are respectively denoted by $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)$ with its elements satisfying $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ and $\boldsymbol{\Phi} = (\phi_{21}, \phi_{31}, ..., \phi_{n(n-1)})$. If the pdf $f_T(\boldsymbol{\Gamma})$ of $\mathbf{T}$ is invariant under the rotation group,* i.e. *$f_T(\boldsymbol{\Gamma}) = f_T(\mathbf{y})$, then the random eigenvalues $\boldsymbol{\lambda}$ and the random eigenrotations $\boldsymbol{\Phi}$ are stochastically independent and respectively have the following marginal distributions:*

$$f_\lambda(\mathbf{y}) \sim f_T(\mathbf{y}) \prod_{i<j} (y_i - y_j) \tag{32}$$

for the n random eigenvalues, and

$$f_\Phi(\boldsymbol{\theta}) \sim \prod_{i=1}^{n-2} \prod_{j=i+2}^{n} (\cos\theta_{ji})^{j-i-1} \tag{33}$$

for the $n(n-1)/2$ random rotations.

For the proof and usage of (32), the reader is referred to Mehta (1990), Girko (1985) or Xu & Grafarend (1996b); (33) is a consequence of Theorem 2 (see also Xu 1999). In particular, it is very important to note that (33) of Theorem 3 indicates that all the random eigenrotations are also stochastically independent.

4.2 Probability distributions of 3D random spectra

For 3D random tensors $\mathbf{T}$ of geo-interest, the three random eigenvalues and three random rotations are denoted by $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ satisfying $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\boldsymbol{\Phi} = (\phi_{21}, \phi_{31}, \phi_{32})$ as in the above n-D case, respectively. Assume that the 3D SRS random tensor $\mathbf{T}$ has a pdf $f_T(\boldsymbol{\tau})$. Then in the similar manner to (29) to (31), for a full 3D random stress/strain tensor $\mathbf{T}$, we can readily obtain the joint and marginal pdfs of the three random principal stress/strain components and three random rotations as follows:

$$f_{\lambda\Phi}(\mathbf{y}, \boldsymbol{\theta}) = \prod_{1\leq i<j\leq 3} (y_i - y_j)\ f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T) \cos\theta_{31}, \tag{34}$$

$$f_\lambda(y_1, y_2, y_3) = \int_{\Omega_\theta} \prod_{1\leq i<j\leq 3} (y_i - y_j)\ f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T) \cos\theta_{31}\, d\Omega_\theta, \tag{35}$$

and

$$f_\Phi(\theta_{21}, \theta_{31}, \theta_{32}) = \int_{\Omega_y} \prod_{1\leq i<j\leq 3} (y_i - y_j)\ f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T) \cos\theta_{31}\, d\Omega_y, \tag{36}$$

where $\mathbf{U}$ and $\mathbf{Y}$ have been defined in (5) and (9), the domains Ω_y and Ω_θ are respectively defined by $-\infty < y_3 \leq y_2 \leq y_1 < \infty$ and $-\pi/2 \leq \theta_{21}, \theta_{31}, \theta_{32} \leq \pi/2$.
If the pdf $f_T(\boldsymbol{\tau})$ takes the form:

$$\begin{aligned} f_T(\boldsymbol{\tau}) &= exp\{-a\, tr(\boldsymbol{\tau}^2) + b\, tr(\boldsymbol{\tau}) + c\} \\ &= exp\{-a \sum_{i=1}^{3} y_i^2 + b \sum_{i=1}^{3} y_i + c\}, \end{aligned}$$

where $a > 0$, b and c must satisfy the condition of unity probability, then the marginal pdfs of the three random principal stress/strain components and three random rotations become:

$$f_\lambda(y_1, y_2, y_3) = 2\pi^2 \prod_{1 \le i < j \le 3} (y_i - y_j) exp\{-a \sum_{i=1}^{3} y_i^2 + b \sum_{i=1}^{3} y_i + c\},$$

and

$$f_\Phi(\theta_{21}, \theta_{31}, \theta_{32}) = \frac{1}{2\pi^2} \cos\theta_{31},$$

respectively (Xu & Grafarend 1996b).

Now we shall further examine the distributions of the random spectra of the 3D deviatoric stress and pure shear random tensors, with numerical demonstrations.

- **3D deviatoric random stress/strain tensors.**
  Given a pdf $f_T(\boldsymbol{\tau})$ for the deviatoric random stress $\mathbf{T}$, and with the corresponding Jacobian in Table 1, we readily obtain the joint pdf of the two random principal stress components and the three random rotations:

$$f_{\lambda\Phi}(\mathbf{y}, \boldsymbol{\theta}) = f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T)(y_1 - y_2)(2y_1 + y_2)(y_1 + 2y_2)\cos^3\theta_{31}(1 + 2\sin^2\theta_{31}), \quad (37)$$

  the marginal pdf of the two random principal stress components:

$$f_\lambda(y_1, y_2) = \int_{\Omega_\theta} f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T)(y_1 - y_2)(2y_1 + y_2)(y_1 + 2y_2)\cos^3\theta_{31}(1 + 2\sin^2\theta_{31})\, d\Omega_\theta, \quad (38)$$

  and the marginal pdf of the three random rotations:

$$f_\Phi(\theta_{21}, \theta_{31}, \theta_{32}) = \int_{\Omega_y} f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T)(y_1 - y_2)(2y_1 + y_2)(y_1 + 2y_2)\cos^3\theta_{31}(1 + 2\sin^2\theta_{31})\, d\Omega_y, \quad (39)$$

  respectively, where $-\infty < y_2 \le y_1 < \infty$.

  For numerical demonstrations of (38) and (39), we construct a 3D deviatoric random stress tensor with the three principal stresses 15.1, -1.6 and -13.5 MPa, based on the real stress data taken from Amadei & Stephasson (1997). The three orientations are generated artificially for making a full deviatoric stress tensor. The five independent stress components are assumed to have a relative error of 15 per cent and to be statistically independent. Then we assume the Gaussian and Laplacian probability models for the generated random stress components. The marginal pdfs of the two random principal stresses are shown in Fig.1 and Fig.2, and the marginal pdfs of the three random rotations in Fig.3. These figures have shown that the pdfs of the random eigenvalues are significantly different from normal and the pdfs of the random rotations significantly different from either normal or standard (Fisher's) pdf model for directional data. For more details on the example, the reader is referred to Xu (1999).

- **3D pure shear random tensors.**
  In a similar manner to (37), (38) and (39), we have the joint pdf of the only random principal stress component and the three random rotations:

$$f_{\lambda\Phi}(y, \boldsymbol{\theta}) = f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T) 2y_1^3 \cos\theta_{31} |det\{\mathbf{M}_{c_{2S}}\}|, \quad (40)$$

  the marginal pdf of the two random principal stress components:

$$f_\lambda(y) = \int_{\Omega_\theta} f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T) 2y_1^3 \cos\theta_{31} |det\{\mathbf{M}_{c_{2S}}\}|\, d\Omega_\theta, \quad (41)$$

and the marginal pdf of the three random rotations:

$$f_{\Phi}(\theta_{21}, \theta_{31}, \theta_{32}) = \int_{\Omega_y} f_T(\mathbf{U}\mathbf{Y}\mathbf{U}^T) 2y_1^3 \cos\theta_{31} |det\{\mathbf{M}_{c_{2S}}\}| \, d\Omega_y, \tag{42}$$

respectively, where $-\infty < y < \infty$.

For illustrative purposes, we have constructed a pure shear random tensor (exactly the same as the DC seismic moment example in Xu (1999) except the difference in unit). The pdfs of the random principal stress with the Gaussian and Laplacian models for the original random tensor components are shown in Fig.4. It is very clear again that the pdfs of the random eigenvalue is significantly different from normal. Fig.5 has plotted the pdfs of the three random rotations, which also have shown that Fisher's model is not representative of the rotations of the 3D pure shear random tensor.

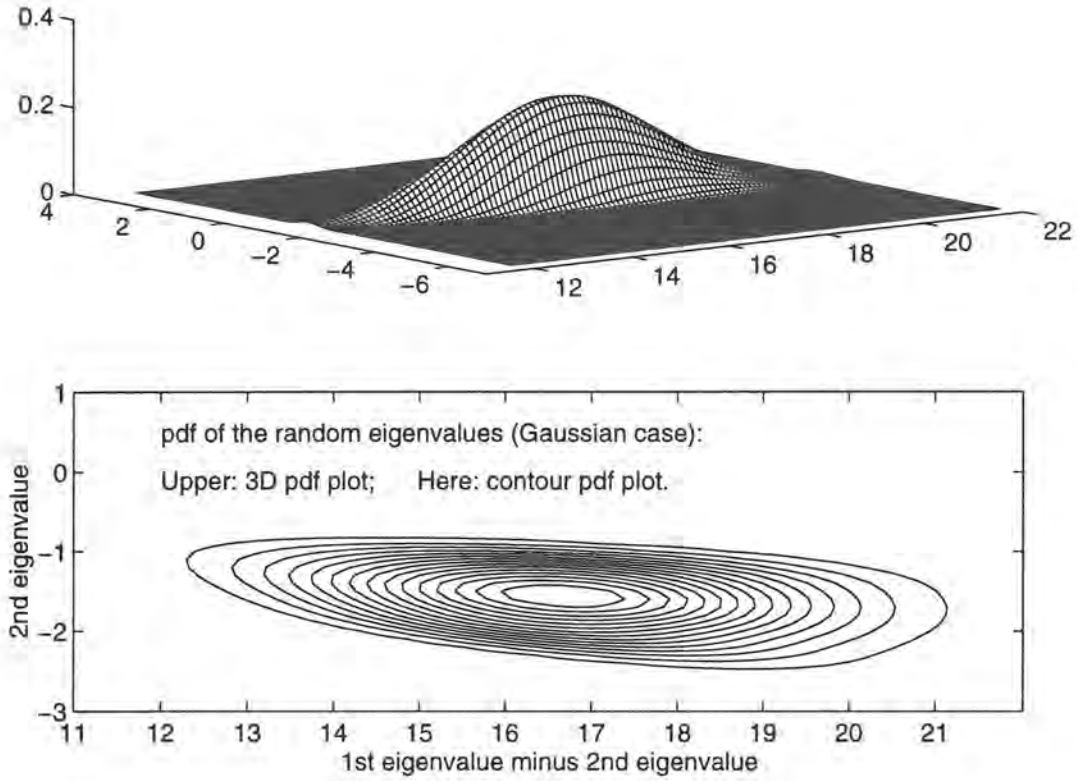


Figure 1: The probability density function of the two random principal stresses of the 3D deviatoric random stress tensor (after Xu 1999).

5 Accuracy and biases of the random eigenspectra

Although 3D SRS stress/strain tensors are practically random, the statistical issue of the random principal stress/strain components and the random orientations of the principal axes has been paid much attention only recently. The variance-covariance matrix of the three random eigenvalues $\boldsymbol{\lambda}$ and three random rotations $\boldsymbol{\Phi}$ with the first order approximation was first hinted at by Angelier et al. (1982) and then further systematically (and independently) worked out by Soler & van Gelder (1991). Since the ratio of stress/strain signal to noise is generally small (see, *e.g.* Amadei & Stephasson 1997), the first-order accuracy estimate can be significantly

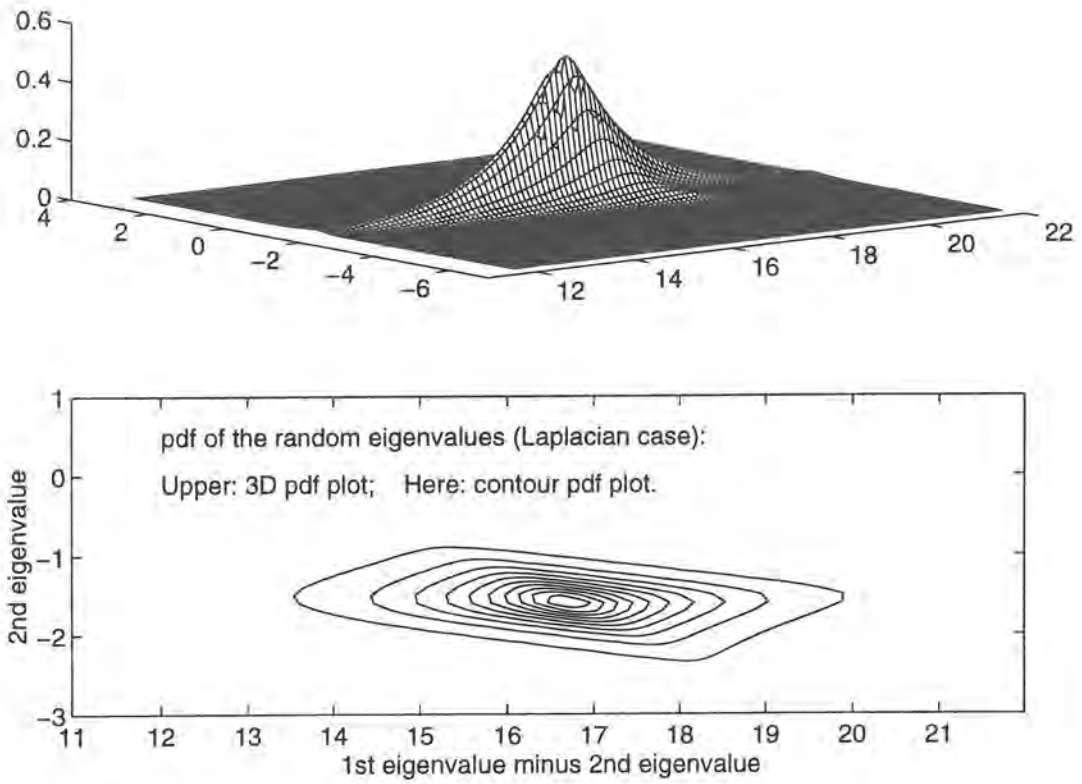


Figure 2: The probability density function of the two random principal stresses of the 3D deviatoric random stress tensor (after Xu 1999).

in error (Xu 1986). Xu & Grafarend (1996a, b) extended the first-order variance-covariance matrix to the second-order approximation. On the other hand, the one-to-one mapping between the (constrained or unconstrained) random tensor $\mathbf{T}$, the eigenvalues $\boldsymbol{\lambda}$ and rotations $\boldsymbol{\Phi}$ is nonlinear; thus biases of the estimated random eigenspectra are expected, the extent of which depends on the signal-to-noise ratio of the original random tensor $\mathbf{T}$. Surprinsingly, it seems that the bias issue of the random eigenspectra, although very important geophysically, has not been derived mathematically. Note, however, that some inequality results have been obtained by Cacoullos (1965), for example. Recently, the biases of the random eigenspectra with second order approximation were worked out by Xu (1996) and Xu & Grafarend (1996a, b).
In this section, we will further extend the statistical measures for 3D SRS random stress/strain tensors by Xu & Grafarend (1996a, b) to the unconstrained n-D case. The same technique can be applied to derive all the corresponding accuracy and biases of the random eigenspectra of 3D constrained random stress/strain tensors; this will not be discussed in this paper however. In principle, the accuracy and biases of the random eigenspectra can be derived using the estimation methods in nonlinear models (see *e.g.* Bates & Watts 1980, 1988; Beale 1960; Box 1971; Clarke 1980; Ratkowsky 1983; Seber & Wild 1989). The spectral decomposition (5) or (10) is very special, though nonlinear, in that there exist no redundant measurements. Thus we can directly derive the accuracy (of first- or second-order) and biases of the random eigenspectra on the basis of spectral decomposition with or without constraints.

