THE CHALLENGE OF COMPUTING THE GEOID IN THE NINETIES

F. SANSÒ and G. SONA

D.I.I.A.R., Politecnico Di Milano, P. za Leonardo Da Vinci 32, I-20133 Milano, Italy

Abstract. This paper deals with the basic definitions and the numerical techniques used nowaday in the estimation of the gradiometric geoid. After a first introductory paragraph, the so-called 3-steps remove/restore method is presented. The computation and the use of global models is analysed in Section 2; the computation of topographic corrections, with particular care to the residual terrain correction, is presented in Section 3; the central solution or estimation of the anomalous potential either by solving a Molodensky's problem, or by applying the collocation theory, is examined in Section 4; the restore step is finally presented in Section 5.

Foreword

From April 22 to April 26, 1991 in Wiesbaden, at the XVI General Assembly of EGS, a Symposium on the Computation of the Geoid was held, convened by B. Heck and F. Sansò. This was somehow a new argument for these meetings, so it was decided to run it in such a way as to present the state of the art in several different branches of geoid computations (global modelling, local modelling, marine geoid, geophysical interpretation, etc.).

The effort was rewarded by a significant success, proved by a conspicuous attendance, so that we thought it was a good idea to collect the presented material and to issue it in a special number of *Surveys in Geophysics*.

This journal is supposed to be directed to all geophysicists, i.e. also to those who are not particularly acquainted with the basics of geoid computations, whence it was felt by the convenors as a duty to write papers which could serve as general introduction to the subject. This is actually one of those papers having at the same time the character of introduction and review of the matter. The paper is certainly not complete, in the sense that due to space limitations several items have been skipped and the discussion could not go too deep into the derivations; by the way we tried to make it selfconsistent and we hope we have done a useful work for the unexperienced reader.

1. Introduction: What is the Geoid and How it Can be Computed

The concept of geoid is a very central one in Geodesy, showing at first glance the tendency of this discipline to use quantities strictly related to both the geometry and the gravity field, in describing the earth.

Essentially the geoid is one particular level surface of the gravity field chosen as a reference in such a way as to be close to the sea surface. In this way the geoid represents on one side a geometrical object serving as a reference for defining several systems of "height" coordinates (geopotential numbers, orthometric heights, etc.), giving body to the intuitive idea of height over the sea level, and on the other side one way of representing the gravity field which, once known over an equipotential surface, can be propagated throughout the external space by using the harmonicity property of the pure gravitational part of the gravity potential and by solving the relevant Dirichlet boundary value problem (B.V.P.).

Naturally the geoid itself is our main unknown object so we must somehow add some information (the first information being that it is an equipotential surface, i.e. the geopotential W is constant on it) to arrive at its complete identification.

One important information useful for this purpose, is the value of the modulus of the gravity field, g, accompanied with the specification of the horizontal position of the measure point.

Another fundamental tool to set up the theory of geoid determination is the knowledge of some reference field to start with, in our approximation process. This is for instance the normal gravity field of Somigliana-Pizzetti type (with potential U and intensity γ) which is composed, like the geopotential W, by a harmonic component and the centrifugal potential (i.e. the potential of the centrifugal force which appears when we describe the earth in a geocentric rotating system) and which accounts for the average ellipsoidal shape of the earth; in fact U is such as to admit a particular ellipsoid of revolution as normal equipotential surface. This ellipsoid is chosen in such a way as to fit as closely as possible the geoid, by suitably adapting the equatorial radius and the eccentricity; the geometrical centre of the ellipsoid is placed at the barycentre of the field, the symmetry axis is taken coinciding with the rotation axis.

The horizontal position of points in space is referred to this ellipsoid by projecting them orthogonally onto it, and is parameterized via the direction angles of the projecting normal, also known as geographic latitude and longitude.

The difference between the actual geopotential and the normal potential is the anomalous (or disturbing) potential $T(P) = W(P) - U(P)$ and it is in fact what we would like to know; loosely speaking we can say that T is about $10^{-5}/10^{-6}$ times smaller than U, so that we can reasonably use it as an infinitesimal quantity of the first order in a linearization process.[1]

So in principle we are led to define our problem in the form of a free B.V.P., namely:

given $W(P) = $ const, and $g(P)$, and given the horizontal position of P,

---

[1] In relating global quantities to perturbations we can follow the rule of thumb that the radius of the earth being of the order of $6 \cdot 10^6$ m, its perturbation, i.e. the geoid, is of the order of 60 m and its second order perturbation $10^{-5} \cdot 60$ m $= 0.6$ mm is in any way negligible.

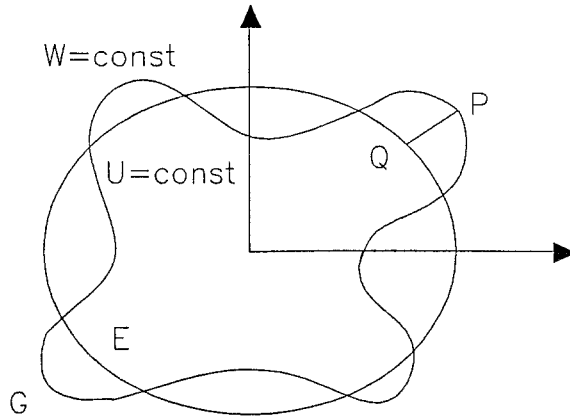Fig. 1.1.   The earth's geoid G and the corresponding ellipsoid E.

(ϕ_P, λ_P) at all the points of the unknown geoid, find the height N_Q of the point P above the ellipsoid (cf. Figure 1.1).

Remark 1.1

The height of the geoid N_Q is also known as geoid undulation.

Remark 1.2

Once N_Q (i.e. the surface G) would be known, then by claiming that $W(P) - \frac{1}{2}(X_P^2 + Y_P^2)$ is a harmonic function known on the known surface G, we would be able to recover $W(P)$ outside G by solving a Dirichlet problem.

This last remark shows that there must be an intimate relation between N_Q and the potential $W(P)$ itself, which is in fact the case as one can see by linearizing the identity:

$$W(P) = U(Q),\tag{1.1}$$

thus arriving at the famous Bruns' relation and (cf. Heiskanen and Moritz, 1981)

$$N(Q) = \frac{T(Q)}{\gamma(Q)}.\tag{1.2}$$

Bruns' relation, taking into account that $\gamma(Q)$ is known by a closed formula (cf. Heiskanen and Moritz, 1981), shows that knowing $N(Q)$ or $T(P)$ is one and the same thing; so we can say that determining the geoid is the same as finding the anomalous potential.

   This remark helps in making one step forward in solving one particular complication, i.e. to take into account that on continental areas the geoid is (almost) always below the physical surface of the earth; this means that there are masses between
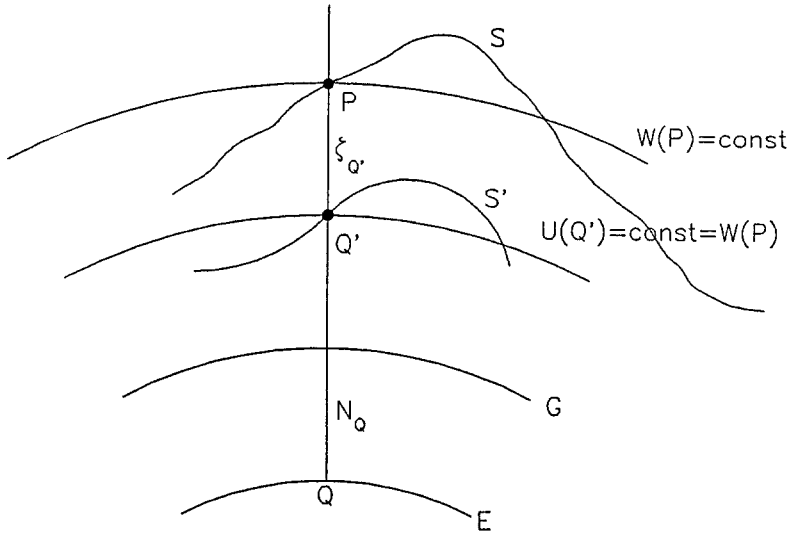
Fig. 1.2. Definition of the telluroid, i.e. the surface S' of Q' points when P varies over S.

the surface where we can perform measurements and the geoid itself so that T cannot be anymore harmonic in this intermediate layer. However as far as our ultimate goal is to know the anomalous potential T outside the masses, we can always use surface measurements to set up a suitable B.V.P. which naturally will refer to a surface more complicated than the simple ellipsoid.

Once T is known we can compute the separation between equipotential surfaces of W and U but at the ground level; this separation ζ_P is called the height anomaly and plays the same role as the geoid undulation, but at the level of the actual surface of the earth.

As it was for N_Q, also ζ_P is related to $T(P)$ by the Bruns' relation, i.e.

$$\zeta = \frac{T}{\gamma} \tag{1.3}$$

with the only difference that ζ is now the vertical distance between P and the footprint of the vertical through P, down to a point Q' where the relation

$$U(Q') = W(P)$$

is satisfied (cfr. Fig. 1.2).

Summarizing we can state the following problem:

> given the horizontal position of P (i.e. the point Q in Fig. 1.2) and the potential $W(P)$ and the gravity modulus $g(P)$ we want to find the surface S, i.e. the ellipsoidal height h_P and the potential W outside S.

This problem is very difficult to be treated in this general formulation, however

we can always linearize it by introducing the telluroid S' (cf. Fig. 1.2) and using the Taylor expansion

$$g(P) \cong \gamma(Q') + \frac{\partial \gamma}{\partial h} \zeta_{Q'} - \frac{\partial T}{\partial h}$$

together with (1.3) to obtain (cf. Heiskanen and Moritz, 1981)

$$-\frac{\partial T}{\partial h} + \frac{\partial \gamma / \partial h}{\gamma} T = \Delta g = g(P) - \gamma(Q') , \qquad (1.4)$$

where Δg is the so-called (surface) free air gravity anomaly.

So the problem has become now to find T harmonic outside S' satisfying the boundary condition (1.4).

With one further simplification, the so-called spherical approximation, the boundary condition (1.4) becomes

$$-\frac{\partial T}{\partial r} - \frac{2}{r} T = \Delta g \text{ (on } S') \qquad (1.5)$$

and the problem is known then as the "simple Molodensky's problem". When T is recovered by solving in some way our B.V.P., ζ can also be computed via (1.3); at this point N_Q, when necessary, can be obtained by observing that

$$h_P = H_P + N_Q = h_{Q'} + \zeta_{Q'} , \qquad (1.6)$$

where H_P is the so-called orthonormal height, i.e. the piece of plumb line lying between P itself and the geoid G.

Remark 1.3

With the present space technology, especially after the development of the GPS system, it is also possible to assume that we know directly the surface S, i.e. Q and h_P as well; in addition we know for instance $g(P)$ and we want to determine $T(P)$. This leads to the so-called fixed boundary, geodetic B.V.P. which after linearization becomes simply

$$-\frac{\partial T}{\partial h} = g(P) - \gamma(P) = \delta g(P) , \qquad (1.7)$$

where $\delta g(P)$ is the so-called true gravity anomaly, or gravity disturbance.

By using the same approximation as in (1.5), formula (1.7) becomes

$$-\frac{\partial T}{\partial r} = \delta g . \qquad (1.8)$$

Once T is derived by solving the B.V.P., it can be downward continued for instance with a linear formula like

$$T(Q) = T(P) + \delta g \cdot h_P , \tag{1.9}$$

so that we can compute N_Q by (1.2) and we finally get

$$H_P = h_P - N_Q . \tag{1.10}$$

This procedure is known as GPS levelling because it allows to compute the orthometric height of P from GPS observations and is therefore very important nowadays.

Summarizing we can say that the problem of computing the geoid has been reduced to the problem of retrieving the anomalous potential by solving a suitable B.V.P.; namely if we know $W(P)$ and $g(P)$ we use the harmonicity of T and the boundary condition (1.4) (or (1.5)) on the telluroid S'; if we know h_P and $g(P)$ we use the harmonicity of T and the boundary condition (1.7) (or (1.8)) on the actual Earth's surface S.

Yet this remains a formidable problem to be solved in practice even in the simplest formulations; in fact in either cases the boundary remains an extremely complicated object which must be defined at least with a resolution of the order of 1 km if one is not willing to accept a significant loss in the gravimetric signal.

In such a case it seems necessary to follow a procedure of successive approximations where the field is reconstructed step by step with different resolutions.

The procedure which is actually standard goes as follows:

1) first we make a global approximation of the field by using global data sets suitably organized in block-wise averages; in this way we try to achieve a knowledge of the gravity field with a maximum resolution down to about 100 km;

2) after subtracting the global model to all available data, the residual field is considered as being more local in nature, in the sense that only those mass anomalies which are closer to the analysed area contribute significantly to set up the "local" anomalous gravity field; predominant in this signal is that part which is due to the specific shape of the topography, known as topographic correction (T.C.). In particular T.C. is responsible for the very high frequency features of the signal with wavelengths ranging from some dozens of kilometers down to some hundreds meters and amplitudes ranging from 10^2 mgal to zero. This part is usually computed by suitable fast numerical techniques applied to Newton integrals over a field of prisms covering a somewhat larger area; the height of the prisms is generally computed between the actual surface and some smoothed reference surface, because this layer of matter, of variable width, is in fact the one which mostly explains the high frequency part of the signal. This approach is known as residual terrain modelling (R.T.M.);

3) the part of the gravity field with spectrum ranging in resolution say between

$100 \sim 200$ km and $5 \sim 20$ km (i.e. the medium wavelength part) as well as all the residual gravity which comes from modelling errors in steps 1 and 2 give rise to the last part of the signal to be suitably interpolated by a harmonic function. The characteristic of this signal is to be "local" (i.e. the long wave-length part should have very little power) and at the same time smooth over the area analyzed, because the very short wavelengths have been subtracted The main interpolation strategies are two:

a) the local B.V.P. approach, which consists in assuming that the anomalous signal is different from zero only in the area where we have it. After reconducting the B.V.P. by some continuation technique to a problem for the sphere, or even for the plane, we can apply explicit integral formulas, where actually the integration is extended only to the local area;

b) the collocation approach in which the field is considered as a stochastic process stationary and isotropic and harmonic through space, down to an internal sphere (the so-called Bjerhammar sphere); a kind of optimal inter-polation is then applied, which is a straightforward generalization of the Wiener–Kolmogorov principle of minimum mean square estimation error.

The rest of the paper is devoted to introducing the basic definitions and well established results necessary to perform steps 1 through 3; a much more specific and updated material will be presented in dedicated papers in the same volume.

2. Global Models

A global model is a mathematical model of the anomalous gravity field represented by means of a harmonic function depending on finitely many degrees of freedom, suitably adapted to fit the actual data.

Many representations are possible, however the ones most commonly used are:

a) truncated series of spherical harmonic expansions;
b) point masses models.

The second type of modelling in particular is often used in describing the dynamics of a satellite, while the first one is almost generally applied to describe the gravity field at the level of the earth's surface or close by, so we shall concentrate here exclusively on this method.

Let us first explain what are the data which are considered as input for the construction of a global model. The first point to be made is that since we are aiming at retrieving just the global features of the gravity field, we don't need to consider in general real pointwise data, but we shall apply some kind of averaging with the purpose of reducing the high frequency noise and compressing the data

set down to some manageable number:[2] the larger the area on which we average, the lower is the resolution of the final estimate of the gravity field.

For practical purposes it is convenient to adopt a procedure of averaging data on equiangular blocks, i.e. geographical rectangles with fixed angular sides $\Delta\phi$, $\Delta\lambda$: nowadays grids of blocksize of $0.5° \times 0.5°$ are the finest available.

It is to be stressed that in this way the blocks are not equi-area, but rather their areas depend on latitude, so that a certain distortion is introduced in estimating the spectrum; how to eliminate this distortion is still a matter of discussion.

The data we are going to consider now are:

a) on land mean potential values (at ground level); to a mean block value $\overline{W}$ will correspond also a mean block value of the telluroid height computed, e.g. to the first order in h from

$$\overline{W} = \overline{U} = \overline{U}_0 - \gamma\bar{h}$$

(U_0 potential of the reference ellipsoid),

so this data set is equivalent to a data set of mean telluroid heights, defining the smooth surface with respect to which the relevant B.V.P. has to be solved.

Note that on sea, once time varying phenomena are suitably averaged, the residual value of $\bar{h}$ is so small (a few meters at most) that it can be generally neglected;

b) on land mean free air gravity anomaly values, $\overline{\Delta g}$; the mean value $\overline{\Delta g}$ is considered as the value of the smoothed field at the center of the block with height $\bar{h}$, according to point a);

c) on sea, apart from small seas like the Mediterranean, the value of the gravity modulus is generally not known, however, if by applying our corrections, we can consider the residual ocean surface as the geoid itself, we can exploit a particular kind of satellite observations, namely the height of the satellite over the ocean as measured by a radar-altimeter, to derive directly the geoid undulation N; in this way by using Bruns' relation written now as

$$T = \gamma \cdot N$$

we achieve the knowledge of T.

At this point a rational approach would be to consider a mixed B.V.P. with Δg given on land and T on the ocean, however this is still too complicated to be applied numerically. Actually what is done is to convert locally T into a $\overline{\Delta g}$

---

[2] On the same time, areas where only few data, even only one, are given, are still represented in our data set and contribute to the computation reducing the area where no data at all are available.

data set:[3] since this operation is improperly posed, we are forced to smooth out the very high frequencies of T, however, this is not too bad since our target is not so much to know Δg but rather block averaged values $\overline{\Delta g}$ which are not too much affected by low-pass filtering. Whence we can say that oceanic areas surveyed with radar altimetry can be filled with mean block values of $\overline{\Delta g}$ data;

d) $\overline{\Delta g}$ files are not complete, for the whole earth, but there are a lot of lacking data especially on land for the euroasiatic continent (e.g. Russia, China, etc.), because they are not yet delivered, and on sea for those regions which have not yet been surveyed by altimetric satellites (especially polar regions); in these areas special procedures are to be applied which combine a priori global models, derived from the observation of satellite orbits only, with relevant geophysical information.

Tests in areas where Δg was known have shown that up to 75% of the gravimetric signal can be recovered (cf. Rapp and Pavlis, 1990).

To conclude we can say that as input for the computation of a global model we can consider two data sets describing respectively the mean heights of the telluroid and the mean gravity anomalies on it.

The telluroid nearly coincides with the ellipsoid on seas, but it is generally above the ellipsoid on lands and since it mirrors the true topography, or better a smoothed form of it, its shape is still quite complicated so that it is impossible to find analytical expressions for the B.V.P. solution.

However if we observe that after all the telluroid deviates from the ellipsoid at most by a few parts in 10^{-4} of the radius of the earth, it becomes handy to think of an approximate solution based on the ellipsoidal boundary; this means that in a sense we imagine to analytically continue our field through masses; on the other hand there are theorems (Keldysh-Lavrentiev, Runge–Krarup, cf. Moritz, 1980) proving that this can be done at any preassigned approximation level.

So the idea is to look for some function T, harmonic down to the ellipsoid and satisfying *on the telluroid S'*, at least approximately, the boundary relation

$$-\frac{\partial T}{\partial h} + \frac{\partial \gamma / \partial h}{\gamma} \, T = \Delta g \,, \tag{2.1}$$

or better its discretized, averaged version, at the centres of $0.5° \times 0.5°$ blocks.