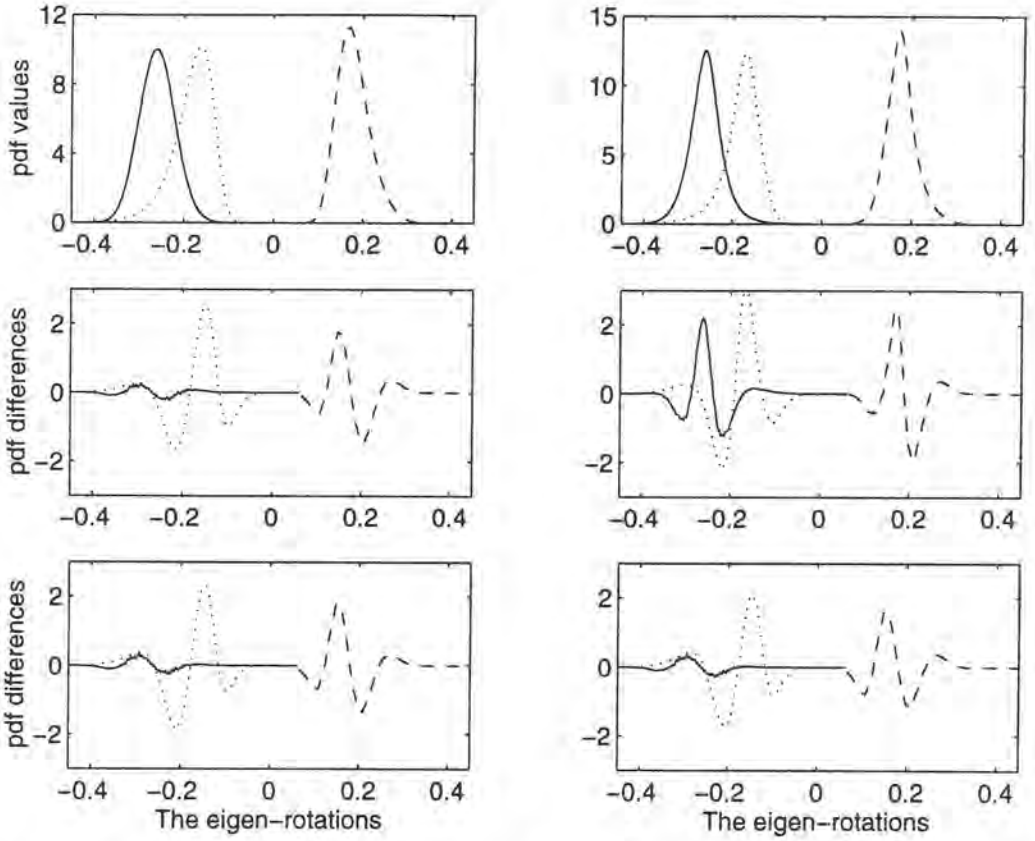


Figure 3: The marginal probability density functions of the three random rotations of the deviatoric random stress tensor (top two subplots), and the differences from their normal (middle two subplots) and Fisher's approximations (lower two subplots): solid line – θ_{21}; dashed line – θ_{31}; and dotted line – θ_{32}. The subplots on the left hand side are with the Gaussian model, and those on the right with the Laplacian model (after Xu 1999).

5.1 The biases of the random eigenspectra

Suppose that the n-D real SRS random tensor $\mathbf{T}$ has the mean $\mathbf{\Gamma}$ and the error $\boldsymbol{\epsilon}_T$, i.e.

$$\mathbf{T} = \mathbf{\Gamma} + \boldsymbol{\epsilon}_T. \tag{43}$$

The vector form of (43) is

$$v(\mathbf{T}) = v(\mathbf{\Gamma}) + \boldsymbol{\epsilon}, \tag{44}$$

where $\boldsymbol{\epsilon} = v(\boldsymbol{\epsilon}_T)$. It is further assumed that $\boldsymbol{\epsilon}$ has a vector of zero mean and a variance-covariance matrix $\mathbf{\Sigma}_\epsilon$. If the tensor $\mathbf{T}$ is constrained, then (44) only collects those functionally independent random components of $\mathbf{T}$. Let $\mathbf{T}$ have the random eigenvalues $\boldsymbol{\lambda}$ and random rotations $\mathbf{\Phi}$, which are functionally independent. Thus we have the following mapping:

$$\begin{bmatrix} \boldsymbol{\lambda} \\ \mathbf{\Phi} \end{bmatrix} = \rho\{v(\mathbf{T})\}, \tag{45}$$

where ρ maps $\mathbf{T}$ to $\boldsymbol{\lambda}$ and $\mathbf{\Phi}$.
Expanding (45) into a Taylor series and taking all the terms up to the second order approximation at the point $v(\mathbf{\Gamma})$, we obtain

$$\begin{bmatrix} \boldsymbol{\lambda} \\ \mathbf{\Phi} \end{bmatrix} = \rho\{v(\mathbf{\Gamma})\} + \dot{\rho}\{v(\mathbf{\Gamma})\}\boldsymbol{\epsilon} + \frac{1}{2}\mathbf{H}_\epsilon\boldsymbol{\epsilon}, \tag{46}$$

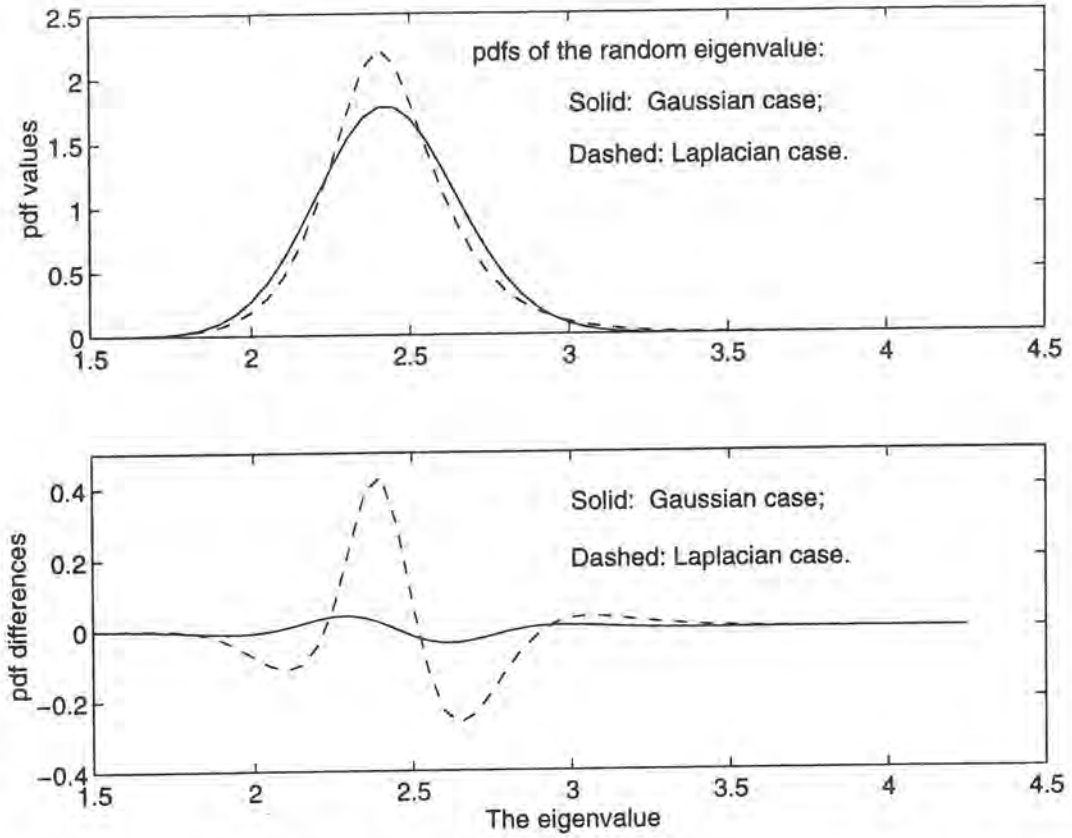


Figure 4: The pdfs of the random eigenvalue of the 3D pure shear random tensor and the differences from their normal approximations (after Xu 1999).

where

$$\dot{\rho}\{v(\mathbf{\Gamma})\} = \left.\frac{\partial(\boldsymbol{\lambda}^T, \boldsymbol{\Phi}^T)^T}{\partial[v(\mathbf{T})]^T}\right|_{\mathbf{T}=\mathbf{\Gamma}}, \tag{47}$$

$$\mathbf{H}_\epsilon = [\ddot{\mathbf{V}}(\lambda_1)\epsilon,\ ...,\ \ddot{\mathbf{V}}(\lambda_i)\epsilon,\ \ddot{\mathbf{V}}(\phi_{kl})\epsilon,\ ...,\ \ddot{\mathbf{V}}(\phi_{mn})\epsilon]^T, \tag{48}$$

with

$$\ddot{\mathbf{V}}(\lambda_i) = \left.\frac{\partial^2 \lambda_i}{\partial v(\mathbf{T})\partial[v(\mathbf{T})]^T}\right|_{\mathbf{T}=\mathbf{\Gamma}},$$

and

$$\ddot{\mathbf{V}}(\phi_{ij}) = \left.\frac{\partial^2 \phi_{ij}}{\partial v(\mathbf{T})\partial[v(\mathbf{T})]^T}\right|_{\mathbf{T}=\mathbf{\Gamma}},$$

being the symmetric matrix of the second derivatives of λ_i and ϕ_{ij} with respect to the elements of $v(\mathbf{T})$, respectively, where the total number of ϕ_{ij} is equal to $[n(n+1)/2 - i - m_c]$. Thus the biases of $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$ can be computed from (46) as follows:

$$\begin{aligned} bias\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\Phi}\end{bmatrix} &= E\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\Phi}\end{bmatrix} - \rho\{v(\mathbf{\Gamma})\} \\ &= \frac{1}{2}E\{\mathbf{H}_\epsilon\epsilon\}. \end{aligned} \tag{49}$$

In order to calculate the biases of $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$, we have to know the second derivatives $\dot{\rho}\{v(\mathbf{\Gamma})\}$, $\ddot{\mathbf{V}}(\lambda_i)$ and $\ddot{\mathbf{V}}(\phi_{ij})$. For a full n-D SRS random tensor $\mathbf{T}$, we have from (16) the first derivative

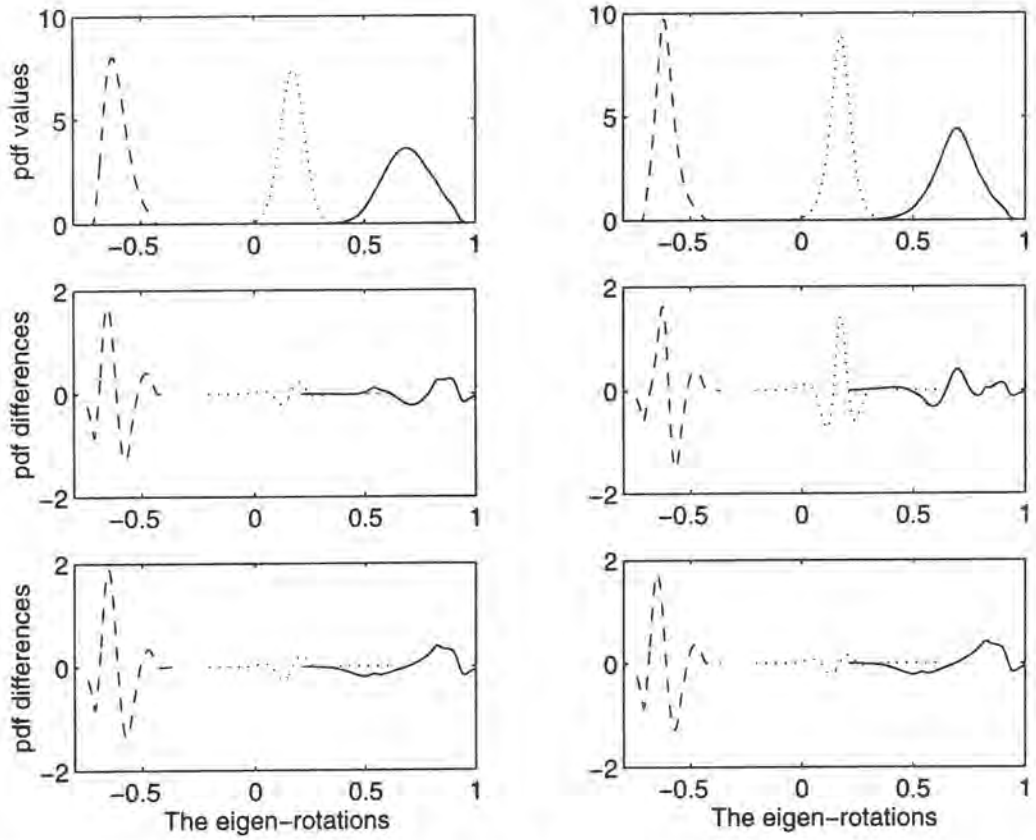


Figure 5: The marginal pdfs of the three random rotations of the 3D pure shear random tensor (top two subplots), and the differences from their normal (middle two subplots) and Fisher's approximations (lower two subplots): solid line – θ_{21}; dashed line – θ_{31}; and dotted line – θ_{32}. The subplots on the left hand side are with the Gaussian model and those on the right with the Laplacian model (after Xu 1999).