The general representation of a harmonic function in the exterior of the ellip-

---

[3] To guess how this is possible one can think of working on the tangent plane with the Fourier transform, so that $T(x, y)$ is transformed into $\tilde{T}(p_x, p_y)$; now it is easy to verify that if $T(x, y, z)$ has to be harmonic, then the Fourier transform of $\partial T/\partial z$ $(x, y, 0)$ is $\sqrt{p_X^2 + p_Y^2} \cdot \tilde{T}(p_X, p_Y)$ which can be obtained from the above, after a suitable filtering of high frequencies. Since $\partial T/\partial z|_{z=0}$ is essentially the gravity anomaly in planar approximation we see that we can get hold of it by back transforming $\sqrt{p_x^2 + p_y^2} \cdot \tilde{T}(p_x, p_y)$.

soid, can be found in analytical terms by separation of variables (cf. Heiskanen and Moritz, 1981) in the form of an ellipsoidal harmonics expansion

$$T = \sum_{0}^{+\infty} {}_n \sum_{-n}^{n} {}_m\, T_{nm} \frac{Q_{nm}\left(i\dfrac{u}{E}\right)}{Q_{nm}\left(i\dfrac{b}{E}\right)} Y_{nm}(\theta, \lambda) ,^4 \tag{2.2}$$

where u, θ, λ are the ellipsoidal coordinates defined through

$$\begin{cases} x = \sqrt{u^2 + E^2} \sin\theta \cos\lambda \\ y = \sqrt{u^2 + E^2} \sin\theta \sin\lambda \\ z = u\cos\theta \end{cases} \tag{2.3}$$

$$(E = \sqrt{a^2 - b^2} = \text{linear eccentricity}) ,$$

However, in order to simplify the computations and in particular to avoid the calculation of Q_{nm} (iu/E) (cf. Thong and Grafarend, 1989), it is possible to perform a perturbative approximation of (2.2) in terms of the small parameter

$$e^2 = \frac{a^2 - b^2}{a^2} \quad (e = \text{eccentricity}) .$$

pushed just to the first order (i.e. neglecting terms in e^4). Our discussion here will be limited to this case while for accurate high degree models it might be advisable to push the perturbative computation to the e^4 terms (cf. Heck, 1991; Jekeli, 1988).

Taking also advantage of the inequality $h/R < 10^{-3}$ which certainly holds on our (smoothed) earth, so that $e^2(h/R)$ is negligible in our approximation, one can see (see Appendix) that (2.2) is essentially equivalent to a "purely spherical" form

$$T = \sum_{0}^{+\infty} {}_n \sum_{-n}^{n} {}_m\, T_{nm} \left(\frac{R}{R + h}\right)^{n+1} Y_{nm}(\theta, \lambda) , \tag{2.4}$$

i.e. the difference of (2.2) minus (2.4) (with equal T_{nm}) is of the order of e^4 or smaller; in (2.4) h has to be understood as the height above the ellipsoid, as shown in Figure 2.1.

It has to be stressed that if one wanted to use geographical coordinates (ϕ, λ)

---

[4] $Y_{nm}(\theta, \lambda)$ are the spherical harmonics

$$Y_{n,m} = \bar{P}_{n|m|}(\sin\theta) \begin{cases} \cos m\lambda & m \geq 0 \\ -\sin m\lambda & m < 0 \end{cases} \quad (|m| \leq n)$$

and $\bar{P}_{nm}(t)$, $Q_{nm}(t)$ are the so called Legendre associated functions of the first kind (defined in $|t| \leq 1$) and of the second kind (defined in $|t| \geq 1$) (cf. Heiskanen and Moritz, 1981).
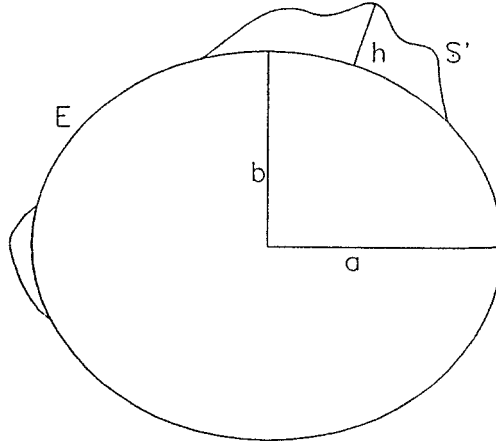
Fig. 2.1. E = ellipsoid; S' = averaged telluroid, $h/a < 10^{-3}$; (a, b) = semimajor and semiminor axes.

instead of the ellipsoidal (θ, λ), corrective terms in e^4 would appear in (2.4), in order to switch to an expansion in $Y_{nm}(\phi, \lambda)$ (cf. Heck, 1991).

We now come to the application of the boundary operator (2.1). Also here, to make things simpler, we approximate the coefficient $(\partial\gamma/\partial h)/\gamma$ to the first order in e^2 and h/R. To this aim we start from a fundamental relation of geometrical geodesy, also due to Bruns (cf. Heiskanen and Moritz, 1981) maintaining that

$$\frac{\partial\gamma}{\partial h} = -2\gamma J - 2\omega^2 , \tag{2.5}$$

where ω is the earth's angular velocity, and J is the mean curvature of the (normal in this case) level surface for the point P. If P is lying on the ellipsoid, we have[5]

$$J_0 = \frac{1}{2}\left(\frac{1}{M} + \frac{1}{N}\right) =$$

$$= \frac{1}{2}\frac{b}{a^2}\{(1 + e'^2 \sin^2\theta)^{3/2} + (1 + e'^2 \sin^2\theta)^{1/2}\} = \tag{2.6}$$

$$\cong \frac{1}{R}\left(1 - \frac{2}{3}e^2 + e^2 \sin^2\theta\right\} .$$

$$(R = \sqrt[3]{a^2 b}) .$$

---

[5] Remember that in our approximation

$$e'^2 = \frac{a^2 - b^2}{b^2} \cong \frac{a^2 - b^2}{a^2} = e^2.$$

Moreover, if we want to transport this curvature at the level h, we can use in our approximation the simple formula

$$J_h = J_0 \frac{R}{R+h} \cong \frac{1}{R}\left(1 - \frac{2}{3}e^2 + e^2 \sin^2\theta - \frac{h}{R}\right). \tag{2.7}$$

Furthermore it is easy to verify that (cf. Heiskanen and Moritz, 1981)

$$\omega^2 = \frac{\gamma}{R}\left(\omega^2 \frac{R}{\gamma}\right) \cong \frac{\gamma}{R}\left(\frac{\omega^2 a^2 b}{GM}\right) \cong \frac{1}{2}e^2 \frac{\gamma}{R}, \tag{2.8}$$

hence, summarizing (2.5), (2.7), (2.8), we can state that

$$\frac{\partial\gamma/\partial h}{\gamma} \cong -\frac{2}{R}\left(1 - \frac{1}{6}e^2 + e^2 \sin^2\theta - \frac{h}{R}\right). \tag{2.9}$$

Substituting (2.4) and (2.9) in (2.1) one gets, after some rearrangement of the coefficients,

$$\sum_{n,m} \frac{(n-1)}{R} T_{nm} Y_{nm}(\theta, \lambda) = \Delta g(\theta, \lambda) + \frac{1}{R}\sum_{n,m} T_{nm} Y_{nm}(\theta, \lambda) \times$$

$$\times \left\{\left[1 - \left(\frac{R}{R+h}\right)^{n+2}\right] \cdot (n-1) + 2e^2\left(\sin^2\theta - \frac{1}{6}\right)\right\}, \tag{2.10}$$

where the first neglected term is of the order of $(h/R)e^2$.

Equation (2.10) is already in a perturbative form which lends itself to a recursive solution (cf. Heck, 1991).

The main concern can be with the term $1 - (R/(R + h))^{n+2}$ which might be significantly close to 1 for very high degrees; on the other hand we already know that ultimately we are not going to use (2.10) for all n because in this step we are not aiming at the solution of our B.V.P. in one shot only, so it's a matter of seeing up to what a degree we really want to use it.

At degree 300 one finds that in those areas where $h \sim 6000$ m the right hand side of (2.10) will contribute $\sim 25\%$ of the signal, which is not a small amount indeed, but these areas are so small that we should not be bothered by them; for instance when $h \sim 1000$ m we see that the r.h.s. operator in (2.10) contributes less then 5% of the signal, which is a fairly small figure.

At degree 600 the same calculations would result respectively in 44% and 9% of the signal, already indicating that convergency might not be achieved in a small number of iterations.

Hence we are left with a very simple problem which is essentially to find the coefficients T_{nm} from

$$\sum_{n,m} \frac{(n-1)}{R} T_{nm} Y_{nm}(\theta, \lambda) = \Delta g(\theta, \lambda) \, , \tag{2.11}$$

which is very well known in the history of Geodesy as the Stokes problem.

Once the coefficients T_{nm} are determined from (2.11) we can feed them into the right hand side of (2.10), compute a corrected Δg known term and then solve again (2.11). As a very last step we can then say that (2.11) can be solved either harmonic coefficient by harmonic coefficient or by a corresponding suitable integral formula.

Since here we are mostly interested in constructing a finite degree model, we prefer the first solution, which is simply

$$T_{nm} = \frac{R}{(n-1)} \frac{1}{4\pi} \int \Delta g(\theta, \lambda) Y_{nm}(\theta, \lambda) \, d\sigma \, ; \tag{2.12}$$

generally the zero degree coefficient is taken as zero because we have adapted the geocentric gravitational constant to be equal for the normal and for the actual field. The first degree terms, corresponding to a shift of the coordinate origin, are also set to zero by choosing a geocentric ellipsoid.

A frequency analysis, exploiting the famous concept of Nyquist frequency adapted to the sphere (cf. Sacerdote and Sansò, 1991), shows that if we have $\delta° \times \delta°$ block averages, the coefficients (2.12) can be meaningfully computed up to a maximum degree

$$N_{\max} = \frac{180}{\delta} \, ; \tag{2.13}$$

so we have for instance models up to degree 180 with $1° \times 1°$ mean data, up to degree 360 with $0.5° \times 0.5°$ mean data etc.

With such coefficients finally a model potential T_m is computed on the ellipsoid ($h = 0$) and in the surrounding space by

$$T_M = \sum_2^{N_{\max}} {}_n \sum_{-n}^{n} {}_m T_{nm} \left(\frac{R}{R+h}\right)^{n+1} Y_{nm}(\theta, \lambda) \, ; \tag{2.14}$$

from such a model all the functionals of T (like height anomalies, deflections of the vertical, gravity anomalies, etc.) can be computed as well.

Remark 2.1

From the practical point of view many errors enter into a model like (2.14) computed from coefficients T_{nm} determined from (2.12); the most important of them however are related to a) gaps of data, which are quite significantly affecting T_{nm}, b) regional biases connected to non-uniform choices of absolute gravity values to which regional gravimetric networks are referred, as well as to other reference frames problems.

On the other hand a very important contribution to the knowledge of T_{nm} for low values of n (e.g. $n \leqslant 36$), comes from satellite observations which can be combined with ground data to strengthen our solution. We cannot treat this subject here, so we refer the interested reader to the very comprehensive review (Rapp, 1989), where also a good bibliography can be found.

3. Topographic Corrections

In this paragraph we study how the very high frequency part of the gravimetric signal is accounted for in terms of some kind of Newton's integration formula: hence we shall assume to analyse here the residual field obtained by subtracting from the actual data a global model.

The first basic statements we must make are:

a) we expect that most of the gravity field signal at low frequency (low degrees) has been subtracted while computing the global model, as we suppose to work here with the residual field,

b) the signal due to the density variations modulated by the topographic heights, namely the topographic signal, is indeed by far the most important component of the high frequency signal; more precisely only the "closest" topographic masses contribute to it.

The first statement is obvious, the second can be better understood by using Newton's integral

$$u(P) = G \int_{B_R} \frac{\rho(Q)}{r_{PQ}} \, dB_Q \, , \tag{3.1}$$

where B_R is a ball of radius R such as to enclose all the masses. Neglecting the ellipsoidal eccentricity, we include in (3.1) the layer of topographic masses, a few kilometers thick, where $\rho(Q)$ undergoes jumps of $\pm 2.67 \, \text{g/cm}^3$, corresponding to the mean density of the crust.

If we use the famous development (cf. Heiskanen and Moritz, 1981)

$$\frac{1}{r_{PQ}} = \sum \frac{r_Q^n}{r_P^{n+1}} \frac{Y_{nm}(P) Y_{nm}(Q)}{2n + 1} \tag{3.2}$$

into (3.1) with $r_P = R \geqslant r_Q$, we find

$$u(P) = G \sum \frac{Y_{nm}(P)}{2n + 1} \int_0^R dr \frac{r^{n+2}}{R^{n+1}} \rho_{nm}(r) \, , \tag{3.3}$$

where

$$\rho_{nm}(r) = \int_{\sigma} \rho(r, \sigma) Y_{nm}(\sigma) \, d\sigma \, . \tag{3.4}$$

We interpret (3.3) and (3.4) by saying that the birth of a harmonic coefficient of u is given by the sum over all depths $D = R - r$ of the corresponding coefficient $\rho_{nm}(r)$ weighted with the weight

$$\left(\frac{r}{R}\right)^{n+2} dr = \left(1 - \frac{D}{R}\right)^{n+2} dr \, . \tag{3.5}$$

Note should be taken that since Y_{nm} is orthogonal to $Y_{00} = 1$, $\rho_{nm}(r)$ are in fact the harmonic coefficients of the density variation $\Delta\rho$, i.e. of $\rho(r, \sigma)$ subtracted the average of ρ at the corresponding depth.

Now if we consider that the biggest density contrasts in the upper layers are due to density discontinuities, like the Moho, also related to some isostatic equilibrium, we can compute the contribution of such a contrast, placed, e.g. at 30 km depth, at degree 300 (or higher).

If the isostatic equilibrium has to hold, the amplitude Δr_M of the Moho wandering, will be related to the thickness Δr_t of the topographic layer, exactly by the relation (cf. Heiskanen and Moritz, 1981)

$$\frac{\Delta\rho_t \, \Delta r_t}{\Delta\rho_M \, \Delta r_M} = 1 \, . \tag{3.6}$$

so that the relative contribution of the two layers to the creation of one coefficient say at degree 360, is

$$\frac{\Delta\rho_t \, \Delta r_t}{\Delta\rho_M \, \Delta r_M \left(1 - \dfrac{D_M}{R}\right)^{380}} = \frac{1}{\left(1 - \dfrac{30}{6378}\right)^{380}} \cong 5.6 \, , \tag{3.7}$$

at degree 600 the same expression would attain the value $\sim$17. This shows that for wavelengths below 100 km by far prevailing is the effect of topographic masses, justifying our effort to make a special separate computation of it.

Furthermore let us compute, as an example, what is the gravimetric effect of a parallelepiped with base 20×20 km and 2 km high, placed, e.g. 100 km distant from the measurement point (cf. Figure 3.1).

Since the large component of the vector $\mathbf{g}$ is in the z direction, it is only the z component of the attraction of this mass, which is contributing to the modulus of g, and we have (cf. Figure 3.1)

$$\delta g_Z \cong \frac{GMZ}{L^3} \cong 1.4 \cdot 10^{-2} \, \text{mgal} \, ,$$
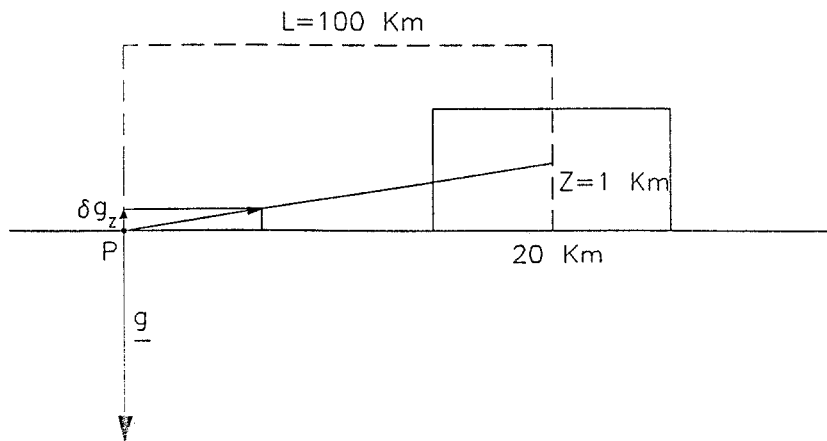
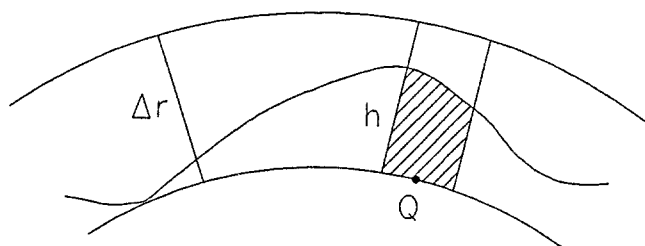Fig. 3.1. The attraction of a large prism at a distant point P.

Fig. 3.2. Mass distribution in the topographic layer.

at a distance of 50 km the same mass would cause a variation in g of 0.1 mgal; at 200 km distance we would get $\delta g_z \cong 1.7\,\mu$gal.

This shows that at 100 Km distance we still have to take into account this mass, although its detailed shape will be of no great importance; while placed further away it gives no significant contribution.

In this discussion, of introductory character, we won't dwell on the problem of the curvature of the earth, which in the illustrated example would cause a shift in Z of the mass at 100 km distance of about 0.8 km; as a matter of fact, when needed, this effect can be easily corrected for. For closer masses however, which give rise to most of the signal, this correction is immaterial as it decreases quadratically with distance; on the other hand these masses need to be represented with greater detail, because even smaller variations of their shape can have a significant effect.

With these observations in mind one can afford to sketch strategies for the computation of the topographic effect, however not before making a very important remark.

Let us consider the topographic masses contained in the uppermost layer of thickness, say, Δr (cf. Figure 3.2); if the mass contained in the cylinder of base

$R^2 \, \mathrm{d}\sigma$ and height $h(Q)$ is spread all over the column of height Δr, the density switches from the average value $\bar{\rho} = 2.67 \, \mathrm{g/cm^3}$ to a value $\rho(Q)$ such that

$$\rho(Q) \, \Delta r = \bar{\rho} h(Q) \, . \tag{3.8}$$

If we substitute this expression in the uppermost part of the integral (3.3) we find a "topographic contribution" to $u(P)$ of the type

$$u_t(P) = GR \sum \frac{Y_{nm}(P)}{2n + 1} \Delta r \, \rho_{t,nm} \tag{3.9}$$

with

$$\Delta r \, \rho_{t,nm} = \int_\sigma \Delta r \, \rho(Q) Y_{nm}(\sigma_Q) \, \mathrm{d}\sigma = \bar{\rho} \int_\sigma h(Q) Y_{nm}(\sigma) \, \mathrm{d}\sigma \, . \tag{3.10}$$

The topographic potential $u_t(P)$ in the space, as opposed to the pure surface expression (3.9), is then

$$u_t(P) = G \sum \frac{Y_{nm}(Q)}{2n + 1} \Delta r \, \rho_{t,nm} \left(\frac{R}{r}\right)^{n+1} \tag{3.11}$$

and its radial derivative, i.e. the topographic contribution to the modulus of the gravity vector at a surface point, is

$$\delta g_t = -\left.\frac{\partial}{\partial r} u_t(P)\right|_{r=R} = G \sum \frac{n + 1}{2n + 1} Y_{nm}(Q) \, \Delta r \, \rho_{t,nm} \, . \tag{3.12}$$

Equations (3.10) and (3.12) show that, when we construct a global model, the effects of topography are absorbed into the determination of the coefficients of the global model, with the same resolution; in other words, at least at the rough level of approximation we are using here, if h is developed to degree and order N to supply some smoothed version $\tilde{h}$

$$\tilde{h}(Q) = \sum_{0}^{N} {}_{n,m} h_{nm} Y_{nm}(\sigma_Q) \, , \tag{3.13}$$

then the gravimetric effects of $\tilde{h}$ go directly into the determined coefficients of the potential model up to degree and order N.