of $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$ with respect to $v(\mathbf{T})$:

$$\dot{\rho}\{v(\boldsymbol{\Gamma})\} = \left.\frac{\partial(\boldsymbol{\lambda}^T, \boldsymbol{\Phi}^T)^T}{\partial[v(\mathbf{T})]^T}\right|_{\mathbf{T}=\boldsymbol{\Gamma}} = \mathbf{B}^{-1}[\mathbf{D}_n^+(\mathbf{G}\otimes\mathbf{G})\mathbf{D}_n]^T\Big|_{\mathbf{T}=\boldsymbol{\Gamma}}, \tag{50}$$

where $\mathbf{B}$ is given by

$$\mathbf{B} = [v(\mathbf{O}_1), v(\mathbf{O}_2), ..., v(\mathbf{O}_n), v(\mathbf{H}_{21}), ..., v(\mathbf{H}_{n1}), v(\mathbf{H}_{32}), ..., v(\mathbf{H}_{n(n-1)})],$$

with $\mathbf{O}_i$ being a zero matrix except for the ith diagonal element taking on the unit value. In the 3D case, the matrix $\mathbf{B}$ becomes:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2-\lambda_1 & 0 & (\lambda_2-\lambda_1)\sin\phi_{31} \\ 0 & 0 & 0 & 0 & (\lambda_3-\lambda_1)\cos\phi_{21} & (\lambda_1-\lambda_3)\sin\phi_{21}\cos\phi_{31} \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (\lambda_3-\lambda_2)\sin\phi_{21} & (\lambda_3-\lambda_2)\cos\phi_{21}\sin\phi_{31} \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

By differentiating (50) and after some lengthy derivation, we obtain the second derivatives of $\boldsymbol{\lambda}$:

$$\ddot{\mathbf{V}}(\lambda_i) = \ddot{\mathbf{V}}_{\rho\lambda_i}\mathbf{B}^{-1}[\mathbf{D}_n^+(\mathbf{G}\otimes\mathbf{G})\mathbf{D}_n]^{-1}\,|_{\mathbf{T}=\boldsymbol{\Gamma}}, \tag{51}$$

and the second derivatives of $\mathbf{\Phi}$:

$$\ddot{\mathbf{V}}(\phi_{ij}) = \ddot{\mathbf{V}}_{\rho\phi_{ij}}\mathbf{B}^{-1}[\mathbf{D}_n^+(\mathbf{G}\otimes\mathbf{G})\mathbf{D}_n]^{-1}\mid_{\mathbf{T}=\mathbf{\Gamma}}, \tag{52}$$

where

$$\begin{aligned}
&\{\ddot{\mathbf{V}}_{\rho\lambda_1}, \ldots, \ddot{\mathbf{V}}_{\rho\lambda_n}, \ddot{\mathbf{V}}_{\rho\theta_{21}}, \ldots, \ddot{\mathbf{V}}_{\rho\theta_{n1}}, \ddot{\mathbf{V}}_{\rho\theta_{32}}, \ldots, \ddot{\mathbf{V}}_{\rho\theta_{n(n-1)}}\}^T \\
= &\{vec[\ddot{\mathbf{V}}_1^T], \ldots, vec[\ddot{\mathbf{V}}_n^T], vec[\ddot{\mathbf{V}}_{21}^T], \ldots, vec[\ddot{\mathbf{V}}_{n1}^T], vec[\ddot{\mathbf{V}}_{32}^T], \ldots, vec[\ddot{\mathbf{V}}_{n(n-1)}^T]\},
\end{aligned}$$

$$\ddot{\mathbf{V}}_i = -\mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\lambda_i}\mathbf{B}^{-1}[\mathbf{D}_n^+(\mathbf{G}\otimes\mathbf{G})\mathbf{D}_n]^{-1}, \quad i = 1, 2, \ldots, n$$

$$\begin{aligned}
\ddot{\mathbf{V}}_{ij} = &-\mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\phi_{ij}}\dot{\rho}\{v(\mathbf{T})\} \\
&-\dot{\rho}\{v(\mathbf{T})\}\mathbf{D}_n^+[(\frac{\partial\mathbf{G}}{\partial\phi_{ij}}\otimes\mathbf{G}) + (\mathbf{G}\otimes\frac{\partial\mathbf{G}}{\partial\phi_{ij}})]\mathbf{D}_n[\mathbf{D}_n^+(\mathbf{G}\otimes\mathbf{G})\mathbf{D}_n]^{-1}, \quad i > j.
\end{aligned}$$

Substituting (51) and (52) into (49), we can readily compute the biases of the random eigenvalues and random rotations. For more details in the 3D case, the reader is referred to Xu & Grafarend (1996b). In the same manner and using the differential relations in Section 3, we can first derive all the required second derivatives and then compute the biases of the random eigenvalues and random rotations of the 3D random deviatoric stress/strain and pure shear tensors; these are omitted here.

5.2 The accuracy of the random eigenspectra

Omitting the second order term of $\boldsymbol{\epsilon}$ from (46) and then applying the error propagation law to it, we obtain the variance-covariance matrix of $\boldsymbol{\lambda}$ and $\mathbf{\Phi}$ as follows:

$$D\begin{bmatrix}\boldsymbol{\lambda}\\ \mathbf{\Phi}\end{bmatrix} = \dot{\rho}\{v(\mathbf{\Gamma})\}\mathbf{\Sigma}_\epsilon\dot{\rho}^T\{v(\mathbf{\Gamma})\}, \tag{53}$$

which is of a first order approximation (see also Angelier et al. 1982; Soler & van Gelder 1991; Xu & Grafarend 1996b).

Since the ratio of signal to noise of random stress/strain tensors in the Earth Sciences may be small, the accuracy of higher order approximation may be necessary. In what follows we will derive the second order accuracy estimate. Applying the definition of variance-covariance to (46), we have

$$D\begin{bmatrix}\boldsymbol{\lambda}\\ \mathbf{\Phi}\end{bmatrix} = \dot{\rho}\{v(\mathbf{\Gamma})\}\mathbf{\Sigma}_\epsilon\dot{\rho}^T\{v(\mathbf{\Gamma})\} + \frac{1}{4}E\{\mathbf{H}_\epsilon\boldsymbol{\epsilon} - E(\mathbf{H}_\epsilon\boldsymbol{\epsilon})\}\{\mathbf{H}_\epsilon\boldsymbol{\epsilon} - E(\mathbf{H}_\epsilon\boldsymbol{\epsilon})\}^T. \tag{54}$$

The second term of (54) can be derived using results of the fourth order moments of quadratic statistics (Searle 1971; Rao & Kleffe 1988; Searle et al. 1992), and is given by

$$E\{\mathbf{H}_\epsilon\boldsymbol{\epsilon} - E(\mathbf{H}_\epsilon\boldsymbol{\epsilon})\}\{\mathbf{H}_\epsilon\boldsymbol{\epsilon} - E(\mathbf{H}_\epsilon\boldsymbol{\epsilon})\}^T = 2\mathbf{M}_{H_\epsilon\epsilon}, \tag{55}$$

where

$$\mathbf{M}_{H_\epsilon\epsilon} = \begin{bmatrix}\mathbf{M}_\lambda & \mathbf{M}_{\lambda\Phi}\\ \mathbf{M}_{\lambda\Phi}^T & \mathbf{M}_\Phi\end{bmatrix}, \tag{56}$$

with

$$\mathbf{M}_\lambda = [tr\{\ddot{\mathbf{V}}(\lambda_i)\mathbf{\Sigma}_\epsilon\ddot{\mathbf{V}}(\lambda_j)\mathbf{\Sigma}_\epsilon\}]_{n\times n},$$

$$\mathbf{M}_{\Phi} = [tr\{\ddot{\mathbf{V}}(\theta_{ij})\mathbf{\Sigma}_{\epsilon}\ddot{\mathbf{V}}(\theta_{ij})\mathbf{\Sigma}_{\epsilon}\}]_{n\times n},$$
$$\mathbf{M}_{\lambda\Phi} = [tr\{\ddot{\mathbf{V}}(\lambda_i)\mathbf{\Sigma}_{\epsilon}\ddot{\mathbf{V}}(\theta_{ij})\mathbf{\Sigma}_{\epsilon}\}]_{n\times n}.$$

Substituting (56) into (54), we obtain the variance-covariance matrix of $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$ with the second order approximation as follows:

$$D\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\Phi}\end{bmatrix} = \dot{\rho}\{v(\mathbf{\Gamma})\}\mathbf{\Sigma}_{\epsilon}\dot{\rho}^T\{v(\mathbf{\Gamma})\} + \frac{1}{2}\mathbf{M}_{H_{\epsilon}\epsilon}. \tag{57}$$

The accuracy of the random principal stress/strain components and random rotations of a full 3D random stress/strain tensor can be found in Xu & Grafarend (1996b). Using the same techniques as in the above, one can also obtain the first- and second-order accuracies of the random eigenvalues and random rotations of the 3D random deviatoric stress/strain and pure shear tensors; they are also omitted here however.
We finally summarize the bias and accuracy results of the random spectra in the following theorem.

Theorem 4 *Let* $\mathbf{T}$ *be a real n-D SRS random tensor, whose random eigenvalues and random rotation parameters are respectively denoted by* $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)$ *with its elements satisfying* $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ *and* $\boldsymbol{\Phi} = (\phi_{21}, \phi_{31}, ..., \phi_{n(n-1)})$. *Assume that* $\mathbf{T}$ *(or more specifically,* $v(\mathbf{T})$*) has the mean* $v(\mathbf{\Gamma})$ *and the variance-covariance matrix* $\mathbf{\Sigma}_{\epsilon}$. *Then the biases and variance-covariance matrix of the second-order approximation of the random eigenvalues and random rotations are given by*

$$bias\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\Phi}\end{bmatrix} = \frac{1}{2}E\{\mathbf{H}_{\epsilon}\boldsymbol{\epsilon}\}, \tag{58}$$

and

$$D\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\Phi}\end{bmatrix} = \dot{\rho}\{v(\mathbf{\Gamma})\}\mathbf{\Sigma}_{\epsilon}\dot{\rho}^T\{v(\mathbf{\Gamma})\} + \frac{1}{2}\mathbf{M}_{H_{\epsilon}\epsilon}, \tag{59}$$

respectively, where the matrix $\mathbf{H}_{\epsilon}$ *of (58) is given in Subsection 5.1.*

6 References

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology*, W.H. Freeman, San Francisco.

Altmann, S.L., 1986. *Rotations, Quaternions, and Double Groups*, Clarendon Press, Oxford.

Amadei, B. & Stephansson, O., 1997. *Rock Stress and its Measurement*, Chapman & Hall, London.

Anderson, T.W., 1958. *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.

Angelier, J., 1984. Tectonic analysis of fault slip data sets, *J. geophys. Res.*, **B89**, 5835-5848.

Angelier, J., 1989. From orientations to magnitudes in paleostress determinations using fault slip data, *J. struct. Geol.*, **11**, 37-50.

Angelier, J., Tarantola, A., Vallete, B. & Manoussis S., 1982. Inversion of field data in fault tectonics to obtain the regional stress — I. single phase fault populations: a new method of computing the stress tensor, *Geophys. J. R. astr. Soc.*, **69**, 607-621.

Bates, D.M. & Watts, D.G., 1980. Relative curvature measures of nonlinearity (with discussions), *J.R. Statist. Soc.*, **B42**, 1-25

Bates, D.M. & Watts, D.G., 1988. *Nonlinear Regression Analysis and its Applications*, John Wiley & Sons, New York

Beale, E.M.L., 1960. Confidence regions in non-linear estimation (with discussions), *J.R. Statist. Soc.*, **B22**, 41-88

Boutet de Monvel, A., Khorunzhy, A. & Vasilchuk, V., 1996. Limiting eigenvalue distribution of random matrices with correlated entries, *Markov Proc. and Related Fields*, **2**, 607-636.

Box, M.J., 1971. Bias in nonlinear estimation (with discussions), *J.R. Statist. Soc.*, **B33**, 171-201

Brekhovskikh, L.M. & Goncharov, V., 1994. *Mechanics of Continua and Wave Dynamics*, 2nd ed., Springer Verlag, Berlin.

Cacoullos, T., 1965. On the bias of functions of characteristic roots of a random matrix, *Biometrica*, **52**, 87-94.

Cichowicz, A., 1993. An automatic S-phase picker, *Bull. seis. Soc. Amer.*, **83**, 180-189.

Clarke, G.P.Y., 1980. Moments of the least squares estimators in a non-linear regression model, *J.R. Statist. Soc.*, **B42**, 227-237

Dai, H. & MacBeth, C., 1997. Application of back-propagation neural networks to identification of seismic arrival types, *Phys. Earth planet. Interior*, **101**, 177-188.

Fang, K.T., Kotz, S. & Ng, K.W., 1990. *Symmetric Multivariate and Related Distributions*, John Wiley & Sons, New York.

Fisher, N.I., Lewis, T. & Embleton, B.J.J., 1993. *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge.

Fisher, R.A., 1953. Dispersion on a sphere, *Proc. R. Soc.*, **A217**, 295-305.

Frank, F.C., 1966. Deduction of earth strains from survey data, *Bull. seismol. Soc. Am.*, **56**, 35-42.

Furstenberg, H. & Kesten, H., 1960. Products of random matrices, *Ann. math. Statist.*, **31**, 457-469.