With this statement in mind we understand that what we are looking for, in order to interpolate the high frequency part of the topographic effects, is not really the computation of the whole effect but rather of its residual part caused by the high frequency features of the difference between actual heights h and smoothed heights $\tilde{h}$.

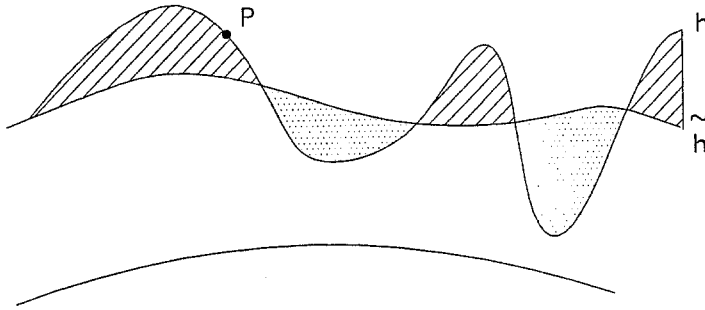So our aim is now to compute a residual terrain correction to the gravity

Fig. 3.3. Residual terrain correction (r.t.c.) corresponding to excess of density ρ (dashed areas) above $\tilde{h}$ or to lack of density ρ (dotted area) below $\tilde{h}$.

modulus; as shown in Figure 3.3, this correction can be positive or negative depending on both the excess or the lack of matter (which in turn depends on the positive or negative sign of $h - \tilde{h}$) and on the height of those masses with respect to the height of P.

There are mainly two strategies for the computation of this residual terrain correction; one is almost exact but requires longer computations, the other is more approximated but it is by far much quicker.

In both strategies the terrain is essentially represented as a collection of parallel-epipeda based on the ellipsoid, while the residual terrain is an analogous collection but now hanging in the air as shown in Figure 3.3; the bases of the parallelepiped (generally squares) define the "resolution" of this model and may vary in general between 100 m to 1 km; the heights of the parallelepiped are assumed to be the mean height of the terrain on the corresponding areas.

If nothing better is available, nowadays there exist worldwide mean elevation models with a resolution of $5' \times 5'$ like, for instance, TUG 87 (cf. Wieser, 1987).

According to the first strategy the attraction in the z direction is computed in P for each prism lying at a preassigned distance from it; this attraction is a function of $x_P - x_Q$, $y_P - y_Q$ and of $z_P - z_Q = h_P - \tilde{h}_Q$, as well as of the (fixed) side of the basis and of the height of the prism $h_Q - \tilde{h}_Q$.

All that is arranged in a lengthy but closed formula, suitable for automatic computation (cf. Sansò, Barzaghi and Tscherning, 1986). The same type of formula holds for the potential generated by the same prism of matter, what will be necessary in the final processing of our data.

As a remark, we notice that since we need to go, say, as far as $\sim 1°$ from P to compute its r.t.c., if we want to perform this computation at all points of an area A we need a terrain model on a larger area enclosing A.

What is not exact in what we said is that the prisms, though placed at the right altitude as shown in Figure 3.4, are slightly twisted as the "vertical" direction does not remain parallel to itself.

The second strategy goes through the computation of the "terrain effect" for
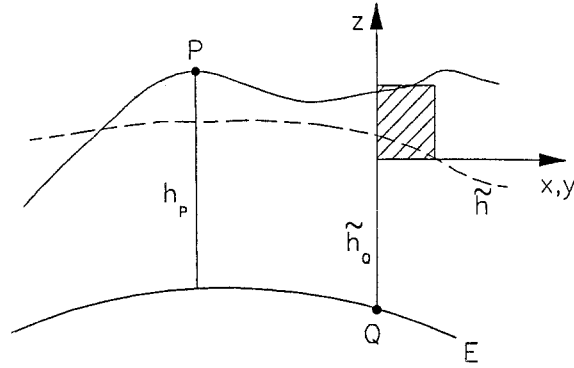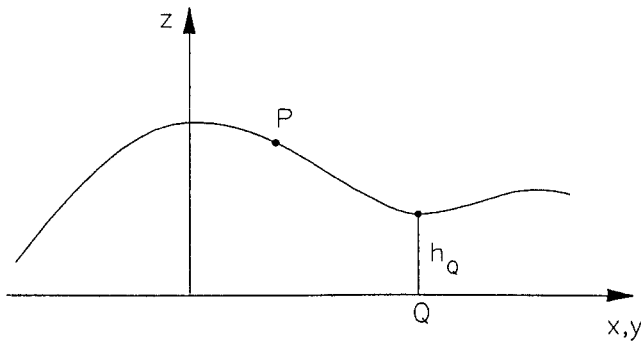
Fig. 3.4. The r.t.c. computed prism-wise.

Fig. 3.5. The terrain correction in planar approximation.

both the surfaces h and $\tilde{h}$ and then retrieves the r.t.c. as the difference between the two; it works in planar approximation.

Referring to Figure 3.5, we see that

$$\delta g_h(P) = -\frac{\partial}{\partial h_P} G\rho \int dx_Q\, dy_Q \int_o^{h_Q} dz\, \frac{1}{\sqrt{l^2 + (h_P - z)^2}} \qquad (3.14)$$

$$(l^2 = (x_P - x_Q)^2 + (y_p - y_Q)^2)$$

and since

$$-\frac{\partial}{\partial h_P}[l^2 + (h_P - z)^2]^{1/2} = \frac{\partial}{\partial z}[l^2 + (h_P - z)^2]^{1/2}\,,$$

after an integration we get

$$\delta g_h(P) = G\rho \int dx_Q\, dy_Q \left[\frac{1}{\sqrt{l^2 + (h_P - h_Q)^2}} - \frac{1}{\sqrt{l^2 + h_P^2}}\right]. \qquad (3.15)$$

Since we can put

$$\frac{1}{\sqrt{l^2 + (h_P - h_Q)^2}} \cong \frac{1}{l} - \frac{1}{2}\frac{(h_P - h_Q)^2}{l^3} \tag{3.16}$$

and since the first term in (3.16) and the second term in (3.15) will be the same for δg_h and for the analogous expression of $\delta g_{\bar{h}}$, which has to be subtracted from the first, we find for the residual terrain correction δg_t,

$$\delta g_t = \delta g_h - \delta g_{\bar{h}}, \cong -\frac{1}{2}G\rho\int \frac{-2h_P(h_Q - \bar{h}_Q) + (h_Q^2 - \tilde{h}_Q^2 4)}{l^3}\,\mathrm{d}x\,\mathrm{d}y\,. \tag{3.17}$$

If we call $\bar{h}$ the mean height in the computation area, with one further approximation we can write

$$h_Q^2 - \tilde{h}_Q^2 = (h_Q - \tilde{h}_Q)(h_Q + \tilde{h}_Q) \cong 2\bar{h}(h_Q - \tilde{h}_Q)\,, \tag{3.18}$$

which substituted in (3.17), after a rearrangement, yields

$$\delta g_t = G\rho(h_P - \bar{h})\int \frac{(h_P - \tilde{h}_Q)}{l^3}\,\mathrm{d}x\,\mathrm{d}y\,. \tag{3.19}$$

The point here is that the integral in (3.19), as well as the integrals into which (3.17) can be broken if we want to maintain that form, is a typical convolution integral in the xy plane as l depends only on $x_P - x_Q$, $y_P - y_Q$; therefore it lends itself quite naturally to a fast computation via a Fourier technique (cf. Schwarz *et al.*, 1990).

Remark 3.1

Some attention has to be payed to the fact that the kernel $1/l^3$ is not straightforwardly integrable in the plane so that its Fourier transform has to be interpreted as a distribution.

This is the effect of the approximations (3.16) and (3.19) combined; in fact the kernel

$$\frac{(h_P - h_Q)^2}{l^3}$$

is still integrable if the slope of the terrain is bounded ($<45°$). On the other hand it is known that the use of a slightly regularized kernel instead of $1/l^3$ gives numerically reliable answers.

We conclude the paragraph by saying that with analogous approximations we can compute the contribution of the residual topographic masses to the gravity potential.

Just as a hint we observe that we can write

$$\delta u_t = G\rho \int dx\, dy \int_{\tilde{h}_Q}^{h_Q} dz \frac{1}{\sqrt{l^2 + (h_P - z)^2}} \cong$$

$$\cong G\rho \int dx\, dy \frac{h_Q - \tilde{h}_Q}{\sqrt{l^2 + (h_P - \bar{h})^2}} \cong$$

$$\cong G\rho \int dx\, dy \frac{h_Q - \tilde{h}_Q}{l}, \tag{3.20}$$

the second approximation being rougher than the first. In any way higher order terms can be computed by developments of the kind used in (3.16), getting more and more fitting evaluations.

4. The Core Solution

Let us summarize briefly what we have stated up to now and make the point in our way toward a solution, i.e. the determination of the anomalous potential T.

We have first observed that with global sets of data, suitably averaged, we can construct global models T_M expressed in a truncated series of spherical harmonics up to some degree and order N (e.g. $N = 360$); this model anomalous potential is such as to account for the long wavelength part of the gravity anomaly field which is our primary datum; so T_M satisfies the boundary relation

$$-\frac{\partial T_M}{\partial h} + \frac{\partial \gamma / \partial h}{\gamma} T_M = \Delta g_M \tag{4.1}$$

in some approximate and averaged way on the averaged boundary. Now assume you have a wealth of data, i.e. of gravity anomalies, at points P_i in a certain area A where we want to derive a very detailed solution T. At first we remove from our data that information which is already contained in the global model T_M

$$\Delta g_l(P_i) = \Delta g(P_i) - \Delta g_M(P_i), \tag{4.2}$$

thus obtaining a local datum $\Delta g_l(P_i)$.

Remark 4.1

There is an important difference between Δg_M in (4.1) and the same symbol in (4.2); in the first case Δg_M is a "datum" derived from averaging real observations and tributing it to points which are for instance, the centers of a geographic grid covering the whole earth, while $\Delta g_M(P_i)$ in (4.2) means exactly

$$\Delta g_M(P_i) = -\frac{\partial T_M}{\partial h} + \frac{\partial \gamma / \partial h}{\gamma} T_M \bigg|_{P = P_i}, \tag{4.3}$$

i.e. it is the expression (4.3) computed from the T_M previously determined, at the point P_i.

Since T_M already accounts for the gross features of T we expect Δg_l to contain mainly medium and short wavelenghts. As for this last part of Δg_l we expect, according to our discussion in Section 3, that it may be explained mainly by the residual topographic signal $\Delta g_t(P_i)$ so that, after subtracting it from Δg_l we are left with a residual field $\Delta g_r(P_i)$

$$\Delta g_r(P_i) = \Delta g_l(P_i) - \Delta g_t(P_i) = \Delta g(P_i) - \Delta g_M(P_i) - \Delta g_t(P_i) . \qquad (4.4)$$

Quite naturally we expect the spectrum of $\Delta g_r(P)$ to be strong mainly on the medium wavelengths (say between 100 km to 20 km) although there will be a residual signal (due to incorrect modelling) on both sides of the spectrum.

The point now is to try to solve the corresponding boundary value problem, using

$$\Delta g_r(P_i) = -\frac{\partial T_r}{\partial h} + \frac{\partial \gamma / \partial h}{\gamma} \left. T_r \right|_{P_i} \qquad (4.5)$$

as observation equations in the area A.

Remark 4.2

It might seem contradictory to say that we want to "solve" a "local" boundary value problem, since by definition such problems in potential theory can be solved only globally. On the other hand the point here is that by subtracting the global model we really "localize" the solution also in the sense that the Green function corresponding to our B.V.P. goes to zero in a much shorter distance if we subtract to it the first N degrees, so that the data $\Delta g_l(P_i)$ from distant areas do not affect too much the value of the solution in A.

There are many studies on the errors committed by truncating to zero the input data beyond a local area A, all of them referring in one way or another to the pioneering work of Kaula, Molodensky and Moritz of the late fifties, early sixties (cf. Heiskanen and Moritz, 1981); all these works introduce an averaging law in order to compute mean square errors, according to which the local area A is considered as a sample from a uniform distribution on the sphere.

This is the basic concept through which statistics enters into the estimation of the gravity field.

Remark 4.3

As we mentioned, in Δg_r and hence in T_r there are left high frequency components as well as low frequency components. The first type of a signal can be retrieved from the data in the local area A only up to a Nyquist frequency depending on the average distance between the observation points. The lower frequency part
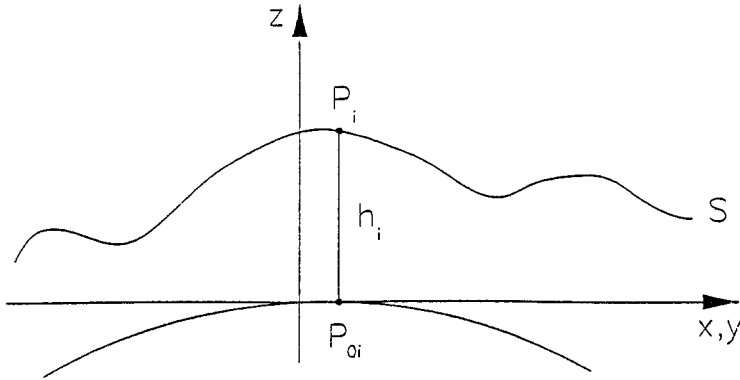
Fig. 4.1. Plane approximation for Molodensky's B.V.P.

can also be retrieved to some extent, depending on the size of the area A. In fact, would A cover the whole earth, we should be able to recover all the low frequencies of T, but then we should treat together millions of data, what we wanted to avoid; for a smaller A we can say, roughly, that we can recover wavelengths of the same size as the diameter of A, while longer wavelengths cause unadjustable distortions. This can be directly seen nowadays since we can test gravimetric geoids along traverses where the geoid is known by comparing the results of spirit levelling with those of GPS levelling.

These traverses have demonstrated the presence of relevant features (distortions of the order of 1 m) starting from 5° or 6°, down to shorter wavelengths. This shows that, with the currently available models, it isn't wise to make local computations in areas much smaller than 5° × 5°.

The "core solution" means that we want to solve (4.5) in the "local" area A; this is actually done in several ways two of which are the most important:

(a) by finding a pseudo-Molodensky solution of the B.V.P. reduced to the tangent plane approximation;
(b) by least squares collocation, i.e. by a statistical technique generalizing the Wiener-Kolmogorov optimal filtering concept.

(a) THE PSEUDO MOLODENSKY APPROACH

Since we can't go too deep here into the theory of Molodensky's B.V.P. we give a formulation of it, which corresponds to the first order Molodensky's solution, as shown in Heiskanen and Moritz, 1981.

Essentially the idea is to work in a plane approximation, as shown in Figure 4.1, where our problem gets the simplified form

$$-\frac{\partial T_r}{\partial z} = \Delta g_r .$$ (4.6)

to be solved from the surface S upward.

To this aim a first solution is elaborated by simply shifting the data $\Delta g_r(P_i)$ to the $z = 0$ plane, i.e. we solve

$$-\left.\frac{\partial T_r}{\partial z}\right|_{P_{0i}} = \Delta g_r(P_i)$$ (4.7)

thus obtaining a zero order approximation T_{0r}.

With this we can compute

$$\left.\frac{\partial \Delta g}{\partial z}\right|_{P_i} \cong -\left.\frac{\partial^2 T_{0r}}{\partial z^2}\right|_{P_i}$$ (4.8)

and we can make a much more realistic backward continuation of $\Delta g_r(P_i)$ by setting

$$\Delta g_r(P_{0i}) \cong \Delta g_r(P_i) - \frac{\partial^2 T_{0r}}{\partial z^2} \cdot h_i .$$ (4.9)

We can now compute a first order approximation T_{1r} by solving

$$-\left.\frac{\partial T_{1r}}{\partial z}\right|_{P_{0i}} = \Delta g_r(P_i) - \frac{\partial^2 T_{0r}}{\partial z^2} \cdot h_i .$$ (4.10)

The whole process can be iterated although generally T_{1r} is often an accurate enough solution. Whence, to put our solution in numbers, we are committed only to solve a B.V.P. of the type

$$-\frac{\partial T}{\partial z} = \Delta g$$ (4.11)

on the x, y plane.

This is easily achieved in terms of an integral formula, since according to the second Green's identity, when P is on the x, y plane,

$$T(P) = \frac{1}{2\pi} \int \left[\frac{1}{r_{PQ}} \left(-\frac{\partial T}{\partial z_Q} \right) - T_0 \left(-\frac{\partial}{\partial z_Q} \frac{1}{r_{PQ}} \right) \right] dx\, dy ;$$ (4.12)

on the other hand

$$-\frac{\partial}{\partial z_Q} \frac{1}{r_{PQ}} = -\frac{z_P - z_Q}{r_{PQ}^3} ,$$

which reduces to zero, when both P, Q belong to the x, y plane (i.e. $z_P = z_Q = 0$).

The sought solution is then simply

$$T(P) = \frac{1}{2\pi} \int \frac{\Delta g(Q)}{r_{PQ}} \, dx \, dy \, . \tag{4.13}$$

If the data $\Delta g_r(P_i)$ are gridded, i.e. the horizontal projections of P_i lie on a regular grid on the x, y plane, (4.13) can be discretized and, being an integral of convolution type, it can be computed quite efficiently by a Fast Fourier technique (cf. Schwarz *et al.*, 1990).

(b) The least squares collocation approach

In this approach the residual gravity potential T_r is considered as a stochastic process homogeneous and isotropic on the Earth's sphere.

How this is possible has been object of long debate and has been fully proved to be consistently acceptable in Sansò, 1986.

One way of making T_r stochastic is to take the harmonic representation

$$T_r = \frac{GM}{R} \sum_{n,m} T_{r,nm} \left(\frac{R}{r}\right)^{n+1} Y_{nm}(\sigma) \tag{4.14}$$

and consider $\{T_{r,nm}\}$ as a sequence of uncorrelated stochastic variables with zero mean and variances

$$\sigma^2(T_{r,nm}) = \frac{1}{2n + 1} \sum_{-n}^{n} {}_m T^2_{r,nm} \,.^{6} \tag{4.15}$$

The constants

$$\sigma^2_n(T_r) = (2n + 1)\sigma^2(T_{r,nm}) = \sum_{-n}^{n} {}_m T^2_{r,nm} \tag{4.16}$$

are called degree variances and have the meaning of the mean power per degree.

With this in mind, we see that the covariance function of T_r has the form

$$C_{TT}(P, Q) = E\{T_r(P)T_r(Q)\} =$$

$$= \left(\frac{GM}{R}\right)^2 \sum_n \sigma^2_n(T_r) \left(\frac{R^2}{r_P r_Q}\right)^{n+1} P_n(\cos \Psi_{PQ}) \,, \tag{4.17}$$

where the summation rule

---

[6] This shows that though uncorrelated $T_{r,nm}$ cannot be independent as the sum of their squares has to be constant.

$$\sum_{-n}^{n} {}_{m} Y_{nm}(\sigma_P) Y_{nm}(\sigma_Q) = (2n + 1)P_n(\cos \Psi_{PQ})$$

has been exploited.