Gephart, J. & Forsyth, D., 1984. An improved method for determining the regional stress tensor using earthquake focal mechanism data: application to the San Fernando earthquake sequence, *J. geophys. Res.*, **B89**, 9305-9320.

Girko, V.L., 1979. Distribution of eigenvalues and eigenvectors of hermitian stochastic matrices, *Ukrain. Math. J.*, **31**, 421-424

Girko, V.L., 1985. Spectral theory of random matrices, *Russian math. Surveys*, **40**, 77-120.

Girko, V.L., 1989. Asymptotic of the distribution of the spectrum of random matrices, *Russian math. Surveys*, **44**, 3-36.

Girko, V.L., 1990a. Limit theorems for distributions of eigenvalues of random symmetrical matrices, *Theory Prob. math. Statist.*, **41**, 25-31.

Girko, V.L., 1990b. *Theory of Random Determinants*, Kluwer Academic Publishers, Amsterdam.

Girko, V.L., 1992. Asymptotic behavior of the eigenvalues of empirical covariance matrices, (part 1), *Theory Prob. math. Statist.*, **44**, 37-44.

Girko, V.L., 1993. Asymptotic behavior of the eigenvalues of empirical covariance matrices, (part 2), *Theory Prob. math. Statist.*, **46**, 33-37.

Girko, V.L., 1997. Strong law for the singular values and eigenvectors of random matrices, *Random Oper. Stoch. Equ.*, **5**, 81-104.

Grafarend, E., 1986. Three-dimensional deformation analysis: Global vector spherical harmonic and local element representation, *Tectonophysics*, **130**, 337-359.

Hansen, K. & Mount, V., 1990. Smoothing and extrapolation of crustal stress orientation measurements, *J. geophys. Res.*, **B95**, 1155-1165.

Helffrich, G.R., 1997. How good are routinely determined focal mechanisms? Empirical statistics based on a comparison of Harvard, USGS and ERI moment tensors, *Geophys. J. Int.*, **131**, 741-750.

Horiuchi, S., Rocco, G. & Hasegawa, A., 1995. Discrimination of fault planes from auxiliary planes based on simultaneous determination of stress tensor and a large number of fault plane solutions, *J. geophys. Res.*, **B100**, 8327-8338.

Johnson, N.L. & Kotz, S., 1970. *Distributions in Statistics: Continuous Univariate Distributions*, John Wiley & Sons, New York.

Johnson, N.L. & Kotz, S., 1972. *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley & Sons, New York.

Kagan, Y.Y., 1987. Point sources of elastic deformation: elementary sources, static displacements, *Geophys. J. R. astr. Soc.*, **90**, 1-34.

Kagan, Y.Y., 1990. Random stress and earthquake statistics: spatial dependence, *Geophys. J. Int.*, **102**, 573-583.

Kagan, Y.Y., 1991. Seismic moment distribution, *Geophys. J. Int.*, **106**, 123-134.

Kagan, Y.Y., 1992a. On the geometry of an earthquake fault system, *Phys. Earth planet. Interior*, **71**, 15-35.

Kagan, Y.Y., 1992b. Correlations of earthquake focal mechanisms, *Geophys. J. Int.*, **110**, 305-320.

Kagan, Y.Y., 1994. Observational evidence from earthquakes as a nonlinear dynamic process, *Physica D*, **77**, 160-192.

Kagan, Y.Y. & Knopoff, L., 1985a. The first order statistical moment of the seismic moment tensor, *Geophys. J. R. astr. Soc.*, **81**, 429-444.

Kagan, Y.Y. & Knopoff, L., 1985b. The two-point correlation of the seismic moment tensor, *Geophys. J. R. astr. Soc.*, **83**, 637-656.

Kostrov, B.V. & Das, S., 1988. *Principles of Earthquakes Source Mechanics*, Cambridge University Press, Cambridge.

Larson, H.L. & Shubert, B.O., 1979. *Probabilistic Models in Engineering Sciences*, John Wiley & Sons, New York.

Lay, T. & Wallace, T.C., 1995. *Modern Global Seismology*, Academic Press, New York.

Lu, Z., Wyss, M & Pulpan, H., 1997. Details of stress directions in the Alaska subduction zone from fault plane solutions, *J. geophys. Res.*, **B102**, 5385-5402.

Magnus, J. & Neudecker, H., 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, New York.

Mehta, M.L., 1990. *Random Matrices and the Statistical Theory of Energy Levels, (extended edition)*, Academic Press, New York.

McKenzie, D.P, 1969. The relation between fault plane solutions for earthquakes and the directions of the principal stresses, *Bull. seis. Soc. Amer.*, **59**, 591-601.

Mood, A., Graybill, F. & Boes, D., 1974. *Introduction to the Theory of Statistics*, McGRAW-Hill, Tokyo.

Oravecz, F. & Petz, D., 1997. On the eigenvalue distribution of some symmetric random matrices, *Acta Sci. Math.*, **63**, 383-395.

Prescott, W.H., 1981. The determination of displacement fields from geodetic data along a strike slip fault, *J. geophys. Res.*, **B86**, 6067-6072.

Rao, C.R. & Kleffe, J., 1988. *Estimation of Variance Components and Applications*, North Holland, Amsterdam

Ratkowsky, D.A., 1983. *Nonlinear Regression Modeling: A Unified Practical Approach*, Marcel Dekker, Inc., New York

Rocco, G., Horiuchi, S. & Hasegawa, A., 1997. Study on the state of stress around earthquake swarm areas using P-wave polarity data: Nikko-Ashio and Teishi Knoll areas, *J. Phys. Earth*, **45**, 29-40.

Samson, J.C., 1977. Matrix and Stokes vector representations of detectors for polarized waveforms: theory, with some applications to teleseismic waves, *Geophys. J. R. astr. Soc.*, **51**, 583-603.

Searle, S.R., 1971. *Linear Models*, John Wiley & Sons, New York

Savage, J.C. & Burford, R.O., 1970. Accumulation of tectonic strain in California, *Bull. seismol. Soc. Am.*, **60**, 1877-1896.

Searle, S.R., Casella, G. & McCullach, C., 1992. *Variance Components*, John Wiley & Sons, New York

Seber, G. & Wild, C., 1989. *Nonlinear Regression*, John Wiley & Sons, New York

Soler T. & van Gelder B., 1991. On covariances of eigenvalues and eigenvectors of second rank symmetric tensors, *Geophys. J. Int.*, **105**, 537-546.

Stroud, A.H., 1971. *Approximate Calculation of Multiple Integrals*, Prentice-Hall, New Jersey.

Wyss, M., Liang, B.Y., Tanagawa, W. & Wu, X., 1992. Comparison of orientations of stress and strain tensors based on fault plane solutions in Kaoiki, Hawaii, *J. geophys. Res.*, **B97**, 4769-4790.

Xu, P.L., 1986. Variance-covariance propagation for a nonlinear function, *J. Wuhan Techn. Uni. Surv. Mapping*, **11**, No.2, 92-99 (in Chinese with English abstract).

Xu, P.L., 1996. Statistical aspects of 2-D symmetric random tensors, Proc. FIG int. Symp. Deformation Measurement, Hong Kong, pp.203-210.

Xu, P.L., 1999. Spectral theory of constrained second-rank symmetric random tensors, *Geophys. J. Int.*, **138**, 1-24.

Xu, P.L. & Grafarend, E., 1996a. Probability distribution of eigenspectra and eigendirections of a two-dimensional, symmetric rank two random tensor, *J. Geod.*, **70**, 419-430.

Xu, P.L. & Grafarend, E., 1996b. Statistics and geometry of the eigenspectra of 3-D second-rank symmetric random tensors, *Geophys. J. Int.*, **127**, 744-756.

Xu, P.L. & Rummel, R., 1994. A simulation study of smoothness methods in recovery of regional gravity fields, *Geophys. J. Int.*, **117**, 472-486.

Xu, P.L. & Shimada, S., 1997. Inferring the relative measure of principal stress components, *Geophys. J. Int.*, **130**, F1-F4.

Geodesy beyond 2000: An attempt to unify geodesy by the geodesic flow in all branches

Rey Jer You

Abstract

According to the idea of Prof. Grafarend, Department of Geodetic Science, Stuttgart University, we attempt to present a unified theory of geodesy by the geodesic flows. In the paper, the plumblines related to the orthometric heights, dynamics of satellite motions, signal transit path in the satellite geodesy, and the geodetic main problems in the geometric geodesy are uniformly interpreted as geodesic flows. Three fundamental elements in the unified theory of geodesy are discussed: First, the interpretation of geodesy problems as geodesic flows. Second, the definition the geodesic manifolds. Third, the determination of the embedding spaces.

Keywords: geodesic flow, metrics, variational principle, embedding spaces

1. Introduction

Geodesics, in particular minimal geodesics, are of focal geodetic interest. For example, the differential equations of geodesic flows are used for solving the geodetic main problems on the ellipsoid of revolution. Hotine (1966) and Marussi (1985, pp.169-172) used the Maupertuis-Euler-Lagrange-Jacobi variational principle of least action to describe the plumbline as geodesics. Some other related discussions in geodesy can be also found in e.g. Heitz, 1988, Moritz, 1994, Schwarz et al., 1993, You, 1995, Zund, 1994. They focused to describe some problems in geodesy as geodesic problems. The plumb line, trajectories of satellite orbital motions and the propagation paths of electromagnetic waves/light rays can be interpreted as geodesics according to their studies. Grafarend presented the unified concept of geodesy in 1973 and then try to establish the unified theory of geodesy (1973, 1989, 1994, 1995, 1997) such as in physics where the gravitation, magnetic force and electronic force etc. are described in a same mathematical form, a unified theory in physics. The central point of the unified theory is the geodesics. In the paper, we try to describe some problems in geodesy by using the same mathematical form, i.e. a unified theory.

2. Lagrange portray, Halmilton portray and Newton form of geodesics

Lagrange portray: If the Riemannian metric of general form

$$ds^2 = g_{\alpha\beta}\, dq^\alpha dq^\beta , \qquad \alpha,\beta \in \{1,2,\cdots,n\} \tag{1}$$

exists, the differential equations of the geodesic in the Riemannian manifold can be derived by the Euler-Lagrange variational principle for the fixed boundary points

$$\delta \int L\, ds = 0$$
$$L = \sqrt{g_{\alpha\beta} \frac{dq^\alpha}{ds} \frac{dq^\beta}{ds}} \tag{2}$$

The sufficient and necessary conditions

$$\frac{\partial L}{\partial q^\alpha} - \frac{d}{ds}\left(\frac{\partial L}{\partial \frac{dq^\alpha}{ds}} \right) = 0 \tag{3}$$

must be satisfied and then lead to the differential equations of the geodesic

$$\frac{d^2 q^\alpha}{ds^2} + \left\{ \begin{matrix} \alpha \\ \beta \quad \gamma \end{matrix} \right\} \frac{dq^\beta}{ds} \frac{dq^\gamma}{ds} = 0, \qquad \alpha, \beta, \gamma \in \{1,2,\cdots,n\} \tag{4}$$

Hamilton portray: One can describe the differential equations of geodesics as a system of differential equations of first order in phase space by using the Hamilton principle for the fixed points:

$$\delta \int (p_\alpha \frac{dq^\alpha}{ds} - \mathrm{H})\, ds = 0$$
$$H = \frac{1}{2} g^{\alpha\beta} p_\alpha p_\beta \tag{5}$$

where p is the general momentum. A minimal geodesic in phase space is now represented by

$$\frac{dq^\alpha}{ds} = \frac{\partial H}{\partial p_\alpha} = g^{\alpha\beta} p_\beta$$
$$\frac{dp_\alpha}{ds} = -\frac{\partial H}{\partial q^\alpha} = -\frac{1}{2} \frac{\partial g^{\beta\gamma}}{\partial q^\alpha} p_\beta p_\gamma \tag{6}$$

If the Riemannian manifold is characterized by conformal coordinates, the metric becomes

$$ds^2 = e^{2\sigma} \delta_{\alpha\beta}\, dq^\alpha dq^\beta, \qquad \alpha, \beta \in \{1,2,\cdots,n\} \tag{7}$$

Then, a geodesic flow takes the form of the Newton law (Grafarend et al. 1995)

$$\frac{d^2 q^\alpha}{dt^2} - \frac{1}{2} \frac{\partial e^{2\sigma}}{\partial q^\alpha} = 0 \tag{8}$$

if the arc length s is replaced at the dynamic time t according to the Maupertuis gauge

$$ds = e^{2\sigma}\, dt \tag{9}$$

The factor of conformality $e^{2\sigma}$ is the source of the "conservative force field".

3. Some metrics in geodesy

In this section, we summarize some metrics that are often used in the geometric geodesy, physical geodesy and satellite geodesy.