On the other hand, since in spherical approximation

$$\Delta g_r = -\frac{\partial T_r}{\partial r} - \frac{2}{r} T_r = \frac{GM}{R^2} \sum_{n,m} (n - 1) \left(\frac{R}{r}\right)^{n+2} Y_{n,m}(\sigma) . \qquad (4.18)$$

we see that the covariance function C_{TT} is also analytically related to the covariance function $C_{\Delta g \Delta g}$ as well as to the crosscovariance $C_{T\Delta g}$. In fact

$$C_{\Delta g \Delta g}(P, Q) = \left(\frac{GM}{R^2}\right)^2 \sum_n (n - 1)^2 \sigma_n^2(T_r) \times$$

$$\times \left(\frac{R^2}{r_P r_Q}\right)^{n+2} P_n(\cos \Psi_{PQ}) , \qquad (4.19)$$

$$C_{T\Delta g}(P, Q) = \frac{GM}{R} \frac{GM}{R^2} \sum_n (n - 1)\sigma_n^2(T_r) \left(\frac{R}{r_p}\right)^{n+1} \times$$

$$\times \left(\frac{R}{r_Q}\right)^{n+2} P_n(\cos \Psi_{PQ}) . \qquad (4.20)$$

As one can see, in particular we have

$$\sigma_n^2(\Delta g) = \left(\frac{n - 1}{R}\right)^2 \sigma_n^2(T) \qquad (4.21)$$

and, given either of these two, the three functions (4.17), (4.19), (4.20) can be computed.

Remark 4.4

Assume that T_r is lacking in the coefficients of some degree n; correspondingly we have $\sigma_n^2(T) = 0$, i.e. the same degree is absent in C_{TT} as well as in $C_{\Delta g \Delta g}$; since T_r has been obtained by subtracting from T the global model T_M, complete up to degree N, we would be pushed by this remark to search the covariance function of Δg in the form (4.19) with n starting from $N + 1$ on. As a matter of fact since T_M contains model errors it is common practice to start the summation of (4.19) from a degree lower than N.

The empirical estimate of the covariance function of Δg_r from data is a delicate step in this approach and it is often performed in the following way: first of all

empirical values of the covariance are estimated by averaging products $\Delta g_r(P_i)\, \Delta g_r(P_i)$ over points P_i, P_j lying at a spherical distance Ψ_{ij},

$$\overline{\Psi} - \Delta \leq \Psi_{ij} \leq \overline{\Psi} + \Delta .$$

In this way a sequence of estimated $\widehat{C}(\overline{\Psi})$ with arguments $\overline{\Psi}$ lagged 2Δ is found; this sequence however is in general not consistent with the properties of positive definiteness implied by the theoretical form (4.19). Then we try to adapt (4.19) to the empirical estimates, i.e. we solve

$$\widehat{C}((2k+1)\Delta) = \left(\frac{GM}{R}\right)^2 \sum_N^{+\infty} n \left(\frac{R}{\bar{r}}\right)^{2n+4} \times$$

$$\times \; \sigma_n^2(\Delta g) P_n[\cos(k+1)\Delta] + \nu_k \tag{4.22}$$

using as shape parameters, N, which is in any case taken smaller than the maximum degree of T_n; $(R/\bar{r})$, i.e. the damping factor where $\bar{r}$ is taken as the mean radius of the earth R_E plus the mean altitude of the area under analysis and R (the so-called Bjerhammar radius) is the unknown parameter that has to be chosen smaller than R_E; $\sigma_n^2(\Delta g)$, the degree variances which are often modelled as rational functions of n (going to zero when $n \to \infty$) depending also on some parameters which are used to adjust (4.22). A typical shape of $\sigma_n^2(\Delta g)$ can be (Tscherning and Rapp, 1974)

$$\sigma_n^2(\Delta g) = \frac{A(n-1)}{(n-2)(n+B)}, \tag{4.23}$$

where A and B are the sought parameters; an advantage of (4.23), or of similar forms, is that the corresponding series can be summed to give a closed expression.

Another point to be watched in modelling the covariance function is that the expression (4.22) computed at the origin be never larger than the empirical estimate $\widehat{C}(O)$, i.e.

$$\left(\frac{GM}{R}\right)^2 \sum_N^{+\infty} n \left(\frac{R}{\bar{r}}\right)^{2n+4} \sigma_n^2(\Delta g) \leq \frac{1}{m} \sum_1^m {}_i \, \Delta g(P_i)^2 . \tag{4.24}$$

(m = number of measuring points);

often (4.24) is in fact imposed as an equality, when the shape of the model allows

[7] In a more refined processing of the covariance function, the theoretical model is represented as two sums, one up to N_{max} (the max degree of T_M) and the other from N_{max} on; in the first sum $\sigma^2(\Delta g_r)$ represent the degree variances of the errors of T_M which are modelled with some a priori formula, while for $n > N_{max}$ models like (4.23) are used. The relative weight between the two sums is a powerful parameter to shape the covariance $C_{\Delta g \Delta g}$ taking the place of N in (4.22).

for this choice, otherwise a discontinuity in the origin is allowed, interpreted as a modelling noise σ_ν^2.

Once the covariance $C_{\Delta g \Delta g}$ has been modelled, the crosscovariance $C_{T\Delta g}$ can also be computed analytically so that the following estimator of T_r at any point P is available:

$$\widehat{T}_r(P) = \sum_{i,k}^{m} C_{T\Delta g}(P, P_i)\{C_{\Delta g\Delta g}(P_i, P_k) + \sigma_\nu^2 \delta_{ik}\}^{(-1)}\Delta g_r(P_k) , \qquad (4.25)$$

where by $\{C_{\Delta g\Delta g}(P_i, P_k) + \sigma_\nu^2 \delta_{ik}\}^{(-1)}$ we mean the i, k element of the matrix inverse of $\{C_{\Delta g\Delta g}(P_i, P_k) + \sigma_\nu^2 \delta_{ik}\}$.

The estimator (4.25) is known to be the "best" among linear estimators, in the sense that it minimizes the mean square estimation error of $T_r(P)$ in that class of estimators.

One of the serious drawbacks of an estimate like (4.25) was the big amount of calculations necessary to compute it, which was limiting its application to the contemporary treatment of no more than 2000 ~ 3000 data at a time; on the other hand it has been shown recently that when the observation points $\{P_i\}$ are gridded as well as the estimation points, one can exploit the particular features of the covariance function to speed up the computation of (4.25) in a way comparable to the spectral analysis realized with Fast Fourier techniques. So this problem is solved at these times and collocation solutions with 20–30.000 data have already been performed.

5. Restoring the Final Solution

If we analyze what we have done in the previous paragraphs, it can be summarized as follows: if we have a set of m observations Δg in an area A of the earth surface on which we want to compute the anomalous potential T, we have started first of all reducing Δg to residual values by subtracting a global model accounting for the long wavelength component of Δg and a topographic correction, to reduce its high frequency spectrum.

These first two steps are known as the "remove" process because it amounts to removing components from Δg in such a way as to be left with a local and smooth gravimetric signal.

Then the residual Δg_r has been elaborated to provide what we called the core solution i.e. T_r.

Finally what we must do to recover T is to add back to T_r those components of the anomalous potential corresponding to what we have subtracted in Δg: this process is known as "restore".

The whole procedure can be represented by the flow diagram in Figure 5.1.

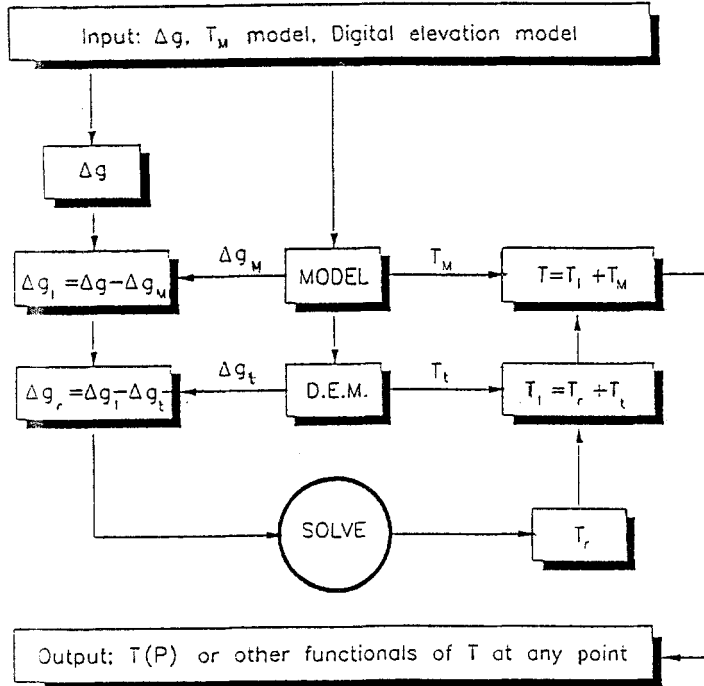Local solutions of this type are then to be patched together, until they will cover

Fig. 5.1. The remove-restore procedure.

all the world; at this point we could say to have accomplished our duty and to have solved completely the problem stated in Section 1 of this paper.

At present we are still far from this final goal, however things are really speeding up in the last years and we are well along this way.

Naturally not everything has been solved, not even theoretically, and an intense research activity is still going on in the area of geoid computations so that we expect in the near future to have more data, more methods, more computing facilities to get closer to our goal.

Appendix

We want to show in this appendix how a rigorous solution of the Laplace equation in the exterior of an ellipsoidal (with semimajor axis a and semiminor axis b) domain can be very simply approximated by a spherical formula when a perturbative computation is performed to the first order in the small parameter (first eccentricity)

$$e^2 = \frac{a^2 - b^2}{a^2} \, (\cong 7 \cdot 10^{-3}) ;$$

throughout this appendix, as we consider only linear terms in e^2, we will put also (second eccentricity)

$$e'^2 = \frac{a^2 - b^2}{b^2} \cong e^2 \,.$$

One starting point is the Laplace equation in ellipsoidal coordinates (cf. Heiskanen and Moritz, 1981)

$$\begin{cases} x = \sqrt{u^2 + E^2} \sin \theta \cos \lambda \\ y = \sqrt{u^2 + E^2} \sin \theta \sin \lambda \\ z = u \cos \theta \end{cases} \tag{A.1}$$

$(E = \sqrt{a^2 - b^2} =$ linear eccentricity, $0 \le \theta \le \pi$, $0 \le \lambda < 2\pi, u \ge b)$

which reads

$$(u^2 + E^2)\frac{\partial^2 T}{\partial u^2} + 2u \frac{\partial T}{\partial u} + \frac{\partial^2 T}{\partial \theta^2} + \cot \theta \frac{\partial T}{\partial \theta} +$$

$$+ \frac{u^2 + E^2 \cos^2 \theta}{(u^2 + E^2) \sin^2 \theta} \frac{\partial^2 T}{\partial \lambda^2} = 0 \,. \tag{A.2}$$

Introducing the dimensionless coordinate

$$s = \frac{u}{b} \qquad (s \ge 1)$$

and noting that

$$\frac{E^2}{b^2} = e'^2 \cong e^2 \,.$$

after developing (A.2) to the first order in e^2, we receive

$$s^2 \frac{\partial^2 T}{\partial s^2} + 2s \frac{\partial T}{\partial s} + \frac{\partial^2 T}{\partial \theta^2} + \cot \theta \frac{1}{\sin^2 \theta} \frac{\partial^2 T}{\partial \lambda^2} +$$

$$+ e^2 \left(\frac{\partial^2 T}{\partial s^2} - \frac{1}{s^2} \frac{\partial^2 T}{\partial \lambda^2} \right) = 0 \,. \tag{A.3}$$

Equation (A.3) is to be integrated in the "spherical" domain (in the space s, θ, λ) exterior to sphere $s = 1$ and is quite suited to a perturbative approach, i.e. to looking for solutions (to the order e^2) in the form $T = T_0 + e^2 T'$. Since (A.3) reduces exactly to a spherical Laplace equation in the triplet (s, θ, λ), if we set

$e^2 = 0$, we see that if we like to find a fundamental set of solutions, it is reasonable to look for functions of the form

$$T_{n,m}(s, \theta, \lambda) = \frac{Y_{nm}(\theta, \lambda)}{s^{n+1}} + e^2 T'_{n,m} , \tag{A.4}$$

which substituted in (A.3) yield, by comparing terms in e^2,

$$s^2 \frac{\partial^2}{\partial s^2} T'_{n,m} + 2s \frac{\partial}{\partial s} T'_{n,m} + \frac{\partial^2}{\partial \theta^2} T'_{n,m} + \cot \theta \frac{\partial}{\partial \theta} T'_{n,m} +$$

$$+ \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \lambda^2} T'_{n,m} =$$

$$= - \frac{(n + 1)(n + 2) + m^2}{s^2} \frac{Y_{nm}(\theta, \lambda)}{s^{n+1}} . \tag{A.5}$$

It is easy to verify that a set of solutions of (A.5) is given by

$$T'_{n,m} = - \frac{(n + 1)(n + 2) + m^2}{4n + 6} \frac{Y_{nm}(\theta, \lambda)}{s^{n+3}} , \tag{A.6}$$

so that we are lefrt with a set of (quasi-) solutions of (A.3) of the form

$$T_{n,m} = \frac{Y_{nm}}{s^{n+1}} \left[1 - \frac{e^2}{s^2} \frac{(n + 1)(n + 2) + m^2}{4n + 6} \right] . \tag{A.7}$$

It is remarkable that any solution

$$T = \sum A_{nm} T_{n,m}(s, \theta, \lambda) \tag{A.8}$$

reduces on the boundary $s = 1$ to a simple development in a series of spherical harmonics, the coefficients of which

$$T(\theta, \lambda) = \sum a_{nm} Y_{nm}(\theta, \lambda) \tag{A.9}$$

can be obtained by the ordinary orthogonality relations of the sequence $\{Y_{nm}(\theta, \lambda)\}$ with respect to the "area" measure $\sin \theta \, d\theta \, d\lambda$. From a_{nm} the true coefficients A_{nm} can be computed by the relation

$$A_{nm} = \left[1 - e^2 \frac{(n + 1)(n + 2) + m^2}{4n + 6} \right]^{-1} a_{nm} . \tag{A.10}$$

If we had now to sum (A.8) only on the ellipsoidal surface $s = 1$, by using (A.10) and (A.7) we see that we go back to (A.9), i.e. we can simply forget about the ellipsoidal shape and perform a simple spherical analysis (determination of a_{nm}) and synthesis (summation of (A.9)). Since however we often need our model

(A.8) for instance at the earth's surface, we can use (A.10) and (A.7) with s properly varying in accordance with the "height" of the computation point.

On the other hand $s = u/b$ can depart from 1 only for at most 10^{-3} as long as we remain at the earth's surface; whence terms like e^2/s^2 can be significantly put equal to e^2 when we neglect e^4-terms, so that our solution becomes simply

$$T = \sum \frac{Y_{nm}(\theta, \lambda)}{s^{n+1}} a_{nm} \,. \tag{A.11}$$

The function resembles strictly a "spherical" solution, were it not for the interpretation of the coordinates (s, θ, λ).

In particular if we go back to (A.1), divide by b, square and sum, we get

$$s^2 + e^2 \sin^2 \theta = \frac{r^2}{b^2} \,;$$

using

$$(s^2 + e^2) \sin^2 \theta = \frac{\rho^2}{b^2} = \frac{x^2 + y^2}{b^2}$$

in this relation and reordering we find

$$s^2 = \frac{\rho^2}{a^2} + \frac{z^2}{b^2} \quad (\rho^2 = x^2 + y^2) \,. \tag{A.12}$$

Substituting in this last expression the well known relations, involving the geographical latitude ϕ,

$$\rho^2 = (N + h)^2 \cos^2 \phi \qquad (N = a(1 - e^2 \sin^2 \phi)^{-1/2})$$

$$z^2 = [(1 - e^2)N + h]^2 \sin^2 \phi \,,$$

developing to the first order in e^2 and neglecting terms of the type $(h/a) \cdot e^2$ and similar, we arrive at the noteworthy relation

$$s^2 \cong \left(1 + \frac{h}{a}\right)^2 \,.$$

This shows essentially that to the first order of approximation in e^2 our solution T can be put in the form (cf. Heck, 1991; Jekeli, 1988)

$$T = \sum a_{nm} Y_{nm}(\theta, \lambda) \left(\frac{R}{R + h}\right)^{n+1}$$

with R being any radius between a and b,[8] for instance the mean radius of the earth $R = \sqrt[3]{a^2 b}$.

Our conclusion is that essentially a spherical approximation shows to be also a good ellipsoidal approximation, to the level of e^2 terms, on condition that the spherical coordinate r is taken as $r = R + h$, with h being the ellipsoidal height of points in space and to use the ellipsoidal colatitude θ. Naturally the same computation can (and often has to) be pushed further to higher powers in e^2; moreover the transition from the ellipsoidal colatitude to the geographic latitude would immediately cause a term of the order of e^2 to appear.

References

Jekeli C.: 1988, 'The Exact Transformation Between Ellipsoidal and Spherical Harmonic Expansion', *Man. Geod.* **13**(2).

Heck, B.: 1991, 'On the Linearized Boundary Value Problems of Physical Geodesy', OSU Report N. 407, February 1991.

Heiskanen, W. A. and Moritz, H.: 1981, *Physical Geodesy*, Repr. Inst. of Physical Geodesy TUG, Graz.

Moritz, H.: 1980, *Advanced Physical Geodesy*, H. Wichmann Verlag, Karlsruhe.

Rapp, R. H.: 1989, 'Combination of Satellite, Altimetric and Terrestrial Gravity Data', in *Theory of Satellite Geodesy and Gravity Field Determination*, Lecture Notes in Earth Science. Springer Verlag.

Rapp, R. H. and Pavlis, N. K.: 1990, 'The Development and Analysis of Geopotential Coefficient Models to Spherical Harmonic Degree 360', *Jour. of Geophysical Research*, **95**(13), Dec. 10.

Sacerdote, F. and Sansò, F.: 1991, 'Spectral Calculus and Moving Average Operators on the Sphere', Contribution to Geodetic Theory and Methodology, IAG Sect IV Report, XX Gen. Assembly of IUGG, Vienna.

Sansò, F., Barzaghi, R. and Tscherning, C. C.: 1986, 'Choice of Norm for the Density Distribution of the Earth', *Geophys. J.R. astr. Soc.* **87**(1).

Sansò, F.: 1986, Statistical Methods in Physical Geodesy', in *Mathematical and Numerical Techniques in Physical Geodesy*. Lecture Notes in Earth Sciences N. 7, (H. Sünkel (ed.)) Springer Verlag.

Schwarz, K. P., Sideris, M. G. and Forsberg, R.: 1990, 'The Use of FFT Techniques in Physical Geodesy', *Geophy. Jour. International* N. 100.

Thong, N. C. and Grafarend, E. W.: 1989, 'A Spheroidal Harmonic Model of the Terrestrial Gravitational Field', *Man. Geod.*, **14**(5).

Tscherning, C. C. and Rapp, R. H.: 1974, 'Closed Covariance Function Expressions for Gravity Anomalies, Geoid Undulations and Deflections of the Vertical Implied by Anomaly Degree Variance Models', O.S.U. Report N. 208, Columbus.

Wieser, M.: 1987, Das globale digitale Höhenmodell TUG 87; Inter. Bericht der Ab.f.Mat. und Datenverarbeitende Geod. TUG, Graz.

---

[8] In fact $s = 1 + \dfrac{h}{a} = 1 + \dfrac{h}{R}\dfrac{R}{a} = 1 + \dfrac{h}{R} + O\left(\dfrac{h}{R}e^2\right) \cong 1 + \dfrac{h}{R}$.