(1) metric of an ellipsoid of revolution

a. in geographic coordinates
(Manifold: ellipsoid)

$$ds^2 = M^2 d\phi^2 + N^2 \cos^2\phi \, d\lambda^2$$

where

$$M = \frac{a(1-e^2)}{(1-e^2\sin^2\phi)^{3/2}}$$

$$N = \frac{a}{(1-e^2\sin^2\phi)^{1/2}} \tag{10}$$

a : semi – major axis

e : eccentricity of ellipsoid

ϕ : geographic latitude

λ : geographic longitude

b. in conformal coordinates
(Manifold: ellipsoid)

$$ds^2 = k^2(y)(dx^2 + dy^2)$$

where (11)

$$x = a(\lambda - \lambda_0)$$

$$y = a\ln\left(\tan\left(\frac{\pi}{4}+\frac{\phi}{2}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right)$$

$$k^2(y) = 1 + c_1 y^2 + c_2 y^4 + \cdots$$

$c_1, c_2, \cdots$: constants (Yang, 1998)

(2) metric of the manifold of plumb line

a. Maupertuis metric (Grafarend et al., 1994, 1995)
(Manifold: conformally flat Riemannian manifold)

$$ds^2 = G^2(W)\left(dx^2 + dy^2 + dz^2\right)$$

x, y, z : 3D cartesian coordinates (12)

$G^2(W)$: any function of terrestrial potential W

b. Hotine metric (Hotine, 1966)
(manifold: conformally flat Riemannian manifold)

$$ds^2 = \| grad\, W \|^2 \; f^2(W)\left(dx^2 + dy^2 + dz^2\right) \tag{13}$$

$f(W)$: any function of terrestrial potential W

c. Marussi metric (Marussi, 1985, pp.169-172, Moritz, 1994)
(Manifold: conformally flat Riemannian manifold)

$$ds^2 = \| grad\, W \|^2 \left(dx^2 + dy^2 + dz^2\right) \tag{14}$$

(3) metric of the manifold of satellite orbits

a. Maupertuis metric of configuration space (Goenner et al., 1994, Ong, 1975, Synge, 1926, You, 1995)
(Manifold: conformally flat Riemannian manifold)

$$ds^2 = 2(E-V)\left(dx^2 + dy^2 + dz^2\right) \tag{15}$$

x, y, z : cartesian coordinates

V : axialsymmetric potential

E : conservative energy

b. Maupertuis metric of impulse space (Moser, 1970, You, 1995)
Manifold: conformally flat Riemannian manifold)

$$ds^2 = \frac{4}{(\vec{p}^2 - 2F)^2}\left((dp^1)^2 + (dp^2)^2 + (dp^3)^2\right) \tag{16}$$

p^1, p^2, p^3 : impulses

F : Kepler energy

c. Space-time metric (e.g. Misner et al., 1973)
(Manifold: pseudo Riemannian manifold)

$$ds^2 = (-1 - \frac{2V}{c^2} - \frac{2V^2}{c^4})c^2dt^2 + (1 - \frac{2V}{c^2})\left(dx^2 + dy^2 + dz^2\right) -$$
$$- \frac{\omega_1}{c^2}(cdt \cdot dx) - \frac{\omega_2}{c^2}(cdt \cdot dy) - \frac{\omega_3}{c^2}(cdt \cdot dz) \tag{17}$$

t, x, y, z : 4D harmonic coordinates

V : potential

c : velocity of light

$\vec{\omega}$: Thirring - Lense vertorial potential

(4) refraction metric

a. non-relativistic case (Born, 1980)
(Manifold: conformally flat Riemannian manifold)

$$ds^2 = n^2\left(dx^2 + dy^2 + dz^2\right)$$

x, y, z : cartesian coordinates (18)

n : index of refraction

b. relativistic case (non-dispersive) (Ehlers, 1967)
(Manifold: pseudo Riemannian manifold)

$$ds^2 = \overline{g}_{\mu\nu} dq^\mu dq^\nu$$

where

$$q^0 = ct,\ q^1 = x,\ q^2 = y,\ q^3 = z \tag{19}$$

$$\overline{g}_{\mu\nu} = g_{\mu\nu} + (1 - \frac{1}{n^2})\frac{u^\mu u^\nu}{c^2}$$

$g_{\mu\nu}$: space - time metric

4. Embedding onto the flat spaces

Consider the metric

$$ds^2 = g_{\alpha\beta}\, dq^\alpha dq^\beta\,, \qquad \alpha, \beta \in \{1,2,\cdots,n\} \tag{20}$$

of an n-dimensional Riemannian manifold. In order to get a better insight into the geometry of a geodesic and its corresponding manifold, it needs to embed the manifold into an m-dimensional flat space whose metric is

$$dS^2 = e_{ii}\, du^i du^i\,, \qquad i \in \{1,2,\cdots,m\} \tag{21}$$

In general, we take the locally isometric embedding theory to find the embedding space, i.e. $dS^2 = ds^2$. Then, the sufficient and necessary conditions

$$e_{ii}\frac{\partial u^i}{\partial q^\alpha}\frac{\partial u^i}{\partial q^\beta} = g_{\alpha\beta} \tag{22}$$

must be satisfied. After solving the above partial differential equations, the embedding space can be performed. But it is usually difficult to solve analytically this system of partial differential equations. Some authors use the transformation method as an alternative approach to find the embedding space, see e.g. Brinkmann, 1923, Goenner et al., 1994 and Rosen, 1965. One can easily obtain the embedding space of a conformally flat manifold by using the transformation of Brinkmann

According of the transformation methods, the embedding spaces of Marussi metric of plumblines in spherical symmetric fields are gives in the following:

(a) exterior space

$$\begin{aligned} ds^2 &= \frac{\mu}{r^4}^2 \left(dx^2 + dy^2 + dz^2\right) \\ &= \frac{\mu}{r^4}^2 (dr^2 + r^2 d\phi^2 + r^2 \cos\phi^2 d\lambda^2) \\ &= (du^1)^2 + (du^2)^2 + (du^3)^2 \end{aligned} \tag{23}$$

Embedding space $E^3(3,0)$:

$$\begin{aligned} u^1 &= \frac{\mu}{r}\cos\phi\ \cos\lambda \\ u^2 &= \frac{\mu}{r}\cos\phi\ \sin\lambda \\ u^3 &= \frac{\mu}{r}\sin\phi \end{aligned} \tag{24}$$

(b) interior space

$$\begin{aligned} ds^2 &= \frac{\mu}{R^4}^2 r^2\left(dx^2 + dy^2 + dz^2\right) \\ &= \frac{\mu}{R^4}^2 r^2(dr^2 + r^2 d\phi^2 + r^2 \cos\phi^2 d\lambda^2) \\ &= (du^1)^2 + (du^2)^2 + (du^3)^2 - (du^4)^2 \end{aligned} \tag{25}$$

R : constant radius

Embedding space $E^4(3,1)$:

$$\begin{aligned} u^1 &= \frac{\mu}{R^3} r^2 \cos\phi\ \cos\lambda \\ u^2 &= \frac{\mu}{R^3} r^2 \cos\phi\ \sin\lambda \\ u^3 &= \frac{\mu}{R^3} r^2 \sin\phi \\ u^4 &= \frac{\sqrt{3}}{2}\frac{\mu}{R^3} r^2 \end{aligned} \tag{26}$$

The embedding space of the Maupertuis manifold of satellite orbits can be found in Goenner (1994) and You (1998).

5. References

Born, M. and E. Wolf (1980): Principles of Optics, Pergamon, Oxford.

Brinkmann, H.W. (1923): On Riemann spaces conformal to Euclidean space. Proc. Nat. Acad. Sci. 9, pp.1-3.

Ehlers, J. (1967): Zum Uebergang von der Wellenoptik zur geometrischen Optik in der allgemeinen Relativitaetstheorie, Z. Naturf. 22A, pp.1328-1333.

Fischer, I. (1975): The figure of the earth-changes in concepts. Geophys. Surveys 2, pp. 3-54.

Grafarend, E.W. (1973): Le theoreme de conservation de la courbure et de la torsion or attempts for a unified theory of geodesy. Bull. Geod. 109, pp. 237-260.

Grafarend, E. W. (1989): Four lectures on special and general relativity. In: Lecture Notes in Earth Sciences 25, F. Sanso and R. Rummel (eds.): Theory of satellite geodesy and gravity field determination, Springer-Verlag, Berlin.

Grafarend, E. W. and R.J. You (1994): The embedding of the plumbline manifold: orthometric heights, III Hotine-Marussi Symp. On Math. Geod., 20. May-03. Jun. 1994, L'Aquila, Italy.

Grafarend, E.W., R. Syffus and R.J. You (1995): Projective heights in geometry and gravity space, AVN 10/1995, pp. 382-403.

Grafarend, E. W. and R.J. You (1995): The Newton form of a geodesic in Maupertuis gauge on the sphere and the biaxial ellipsoid – part one-. ZfV 120, pp. 68-80.

Goenner, H, E.W. Grafarend and R.J. You (1994): Newton mechanics as geodesic flow on Maupertuis' manifolds: the local isometric embedding into spaces, Manu. Geod. 19, pp. 339-345.

Heitz, S. (1988): Geodetische Hauptaufgaben in klassischen und relativistischen Modellen, DGK Reihe B 287.

Hotine, M. (1966): Geodetic applications of conformal transformation, Bull. Geod. 80, pp. 123-140.

Marussi, A. (1985): Intrinsic geodesy, pp.132, Springer-Verlag, Berlin.

Misner, C.W., K.S. Thorne, J.A. Wheeler (1973): Gravitation, W.H. Freeman and Company, New York.

Moritz, H. (1994): The Hamiltonian structure of refraction and of the gravity field, Manu. Geod. 20, pp.52-60.

Moser, J. (1970): Regularization of Kepler's problem and the averaging method on a manifold. Commun. On Pure and appl. Math. 23, pp.609-636.

Ong, C.P. (1975): Curvature and mechanics, Adv. Math, 15,pp. 269-311.

Rosen, J. (1965): Embedding of various relativistic Riemannian spaces in pseudo Euclidean spaces. Rev. mod. Phys. 37, pp. 204-214.

Schouten, J.A. (1954): Ricci-Calculus: An introduction to tensor analysis and its geometrical applications. Springer-Verlag, Berlin.

Schwarze, V.S., T. Hartmann, M. Leins, M.H. Soffel (1993): Relativistic effects in satellite positioning. Manu. Geod. 18, pp.306-316.

Synge, J.L. (1926): On the geometry of dynamics. Philo. Trans. Roy. Soc. London, A226, pp. 31-106.

Yang, D.W. (1998): A study on solving geodetic problems on the universal Mercator projection maps with the aid of the variational principle of least action, master thesis, Department of Surveying Engineering, National Cheng Kung University.

You, R.J. (1995): Zur analytischen Bahnberechnung kuenstlicher Erdsatelliten mittels konformer Transfromationen, DGK, Reihe C 440, Muenchen.

You, R.J. (1998): Geodesic motion of an Earth's artificial satellite in an axial-symmetrically gravitational field. Boll. Di geod. E Sci. Aff. LVII, pp. 257-274.

Zund, J. (1994): Foundations of differential geodesy, Springer-Verlag, Berlin.

A solution of Stokes' problem for the ellipsoidal Earth by means of Green's function

M.I.Yurkina

A solution of Stokes' boundary problem for the ellipsoidal (spheroidal) Earth accounting for terms of the order of the Earth's flattening can be obtained in spheroidal coordinates u, v, w, which are connected with cartesian ones by formulas

$$x=c\ \mathrm{sin}u\ \mathrm{cos}v\ \mathrm{ch}w,\ y=c\ \mathrm{sin}u\ \mathrm{sin}v\ \mathrm{ch}w,\ z=c\ \mathrm{cos}u\ \mathrm{sh}w,$$

Lamé's coefficients being

$$h_u=h_w=c(\mathrm{ch}^2w-\mathrm{sin}^2u)^{1/2},\ h_v=c\ \mathrm{sin}u\ \mathrm{ch}w.$$

On the Earth's spheroid we have

$$h_u=h_v=a\ (1-\frac{c^2}{a^2}\sin^2 u)^{1/2},\ \ h_v=a\ \mathrm{sin}u,$$

where $c=(a^2-b^2)^{1/2}$ and a, b are the semi-axes of the spheroid. Green's function of the Stokes' problem for the ellipsoidal Earth can be established by means of Green's formula

$$4\pi T=\int_S (T\frac{\partial}{\partial\nu}\cdot\frac{1}{r_{Tds}}-\frac{1}{r_{Tds}}\cdot\frac{\partial T}{\partial\nu})dS, \qquad (1)$$

T being the disturbing potential, ν the external normal to the ellipsoidal surface S, r_{Tds} being the distance between the fixed point out of the spheroid and the element dS. The Green's function can be found as a sum $\frac{1}{r_{Tds}}+\varphi$, the function φ being a harmonic function out of the spheroid. The derivative $\frac{\partial T}{\partial\nu}$ can be excluded from the formula (1) by means of the boundary condition

$$\frac{\partial T}{\partial\nu}-\frac{T}{\gamma}\cdot\frac{\partial\gamma}{\partial\nu}=-(g-\gamma),$$

g being the gravity, γ its normal value. Then

$$T=\frac{1}{4\pi}\int_S \left[T\frac{\partial}{\partial\nu}(\frac{1}{r_{Tds}}+\varphi)-\frac{1}{\gamma}\cdot\frac{\partial\gamma}{\partial\nu}(\frac{1}{r_{Tds}}+\varphi)\ \right]+(\frac{1}{r_{Tds}}+\varphi)(g-\gamma)\Big]dS,$$

the unknown part φ of the Green's function should satisfy the boundary condition

$$\frac{\partial\varphi}{\partial\nu}-\frac{\varphi}{\gamma}\cdot\frac{\partial\gamma}{\partial\nu}+\frac{\partial}{\partial\nu}\ \frac{1}{r_{Tds}}-\frac{1}{\gamma}\ \frac{\partial\gamma}{\partial\nu}\cdot\frac{1}{r_{Tds}}=0. \qquad (2)$$

This approach is suggested by Ostach (1969). It can be written

$$4\pi\varphi = \int_S \left[\varphi \frac{\partial}{\partial \nu}\frac{1}{r_{\varphi ds}} - \frac{1}{r_{\varphi ds}}\cdot\frac{\partial \varphi}{\partial \nu}\right] dS \qquad (3)$$

The derivative $\frac{\partial \varphi}{\partial \nu}$ can be excluded from (3) by means of the boundary condition (2) and an integral equation for function φ will be obtained

$$4\pi\varphi - \int_S \varphi\left[\frac{\partial}{\partial \nu}\frac{1}{r_{\varphi ds}} - \frac{1}{\gamma}\cdot\frac{\partial \gamma}{\partial \nu}\cdot\frac{1}{r_{\varphi ds}}\right] dS = \int_S \frac{1}{r_{\varphi ds}}\left[\frac{\partial}{\partial \nu}\frac{1}{r_{Tds}} - \frac{1}{r_{Tds}}\cdot\frac{1}{\gamma}\cdot\frac{\partial \gamma}{\partial \nu}\right] dS$$

or

$$4\pi\varphi - \int_S \varphi\left[\frac{\partial}{\partial w}\frac{1}{r_{\varphi ds}} - \frac{1}{\gamma}\cdot\frac{\partial \gamma}{\partial \nu}\cdot\frac{1}{r_{\varphi ds}}\right] h_\nu du d\nu = \int_S \frac{1}{r_{\varphi ds}}\left[\frac{\partial}{\partial w}\cdot\frac{1}{r_{Tds}} - \frac{1}{r_{Tds}}\cdot\frac{1}{\gamma}\cdot\frac{\partial \gamma}{\partial \nu}\right] h_\nu du d\nu . \qquad (4)$$

As it is known

$$\frac{1}{r_{\varphi ds}} = \frac{i}{c}\sum_{n=0}^{\infty} (2n+1)\left[P_n(\cos u) P_n(\cos u_\varphi) Q_n(ishw_\varphi) P_n(ishw) + \right.$$

$$\left. +2\sum_{m=1}^{\infty} (-1)^m \left(\frac{(n-m)!}{(n+m)!}\right)^2 P_{nm}(\cos u) P_{nm}(\cos u_\varphi) Q_{nm}(ishw_\varphi) P_{nm}(ishw)\cos m(\nu - \nu_\varphi)\right].$$

We have for the current point of the surface S

$$\frac{\partial}{\partial w}\frac{1}{r_{\varphi ds}} = \frac{i}{c}\sum_{n=0}^{\infty} (2n+1)\left[P_n(\cos u)\, P_n(\cos u_\varphi) Q_n(ishw_\varphi)\frac{dP_n(ishw)}{dw} + \right.$$

$$\left. +2\sum_{m=1}^{n} (-1)^m \left(\frac{(n-m)}{(n+m)}\right)^2 P_{nm}(\cos u) P_{nm}(\cos u_\varphi) Q_{nm}(ishw_\varphi)\frac{dPnm(ishw)}{dw}\cos m(\nu - \nu_\varphi)\right],$$

where

$$\left(\frac{dP_{nm}(ishw)}{dw}\right)_{shw=\frac{b}{c}} = -\frac{ci}{a}\left[(n-m+1)P_{n+1,m}(i\tfrac{b}{c}) - (n+1)i\tfrac{b}{c}P_{nm}(i\tfrac{b}{c})\right].$$

It can be obtained by means of known formulas (Gradshtein, Ryzhik 1965)

$$Q_{nm}(i\tfrac{b}{c})\left(\frac{dP_{nm}(ishw)}{dw}\right)_{shw=\frac{b}{c}} \approx -(-1)^m \frac{ci}{a}\left(\frac{a}{b}\right)^{2m}\frac{(n+m)!}{(n-m)!}\,\frac{n}{2n+1}\times$$

$$\times\left[1 + \frac{c^2}{b^2}\left(\frac{(n-m)(n^2-mn+n+2m)}{2n(2n-1)} - \frac{(n+m+1)(n+m+2)}{2(2n+3)}\right)\right]$$

and

$$Q_{nm}(i\tfrac{b}{c})P_{nm}(i\tfrac{b}{c}) \approx (-1)^m \frac{c}{ib}\left(\frac{a}{b}\right)^{2m}\frac{1}{2n+1}\frac{(n+m)!}{(n-m)!}\times$$

$$\times\left[1 + \frac{c^2}{b^2}\left(\frac{(n-m-1)(n-m)}{2(2n-1)} - \frac{(n+m+2)(n+m+1)}{2(2n+3)}\right)\right].$$

It can be written with the assumed accuracy

$$\gamma = \frac{GM}{b^2}\left(1 - \frac{1}{2}\frac{c^2}{b^2} - \frac{3}{2}p - \frac{1}{2}\frac{c^2}{b^2}\cos^2 u + \frac{5}{2}p\cos^2 u\right),$$

$$\frac{\partial \gamma}{\partial w} = -\frac{2GM}{b^2}\left(1 - \frac{1}{2}\frac{c^2}{b^2} - \frac{c^2}{b^2}\cos^2 u - \frac{1}{2}p + \frac{5}{2}p\cos^2 u\right),$$

$$\frac{1}{\gamma}\cdot\frac{\partial \gamma}{\partial w} = -2\left(1 - \frac{1}{2}\frac{c^2}{b^2}\cos^2 u + p\right),$$

where $p = \frac{\omega^2 b^3}{GM}$. The function φ for the spheroidal Earth can be represented by an expansion

$$\varphi = \frac{1}{c}\sum_{n=0}^{\infty}(2n+1)\Big[A_n P_n(\cos u_T)P_n(\cos u)Q_n(ishw_T)P_n(ishw) +$$
$$+2\sum_{m=1}^{n}(-1)^m\left(\frac{(n-m)!}{(n+m)!}\right)^2\left(\frac{b}{a}\right)^{4m}A_{nm}P_{nm}(\cos u_T)P_{nm}(\cos u)\times$$
$$\times Q_{nm}(ishw_T)P_{nm}(ishw)\cos m(\nu_T-\nu)\Big]$$

or

$$\varphi = \frac{1}{b}\sum_{n=0}^{\infty}\Big[A_n P_n(\cos u_T)P_n(\cos u)\Big[1+\frac{c^2}{b^2}\left(\frac{n(n-1)}{2(2n-1)}-\frac{(n+1)(n+2)}{2(2n+3)}\right)\Big]+$$
$$+2\sum_{m=1}^{n}A_{nm}\frac{(n-m)!}{(n+m)!}\left(\frac{b}{a}\right)^{2m}P_{nm}(\cos u_T)P_{nm}(\cos u)\times \qquad (5)$$
$$\times\Big[1+\frac{c^2}{b^2}\left(\frac{(n-m-1)(n-m)}{2(2n-1)}-\frac{(n+m+1)(n+m+2)}{2(2n+3)}\right)\Big]\cos m(\nu_T-\nu)\Big].$$

It will be possible to express coefficients A_{nm} after substitution of formula (5) into the equation (4) and collecting terms with $P_{nm}(\cos u_t)$ $P_{nm}(\cos u_\varphi)$. Since

$$\int_0^{2\pi}\cos m(\nu-\nu_T)\cos m(\nu-\nu_\varphi)d\nu = \pi\cos m(\nu_T-\nu_\varphi),$$
$$\int_s \varphi\frac{\partial}{\partial w}\frac{1}{r_{\varphi ds}}h_\nu du d\nu = \frac{4\pi}{b}\sum_{n=0}^{\infty}\Big\{\frac{A_n n}{2n+1}P_n(\cos u_T)P_n(\cos u_\varphi)\Big[1+\frac{c^2}{b^2}\left(\frac{n^2}{2n-1}-\frac{(n+1)(n+2)}{2n+3}\right)\Big]+$$
$$+2\sum_{m=1}^{n}\frac{A_{mn}n}{2n+1}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)\frac{(n-m)!}{(n+m)!}\Big[1+\frac{c^2}{b^2}\left(\frac{(n-m)(n^2-mn+m)}{n(2n-1)}-\frac{(n+m+1)(n+m+2)}{2n+3}\right)\Big]\times$$
$$\times\cos m(\nu_T-\nu_\varphi)\Big\}.$$

The next integral can be expressed by means of the known relation

$$\cos^2 u\,P_{nm}(\cos u) = \frac{(n+m)(n+m-1)}{(2n-1)(2n+1)}P_{n-2,m}(\cos u)+\frac{2n^2-2m^2+2n-1}{(2n-1)(2n+3)}P_{nm}(\cos u)+$$
$$+\frac{(n-m+1)(n-m+2)}{(2n+1)(2n+3)}P_{n+2,m}(\cos u).$$

Then

$$\int_s \frac{\varphi}{r_{\varphi ds}}\cdot\frac{1}{\gamma}\cdot\frac{\partial\gamma}{\partial w}h_\nu du d\nu = -2(1+p)\int_s\frac{\varphi}{r_{\varphi ds}}h_\nu du d\nu+\frac{c^2}{b^2}\int_s\frac{\varphi}{r_{\varphi ds}}h_\nu\cos^2 u\,du d\nu =$$
$$=-\frac{8\pi a}{b^2}(1+p)\sum_{n=0}^{\infty}\Big\{\frac{A_n}{2n+1}P_n(\cos u_T)P_n(\cos u_\varphi)\Big[1+\frac{c^2}{b^2}\left(\frac{n(n-1)}{2n-1}-\frac{(n+1)(n+2)}{(2n+3)}\right)\Big]+$$
$$+2\sum_{m=1}^{n}\frac{A_{nm}}{2n+1}\cdot\frac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)\times$$
$$\times\Big[1+\frac{c^2}{b^2}\left(\frac{(n-m-1)(n-m)}{2n-1}-\frac{(n+m+1)(n+m+2)}{2n+3}\right)\Big]\cos m(\nu_T-\nu_\varphi)\Big\}+$$

$$+\frac{4\pi}{b}\cdot\frac{c^2}{b^2}\sum_{n=0}^{\infty}\left\{\frac{A_n}{2n+1}P_n(\cos u_T)P_n(\cos u_\varphi)E_n+\right.$$

$$\left.+2\sum_{m=1}^{n}\frac{Anm}{2n+1}\cdot\frac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)E_{nm}\cos m(\nu_T-\nu_\varphi)\right\},$$

where

$$E_{nm}=\frac{(n+m+2)(n+m+1)}{(2n+3)(2n+5)}+\frac{2n^2-2m^2+2n-1}{(2n-1)(2n+3)}+\frac{(n-m-1)(n-m)}{(2n-3)(2n-1)},\ E_n=(E_{nm})_{m=0}$$

We get also

$$a\int_s\frac{1}{r_{\varphi ds}}\cdot\frac{\partial}{\partial w}\frac{1}{r_{Tds}}\sin u\,du\,dv=\frac{4\pi}{b}\sum_{n=0}^{\infty}\frac{n}{2n+1}\left\{P_n(\cos u_T)P_n(\cos u_\varphi)\times\right.$$

$$\times\left[1+\frac{c^2}{b^2}\left(\frac{n^2}{2n-1}-\frac{(n+1)(n+2)}{2n+3}\right)\right]+2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)\left(\frac{a}{b}\right)^{4m}\times$$

$$\left.\times\left[1+\frac{c^2}{b^2}\left(\frac{(n-m)(n^2-nm+m)}{n(2n-1)}-\frac{(n+m+1)(n+m+2)}{2n+3}\right)\right]\cos m(\nu_T-\nu_\varphi)\right\},$$

$$a\int_s\frac{1}{r_{\varphi ds}}\cdot\frac{1}{r_{Tds}}\cdot\frac{1}{\gamma}\cdot\frac{\partial\gamma}{\partial\nu}\sin u\,du\,dv=-2a(1+p)\int_s\frac{1}{r_{\varphi ds}}\cdot\frac{1}{r_{Tds}}\sin u\,du\,dv+$$

$$+a\frac{c^2}{b^2}\int_s\frac{1}{r_{\varphi ds}}\cdot\frac{1}{r_{Tds}}\cos^2 u\sin u\,du\,dv=$$

$$=-\frac{8\pi a}{b^2}(1+p)\sum_{n=0}^{\infty}\frac{1}{2n+1}\left\{P_n(\cos u_T)P_n(\cos u_\varphi)\left[1+\frac{c^2}{b^2}\left(\frac{n(n-1)}{2n-1}-\frac{(n+1)(n+2)}{2n+3}\right)\right]+\right.$$

$$+2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)\left(\frac{a}{b}\right)^{4m}\times$$

$$\left.\times\left[1+\frac{c^2}{b^2}\left(\frac{(n-m-1)(n-m)}{2n-1}-\frac{(n+m+1)(n+m+2)}{2n+3}\right)\right]\cos m(\nu_T-\nu_\varphi)\right\}+$$

$$+\frac{4\pi}{b}\cdot\frac{c^2}{b^2}\sum_{n=0}^{\infty}\frac{1}{2n+1}\left\{E_nP_n(\cos u_T)P_n(\cos u_\varphi)+\right.$$

$$\left.+2\sum_{m=1}^{n}E_{nm}\frac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)\cos m(\nu_T-\nu_\varphi)\right\}.$$

We get

$$K_{nm}A_{nm}=L_{nm}$$

where

$$K_{nm} = (n-1) - 2(\tfrac{a}{b}(1+p)-1) + (2n+1)\ (\ (\tfrac{b}{a})^{2m} - 1) +$$
$$+\tfrac{c^2}{b^2}\Big\{(2n-3)(\tfrac{(n-m-1)(n-m)}{2(2n-1)} - \tfrac{(n+m+1)(n+m+2)}{(2n+3)}) +$$
$$+(\tfrac{(n+m+1)(n+m+2)}{2n+3} - \tfrac{(n-m)(n^2-mn+m)}{n(2n-1)}) + E_{nm}\ \Big\},$$
$$L_{nm} = n+2-2(1-(1+p)(\tfrac{a}{b})^{4m+1} + n((\tfrac{a}{b})^{4m} - 1) +$$
$$+\tfrac{c^2}{b^2}\Big\{n\ (\tfrac{(n-m)(n^2-mn+m)}{n(2n-1)} - \tfrac{(n+m+1)(n+m+2)}{2n+3}) + 2\ (\tfrac{(n-m-1)(n-m)}{2n-1} - \tfrac{(n+m+1)(n+m+2)}{2n+3}) - E_{nm}\ \Big\}.$$

The coefficient A_n cannot be determined by n=1. It can be expressed

$$A_{nm} = \tfrac{L_{nm}}{K_{nm}} = \tfrac{n+2}{n-1}\Big\{1 + 2p\tfrac{2n+1}{(n-1)(n+2)} + \tfrac{c^2}{b^2}\Big\{\tfrac{4mn^2+7mn+2n-2m+1}{(n-1)(n+2)} +$$
$$+\tfrac{(n-m)(n^2-mn+m)(2n+1)}{(n-1)(n+2)(2n-1)} - \tfrac{(n-m-1)(n-m)(2n^2-3n-2)}{2(n-1)(n+2)(2n-1)} +$$
$$-\tfrac{(n+m+1)(n+m+2)(2n+1)}{2(n-1)(2n+3)} - \tfrac{(2n+1)}{(n-1)(n+2)}E_{nm}\ \Big\}.$$

By excluding differences of great numbers we obtain

$$\varphi = \tfrac{1}{b}\sum_{\substack{n=0\\ n\neq 1}}^{\infty} \tfrac{n+2}{n-1}\Big\{1 + \tfrac{2n+1}{(n-1)(n+2)}(2p - \tfrac{c^2}{b^2}E_n) +$$
$$-\tfrac{1}{2}\tfrac{c^2}{b^2}\Big\{1 - \tfrac{4n+2}{(n-1)(n+2)} - \tfrac{3n^2+43n-22}{4(2n^3+n^2-5n+2)} - \tfrac{27n+45}{4(2n^2+n-3)}\ \Big\}P_n(\cos u_T)P_n(\cos u_\varphi)\ \Big\} +$$
$$+\tfrac{2}{b}\sum_{\substack{n=0\\ n\neq 1}}^{\infty} \tfrac{(n+2)}{(n-1)}\sum_{m=1}^{n}\ \Big\{1 + \tfrac{2n+1}{(n-1)(n+2)}(2p - \tfrac{c^2}{b^2}E_{nm}) +$$
$$-\tfrac{1}{2}\tfrac{c^2}{b^2}\Big\{1 - \tfrac{4n^2+6mn+2n+12m}{n(n^2+n-2)} - \tfrac{2m^2n^2+3n^3-24mn^2+22m^2n-36mn+43n^2-12m^2-22n+24m}{4n(2n^3+n^2-5n+2)} +$$
$$+\tfrac{6m^2n-27n^2-24mn-18m^2-45n-36m}{4n(2n^2+n-3)}\ \Big\}\tfrac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)\cos m(v_T - v_\varphi)\ \Big\}.$$

Since

$$\tfrac{1}{r} = \tfrac{1}{b}\sum_{n=0}^{\infty}\ \Big\{\ P_n(\cos\psi) - \tfrac{1}{2}\tfrac{c^2}{b^2}(1 + \tfrac{1}{4(2n-1)} - \tfrac{1}{4(2n+3)})P_n(\cos u_T)P_n(\cos u_\varphi) +$$
$$-\tfrac{c^2}{b^2}\sum_{m=1}^{n}\ \tfrac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)(1 - \tfrac{2m^2-n}{4n(2n-1)} - \tfrac{6m^2+n}{4n(2n+3)})\cos m(v_T - v_\varphi)\ \Big\}.$$

Green's function S_{sph} for Stokes' problem of the spheroidal Earth can be written

$$S_{sph} = S(\cos\psi) - 1 + \tfrac{1}{b}\sum_{\substack{n=0\\ n\neq 1}}^{\infty}\ \Big[\ \tfrac{n+2}{n-1}\Big\{\tfrac{2n+1}{(n-1)(n+2)}(2p - \tfrac{c^2}{b^2}E_n) - \tfrac{1}{2}\tfrac{c^2}{b^2}\Big\{\ 1 - \tfrac{4n+2}{(n-1)(n+2)} +$$
$$-\tfrac{3n^2+43n-22}{4(2n^3+n^2-5n+2)} - \tfrac{27n+45}{4(2n^2+n-3)}\ \Big\}\Big\} - \tfrac{1}{2}\tfrac{c^2}{b^2}(1 + \tfrac{1}{4(2n-1)} - \tfrac{1}{4(2n+3)})\ \Big]P_n(\cos u_T)P_n(\cos u_\varphi) +$$

$$+\frac{2}{b}\sum_{\substack{n=0\\n\neq 1}}^{\infty}\sum_{m=1}^{n}\left\{\frac{n+2}{n-1}\left[\frac{2n+1}{(n-1)(n+2)}\left(2p-\frac{c^2}{b^2}E_{nm}\right)-\frac{1}{2}\frac{c^2}{b^2}\left[1-\frac{4n^2+6mn+2n+12m}{n(n^2+n-2)}+\right.\right.\right.$$

$$\left.-\frac{2m^2n^2+3n^3-24mn^2+22m^2n-36mn+43n^2-12m^2-22n+24m}{4n(2n^3+n^2-5n+2)}+\right.$$

$$\left.\left.+\frac{6m^2n-27n^2-24mn-18m^2-45n-36m}{4n(2n^2+n-3)}\right]\right]+$$

$$\left.-\frac{1}{2}\frac{c^2}{b^2}\left[1+\frac{n-2m^2}{4n(2n-1)}-\frac{n+6m^2}{4n(2n+3)}\right]\right\}\frac{(n-m)!}{(n+m)!}P_{nm}(\cos u_T)P_{nm}(\cos u_\varphi)\cos m(\nu_T-\nu_\varphi),$$

where $$S(\cos\psi)=\frac{1}{b}\sum_{n=2}^{\infty}\frac{2n+1}{n-1}P_n(\cos\psi).$$

If the gravity anomalies are expressed by

$$g-\gamma=\sum_{n=0}^{\infty}A_nP_n(\cos u)+\sum_{n=2}^{\infty}\sum_{m=1}^{\infty}\begin{matrix}A_{nm}\cos m\nu\\B_{nm}\sin m\nu\end{matrix}P_{nm}(\cos u),$$

their influence upon the disturbing potential is

$$\delta T=b\sum_{\substack{n=0\\n\neq 1}}^{\infty}\sum_{m=1}^{n}\begin{matrix}A_{nm}\cos m\nu\\B_{nm}\sin m\nu\end{matrix}\left[\frac{n+2}{n-1}\left\{\frac{1}{(n-1)(n+2)}\left(2p-\frac{c^2}{b^2}E_{nm}\right)-\frac{1}{2(2n+1)}\frac{c^2}{b^2}\left\{1-\frac{4n^2+6mn+2n+12m}{n(n^2+n-2)}+\right.\right.\right.$$

$$\left.\left.-\frac{2m^2n^2+3n^3-24mn^2+22m^2n-36mn+43n^2-12m^2-22n+24m}{4n(2n^3+n^2-5n+2)}+\frac{6m^2n-27n^2-24mn-18m^2-45n-36m}{4n(2n^2+n-3)}\right\}\right\}+$$

$$\left.-\frac{1}{2(2n+1)}\cdot\frac{c^2}{b^2}\left(1+\frac{n-2m^2}{4n(2n-1)}-\frac{n+6m^2}{4n(2n+3)}\right)\right]P_n(\cos u_T).$$

The terms of this expression decrease as 1: n. Some useful indications can be found at Thông and Grafarend (1989).

Acknowledgement

This work was financially supported by the Russian Fund of Fundamental Research, Project 97-05-65054.

References

Gradshtein I.S., Ryzhik I.M. (1965): Table of integrals, series, and products. 4th ed., Translated from the Russian. New York Academic Press, XIV+1086 pp.

Ostach O.M. (1969): Determination of the disturbing potential and Stokes ' constants of the Earth by means of Green ' s function. Trudy CNIIGAiK N 176, 26-42. In the Russian.

Thông Nguyen Chi, Grafarend E.W. (1989): A spheroidal harmonic model of the terrestrial gravitational field. Manuscripta geodaetica 14, 285-304.

List of Authors

Ádám József: Department of Geodesy and Surveying, Technical University of Budapest, P.O.Box 91, H-1521 Budapest, Hungary, email: jadam@epito.bme.hu

Aduol Francis W O: Department of Surveying, University of Nairobi, P.O.Box 30197, Nairobi, Kenya

Ardalan Alireza A: Institute of Geodesy, Stuttgart University, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart, Germany, email: ardalan@gis.uni-stuttgart.de

Awange Joseph L: Institute of Geodesy, Stuttgart University, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart, Germany, email: awange@gis.uni-stuttgart.de

Bähr Hans-Peter: Institut für Photogrammetrie und Fernerkundung, Universität Karlsruhe, Englerstr. 7, D-76131 Karlsruhe, Germany, email: baehr@ipf.bau-verm.uni-karlsruhe.de

Blais J A Rod: University of Calgary, Department of Geomatics Engineering, 2500 University Drive, N.W., Calgary, Alberta, Canada T2N 1N4, email: blais@ucalgary.ca

Caputo Michele: Department of Physics, University "La Sapienza", Piazzale A. Moro, 2, I-00185 Roma, Italy, email: mcaputo@axcasp.caspur.it

Chai Goh Pong: School of Civil & Structural Engineering, Nanyang Technological University, Singapore 639789, email: cpcgoh@ntu.edu.sg

Chen Horng-Yue: School of Geomatic Engineering, The University of New South Wales, Sydney NSW 2052, Australia, email: Z2196150@student.unsw.edu.au

Crosilla Fabio: Dipartimento di Georisorse e Territorio, Università di Udine, Via Cotonificio, 114, I-33100 Udine, Italy, email: crosilla@dgt.uniud.it

Cui Chunfang: Institut für Geodäsie und Geoinformatik, Sekr. H 12, TU Berlin, Straße des 17. Juni 135, D-10623 Berlin, Germany, email: cui@lego.bv.tu-berlin.de

Dermanis Athanasios: Department of Geodesy and Surveying, The Aristotle University of Thessaloniki, University Box 503, 54006 Thessaloniki, Greece, email: dermanis@topo.auth.gr

Dorrer Egon: Institut für Photogrammetrie und Kartographie, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, D-85579 Neubiberg, Germany, email: egon.dorrer@unibw-muenchen.de

Featherstone Will E: School of Spatial Sciences, Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia, email: tfeather@cc.curtin.edu.au

Fei Zhiling L: University of Calgary, Department of Geomatics Engineering, 2500 University Drive, N.W., Calgary, Alberta, Canada T2N 1N4, email: zlfei@ucalgary.ca

Förstner Wolfgang: Institut für Photogrammetrie, Universität Bonn, Nussallee 15, D-53115 Bonn, Germany, email: wf@ipb.uni-bonn.de

Fritsch Dieter: Institut für Photogrammetrie, Universität Stuttgart, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart, Germany, email: dieter.fritsch@ifp.uni-stuttgart.de

Groten Erwin: Institute for Physical Geodesy, Darmstadt University of Technology, Petersenstrasse 13, D-64287 Darmstadt, Germany, email: groten@ipgs.ipg.verm.tu-darmstadt.de

Han Shaowei: School of Geomatic Engineering, The University of New South Wales, Sydney NSW 2052, Australia, email: s.han@unsw.edu.au

Hannah Bruce M: Cooperative Research Centre for Satellite Systems, Queensland University of Technology, 2 George Street, GPO BOS 2434, Brisbane Q 4001, Australia, email: b.hannah@ieee.org

Hartung Joachim: Department of Statistics, SFB 475, University of Dortmund, Vogelpothsweg 87, D-44221 Dortmund, Germany, email: hartung@omega.statistik.uni-dortmund.de

Heck Bernhard: Geodetic Institute, University of Karlsruhe, Englerstr. 7, D-76128 Karlsruhe, Germany, email: heck@gik.uni-karlsruhe.de

Hehl Friedrich W: Institute for Theoretical Physics, University of Cologne, Zülpicher Str. 77, D-50923 Köln, Germany, email: hehl@thp.Uni-Koeln.DE

Hils Alfred: Vermessungsbüro Hils, Heusteigstr. 28, D-70180 Stuttgart, Germany, email: vermessung@hils.ingenieure.de

Hofmann-Wellenhof Bernhard: Institute of Applied Geodesy, Technical University Graz, Dept. of Positioning & Navigation, Steyrergasse 30, A-8010 Graz, Austria, email: howe@tu-graz.ac.at

Ilk Karl Heinz: Institut für Theoretische Geodäsie, Universität Bonn, Nussallee 17, D-53115 Bonn, Germany, email: ilk@theor.geod.uni-bonn.de

Jurisch Ronald: Hochschule Anhalt (FH), Fachbereich Vermessungswesen, Bauhausstr. 6, D-06846 Dessau, Germany, email: jurisch@vw.hs-anhalt.de

Kakkuri Juhani: Finnish Geodetic Institute, Geodeetinrinne 2, FIN-02430 Masala, Finland, email: jka@beta.fgi.fi

Kampmann Georg: Hochschule Anhalt (FH), Fachbereich Vermessungswesen, Bauhausstr. 6, D-06846 Dessau, Germany, email: kampmann@vw.hs-anhalt.de

Keller Wolfgang: Institute of Geodesy, Stuttgart University, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart, Germany, email: keller@gis.uni-stuttgart.de

Kleusberg Alfred: Institute of Navigation, Stuttgart University, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart, Germany, email: kleusberg@nav.uni-stuttgart.de

Koch Karl-Rudolf: Institut für Theoretische Geodäsie, Universität Bonn, Nussallee 17, D-53115 Bonn, Germany, email: koch@theor.geod.uni-bonn.de

Kubik Kurt: Department of Geographical Sciences and Planning. The University of Queensland, Brisbane Qld 4072, Australia, email: k.kubik@mailbox.uq.edu.au

Leick Alfred: Department of Spatial Information, University of Maine, Orono, ME, USA, email: Leick@spatial.maine.edu

Lelgemann Dieter: Institut für Geodäsie und Geoinformatik, Sekr. H 12, TU Berlin, Straße des 17. Juni 135, D-10623 Berlin, Germany, email: lelge@mca.bv.tu-berlin.de

Linke Janette: Hochschule Anhalt (FH), Fachbereich Vermessungswesen, Bauhausstr. 6, D-06846 Dessau, Germany, email: linke@vw.hs-anhalt.de

Livieratos Evangelos: Department of Geodesy and Surveying, The Aristotle University of Thessaloniki, University Box 503, 54006 Thessaloniki, Greece, email: livier@topo.auth.gr

Linkwitz Klaus: Obere Tannenbergerstr. 24, D-71229 Leonberg, Germany

Marchenko Alexander N: State University "Lviv Polytechnic", P.Pancha St. 11/93, 290020 Lviv, Ukraine, email: march@polynet.lviv.ua

Martinĕc Zdenek: Department of Geophysics, Charles University, V Holešovičkách 2, 180 00 Prague 8, Czech Republic, email: zdenek@hervam.troja.mff.cuni.cz

Meier Siegfried: Institut für Planetare Geodäsie, TU Dresden, Mommsenstraße 13, D-01062 Dresden, Germany, email: meier@ipg.geo.tu-dresden.de

Mira Sjamsir: Pusat Penelitian Kelantan, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung 40132, Indonesia, email: sjamsir@rad.net.id

Moonen Boudewijn: Institut für Photogrammetrie, Universität Bonn, Nussallee 15, D-53115 Bonn, Germany, email: bm@ipb.uni-bonn.de

Moritz Helmut: Technische Universität Graz, Abteilung für Physikalische Geodäsie, Steyrergasse 30, A-8010 Graz, Austria, email: moritz@phgg.tu-graz.ac.at

Obukhov Yuri N: Institute for Theoretical Physics, University of Cologne, Zülpicher Str. 77, D-50923 Köln, Germany, (permanent address: Moscow State University), yo@thp.Uni-Koeln.DE,

Papo Haim B: Civil Engineering Technion, Israel Institute of Technology, Technion City, 32000 Haifa, Israel, email: haimp@techunix.technion.ac.il

Plastino Wolfango: Department of Physics, University "Roma Tre", Via della Vasca Navale, 84, I-00146 Roma, Italy, email: plastino@fis.uniroma3.it

Reilly Ian W: Department of Surveying, University of Otago, Dunedin, New Zealand; mailto: 1/264 Riccarton Road, Christchurch 8004, New Zealand, email: ireilly@southern.co.nz

Rizos Chris: School of Geomatic Engineering, The University of New South Wales, Sydney NSW 2052, Australia, email: c.rizos@unsw.edu.au

Rozanov Youri: IAMI-CNR, via Ampere, 56, 20100 Milano, Italy

Rowe Clive H: Department of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch, New Zealand, email: C.Rowe@phys.canterbury.ac.nz

Rubilar Guillermo F: Institute for Theoretical Physics, University of Cologne, Zülpicher Str. 77, D-50923 Köln, Germany, email: gr@thp.Uni-Koeln.DE

Rummel Reiner: Institut für Astronomische und Physikalische Geodäsie, TU München, Arcisstr. 21, D-80290 München, Germany, email: rummel@step.iapg.verm.tu-muenchen.de

Sansò Fernando: Istituto di Ingegneria Idraulica, Ambientale e del Rilevamento, Sezione Rilevamento, Piazza L. da Vinci, 32, I-20133 Milano, Italy, email: fsanso@ipmtf4.topo.polimi.it

Schäfer Volker: Wirtschaftsministerium Baden-Württemberg, Postfach 103451, 70029 Stuttgart

Schaffrin Burkhard: Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Ave., Columbus, OH 43210-1275, USA, email: schaffrin1@osu.edu
Schlüter Wolfgang: Bundesamt für Kartographie und Geodäsie, Fundamentalstation Wettzell, D-93444 Kötzting, Germany, email: hase@wettzell.ifag.de
Schmitz-Hübsch Harald: Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, D-80539 München, Germany, email: schmihue@dgfi.badw.de
Schneider Manfred: Forschungseinrichtung Satellitengeodäsie, Fundamentalstation Wettzell, D-93444 Kötzting, Germany, email: mxsx@alpha.fesg.tu-muenchen.de
Schreiber Ulrich: Forschungseinrichtung Satellitengeodäsie, Fundamentalstation Wettzell, D-93444 Kötzting, Germany, email: schreiber@wettzell.ifag.de
Schuh Harald: Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, D-80539 München, Germany, email: schuh@dgfi.badw-muenchen.de
Seitz Kurt: Geodetic Institute, University of Karlsruhe, Englerstr. 7, D-76128 Karlsruhe, Germany, email: seitz@gik.uni-karlsruhe.de
Sideris Michael G: University of Calgary, Department. of Geomatics Engineering, 2500 University Drive, N.W., Calgary, Alberta, Canada T2N 1N4, email: sideris@ucalgary.ca
Sjöberg Lars E: Royal Institute of Technology, Department of Geodesy and Photogrammetry, S-10044 Stockholm, Sweden, email: sjoberg@geomatics.kth.se
Stedman Geoffrey E: Department of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch, New Zealand, email: G.Stedman@phys.canterbury.ac.nz
Svensson S Leif: Centre for Mathematical Sciences, Lund University, Box 118, S-221 00 Lund, Sweden, email: Leif.Svensson@math.lth.se
Teunissen Peter J G: Department of Mathematical Geodesy and Positioning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Thijsseweg 11, NL-2629 JA Delft, The Netherlands, email: p.j.g.teunissen@geo.tudelft.nl
van Gelderen Martin: Delft Institute for Earth-Oriented Space Research, Technical University Delft, Thijsseweg 11, NL-2629 JA Delft, The Netherlands, email: M.vangelderen@geo.tudelft.nl
Varga Péter: Geodetic and Geophysical Research Institute, Hungarian Academy of Sciences, H-9400 Sopron, Csatkai E.u. 6-8, Hungary, email: varga@ggki.hu
Vermeer Martin: Finnish Geodetic Institute, Geodeetinrinne 2, FIN-02430 Masala, Finland, email: mv@fgi.fi
Walker Rodney A: Cooperative Research Centre for Satellite Systems, Queensland University of Technology, 2 George Street, GPO BOS 2434, Brisbane Q 4001, Australia, email: ra.walker@qut.edu.au
Welsch Walter: Institut für Geodäsie, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, D-85579 Neubiberg, Germany, email: walter.welsch@unibw-muenchen.de
Wittenburg Rüdiger: Institut für Markscheidewesen, Bergakademie Freiberg, Agricolastr. 1, D-09599 Freiberg/Sachsen, Germany, email: wittenbu@mabb.tu-freiberg.de
Xu Peiliang: Research Centre for Earthquake Prediction, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan, email: pxu@rcep.dpri.kyoto-u.ac.jp
You Rey-Jer: Department of Surveying Engineering, National Cheng Kung University, 1, University Road, Tainan, Taiwan, Republic of China, email: rjyou@mail.ncku.edu.tw
Yurkina M I: TsNIIGAIK, Onejskaya 26, 125413 Moscow, Russia, email: gleb@space.ru

Schriftenreihe Institut für Photogrammetrie der Universität Stuttgart, Stuttgart University, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart
Fax: ++49 711 121 3297; Internet: http://www.ifp.uni-stuttgart.de

---

Nr.	1	(1976)	Vorträge des Lehrgangs Numerische Photogrammetrie (III), Esslingen 1975 - vergriffen
Nr.	2	(1976)	Vorträge der 35. Photogrammetrischen Woche Stuttgart 1975
Nr.	3	(1976)	Contributions of the XIIIth ISP-Congress of the Photogrammetric Institute, Helsinki 1976 - vergriffen
Nr.	4	(1977)	Vorträge der 36. Photogrammetrischen Woche Stuttgart 1977
Nr.	5	(1979)	E. Seeger: Das Orthophotoverfahren in der Architekturphotogrammetrie, Dissertation
Nr.	6	(1980)	Vorträge der 37. Photogrammetrischen Woche Stuttgart 1979
Nr.	7	(1981)	Vorträge des Lehrgangs Numerische Photogrammetrie (IV): Grobe Datenfehler und die Zuverlässigkeit der photogrammetrischen Punktbestimmung, Stuttgart 1980 - vergriffen
Nr.	8	(1982)	Vorträge der 38. Photogrammetrischen Woche Stuttgart 1981
Nr.	9	(1984)	Vorträge der 39. Photogrammetrischen Woche Stuttgart 1983
Nr.	10	(1984)	Contributions to the XVth ISPRS-Congress of the Photogrammetric Institute, Rio de Janeiro 1984
Nr.	11	(1986)	Vorträge der 40. Photogrammetrischen Woche Stuttgart 1985
Nr.	12	(1987)	Vorträge der 41. Photogrammetrischen Woche Stuttgart 1987
Nr.	13	(1989)	Vorträge der 42. Photogrammetrischen Woche Stuttgart 1989
Nr.	14	(1989)	Festschrift - Friedrich Ackermann zum 60. Geburtstag, Stuttgart 1989
Nr.	15	(1991)	Vorträge der 43. Photogrammetrischen Woche Stuttgart 1991
Nr.	16	(1992)	Vorträge zum Workshop "Geoinformationssysteme in der Ausbildung", Stuttgart 1992

Technical Reports Department of Geodetic Science, Stuttgart University, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart
Fax: ++49 711 121 3285; Internet: http://www.uni-stuttgart.de/gi

---

Nr.	1	(1987)	K. Eren: Geodetic Network Adjustment Using GPS Triple Difference Observations and a Priori Stochastic Information
Nr.	2	(1987)	F.W.O. Aduol: Detection of Outliers in Geodetic Networks Using Principal Component Analysis and Bias Parameter Estimation
Nr.	3	(1987)	M. Lindlohr: SIMALS; SIMulation, Analysis and Synthesis of General Vector Fields
Nr.	4	(1988)	W. Pachelski, D. Lapucha, K. Budde: GPS-Network Analysis: The Influence of Stochastic Prior Information of Orbital Elements on Ground Station Position Measures
Nr.	5	(1988)	W. Lindlohr: PUMA; Processing of Undifferenced GPS Carrier Beat Phase Measurements and Adjustment Computations
Nr.	6	(1988)	R.A. Snay, A.R. Drew: Supplementing Geodetic Data with Prior Information for Crustal Deformation in the Imperial Valley, California 1988
Nr.	7	(1989)	H.-W. Mikolaiski, P. Braun: Dokumentation der Programme zur Behandlung beliebig langer ganzer Zahlen und Brüche
Nr.	8	(1989)	H.-W. Mikolaiski: Wigner 3j Symbole, berechnet mittels Ganzzahlarithmetik
Nr.	9	(1989)	H.-W. Mikolaiski: Dokumentation der Programme zur Multikplikation nach Kugelfunktionen entwickelter Felder
Nr.	10	(1989)	H.-W. Mikolaiski, P. Braun: Dokumentation der Programme zur Differentiation und zur Lösung des Dirichlet-Problems nach Kugelfunktionen entwickelter Felder
Nr.	11	(1990)	L. Kubácková, L. Kubácek: Elimination Transformation of an Observation Vector preserving Information on the First and Second Order Parameters
Nr.	12	(1990)	L. Kubácková: Locally best Estimators of the Second Order Parameters in Fundamental Replicated Structures with Nuisance Parameters
Nr.	13	(1991)	G. Joos, K. Jörg: Inversion of Two Bivariate Power Series Using Symbolic Formula Manipulation
Nr.	14	(1991)	B. Heck, K. Seitz: Nonlinear Effects in the Scalar Free Geodetic Boundary Value Problem
Nr.	15	(1991)	B. Schaffrin: Generating Robustified Kalman Filters for the Integration of GPS and INS
Nr.	16	(1992)	Z. Martinec: The Role of the Irregularities of the Earth's Topography on the Tidally Induced Elastic Stress Distribution within the Earth
Nr.	17	(1992)	B. Middel: Computation of the Gravitational Potential of Topographic-Isostatic Masses
Nr.	18	(1993)	M.I. Yurkina, M.D. Bondarewa: Einige Probleme der Erdrotationsermittlung
Nr.	19	(1993)	L. Kubácková: Multiepoch Linear Regression Models
Nr.	20	(1993)	O.S. Salychev: Wave and Scalar Estimation Approaches for GPS/INS Integration

Schriftenreihe der Institute des Studiengangs Geodäsie und Geoinformatik (ehemals Fachbereich Vermessungswesen); Technical Reports Department of Geodesy and GeoInformatics, Stuttgart University, Geschw.-Scholl-Str. 24/D, D-70174 Stuttgart
Fax: ++49 711 121 3285; Internet: http://www.uni-stuttgart.de/gi

---

Nr. 1994.1 (1994)	H.-J. Euler: Generation of Suitable Coordinate Updates for an Inertial Navigation System
Nr. 1994.2 (1994)	W. Pachelski: Possible Uses of Natural (Barycentric) Coordinates for Positioning
Nr. 1995.1 (1995)	J. Engels, E.W. Grafarend, P. Sorcik: The Gravitational Field of Topographic-Isostatic Masses and the Hypothesis of Mass Condensation - Part I & II
Nr. 1995.2 (1995)	Minutes of the ISPRS Joint Workshop on Integrated Acquisition and Interpretation of Photogrammetric Data
Nr. 1996.1 (1996)	Festschrift für Klaus Linkwitz anläßlich der Abschiedsvorlesung im Wintersemester 1995/96; herausgegeben von Eberhard Baumann, Ulrich Hangleiter und Wolfgang Möhlenbrink
Nr. 1996.2 (1996)	J. Shan: Edge Detection Algorithms in Photogrammetry and Computer Vision
Nr. 1997.1 (1997)	Erste Geodätische Woche Stuttgart, 7.-12. Oktober 1996; herausgegeben von A. Gilbert und E.W. Grafarend
Nr. 1997.2 (1997)	U. Kälberer: Untersuchungen zur flugzeuggetragenen Radaraltimetrie
Nr. 1998.1 (1998)	L. Kubácek, L. Kubácková: Regression Models with a weak Nonlinearity
Nr. 1999.1 (1999)	GIS-Forschung im Studiengang Geodäsie und Geoinformatik der Universität Stuttgart; herausgegeben von M. Sester und F. Krumm
Nr. 1999.2 (1999)	Z. Martinec: Continuum Mechanics for Geophysicists and Geodesists. Part I: Basic Theory
Nr. 1999.3 (1999)	J. H. Dambeck: Diagnose und Therapie geodätischer Trägheitsnavigationssysteme. Modellierung – Systemtheorie – Simulation – Realdatenverarbeitung
Nr. 1999.4 (1999)	G. Fotopoulos, C. Kotsakis, M. G. Sideris: Evaluation of Geoid Models and Their Use in Combined GPS/Levelling/Geoid Height Network Adjustment
Nr. 1999.5 (1999)	Ch. Kotsakis, M. G. Sideris: The Long Road from Deterministic Collocation to Multiresolution Approximation
Nr. 1999.6 (1999)	Quo vadis geodesia...? Festschrift for Erik W. Grafarend on the occasion of his 60th birthday; herausgegeben von F. Krumm und V.S. Schwarze