Journal of Geodesy (1995) 70:117–126

Journal of Geodesy

Numerical problems in the computation of ellipsoidal harmonics

G. Sona

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

Received 22 March 1994; Accepted 4 January 1995

Abstract

The correct use of ellipsoidal coordinates and related ellipsoidal harmonic functions can provide a representation of linearized Geodetic Boundary Value Problems (GBVP) much closer to the exact ones than what is usually done in spherical approximation: this becomes important in the present age, since terms of the type e^2N, possibly amounting to several dozens of centimetres, are nowadays observable.
Although the theory of ellipsoidal harmonics has been introduced into geodesy by several authors to treat gravity global models, the numerical computation of ellipsoidal harmonics of high degree and order seems to be more critical than it has been recognized. In particular, exact recursive relations display a quite unstable behaviour, no matter what normalization constants are used; it is only through particular representation of hypergeometric functions that it is possible to find a sound method for numerical manipulation. Also the asymptotic approximations, exploiting the smallness of the eccentricity, e^2, are analysed in relation to their critical behaviour for particular values of degree and order; it is shown that a limit layer theory can provide a simpler, better, and stable approximation of the exact values of ellipsoidal harmonics.

1 Introduction

The earth's anomalous gravitational potential T is commonly represented in the outer space as a truncated series of spherical harmonic functions :

$$T(r,\theta,\lambda)=\sum_{nm} T_{nm}\left(\frac{R}{r}\right)^{n+1} Y_{nm}(\theta,\lambda) \qquad (1)$$

whose coefficients T_{nm} can be obtained, at a first order of approximation in e^2 (where $e^2=(a^2-b^2)/a^2$, is the first eccentricity, a and b being the semimajor and semiminor ellipsoidal axes respectively) from surface measurements of gravity anomaly Δg and of gravitational potential W.
In order to use this representation and the surface (non spherical) data properly it is necessary to perform some approximations to refer to a formally spherical problem. Spherical approximations in solving GBVPs are well known, and it is also well known that in most cases surface data are reduced in one way or another to spherical data. The coefficients obtained with such a procedure (analysis) are therefore the spherical spectrum of the earth's anomalous potential.
Since terms of the order e^2N can amount to several tens of centimetres, and in the present age it becomes more and more important to work with the most precise representation available of the potential, they cannot be neglected anymore.
In order to reduce the errors of these "spherical" global models, different authors followed various procedures. The first one consists in pushing the approximations up to the first terms $O(e^2)$, to calculate corrective terms, named "ellipsoidal corrections", and add them to spherical models (Cruz,1986; Heck,1991; Rapp and Pavlis,1990). This method yields different kinds of corrections that in terms of geoid undulation can globally attain values of 50-80 cm. (over polar regions).
Another procedure starts from the GBVP expressed in ellipsoidal coordinates, so that the solution is harmonic outside a rotational ellipsoidal surface, and this leads to a potential global model written in terms of ellipsoidal harmonic functions. In this way, although the only approximation involved is the reduction from surface data to the ellipsoid, while the spherical approximation is avoided, the computation of the truncated series becomes more complicated, requiring the introduction of Legendre functions of second kind Q_{nm} (as we will see later on). To avoid the direct use of Q_{nm}, Jekeli (1988) and Gleason (1988, 1989) proposed to transform ellipsoidal coefficients in spherical coefficients, with an exact transformation that allows to use again spherical harmonic expansions.

This is however complicated enough to urge us to study again the direct use of ellipsoidal harmonic expansions and to look for a feasible and efficient way of computing them, in an exact way or in an approximate one (to a predetermined precision).
In literature basically two methods can be found to manipulate (truncated) series of ellipsoidal harmonics; one is the expansion of Q_{nm} in terms of e^2 (or E^2), which has the drawback of requiring an increasing number of perturbative terms when also the maximum degree is increasing (Heck, 1991); the other one is the expansion of Q_{nm} in terms of the inverse of its argument (use of hypergeometric series) which has in fact solved the problem, although with a large numerical effort (Jekeli,1988; Thong and Grafarend,1989; Thong,1989).
In this paper we collect a series of items concerning ellipsoidal harmonics manipulation and Q_{nm} computation which have been discovered in the effort of systematise the matter and, we hope, contribute to a better understanding of the subject.
After two short introductory paragraphs (§2, §3) we first analyse (§4) the problem of using iterative (recursive) relations, showing in a clear way why the instability inherent in this approach prevents us from practical implementation.
Then (§5) we review the hypergeometric series expansion of Q_{nm} : in this respect a solution has already been found and implemented (Thong and Grafarend,1989; Moritz,1980); yet we have found another solution, always based on analytical properties of hypergeometric functions, displaying superior convergence properties.
Finally, in §6, we tackle the method of perturbative expansions. Here, by exploiting an approach similar to that of the so-called "limit layer method", used in hydrology, we derive an approximate representation of Q_{nm} which combines an extreme simplicity with a very stable behaviour, avoiding the instabilities typical of all the other perturbative expansions.
Numerical comparisons and final comments are presented in §7.

2 Laplace Equation in Ellipsoidal Coordinates and its Solutions

We report here well known basic facts about ellipsoidal harmonics for the ease of the reader.
Using the orthogonal system of ellipsoidal coordinates (u, θ,λ) (see Fig.1; Heiskanen, Moritz, 1990), the Laplace equation, $\nabla^2 V=0$ can be written as:

$$\nabla^2 V=\frac{1}{u^2+E^2\cos^2\theta}\left[(u^2+E^2)\frac{\partial^2 V}{\partial u^2}+2u\frac{\partial V}{\partial u}+\frac{\partial^2 V}{\partial\theta^2}+\right.$$

$$\left.+\cot\theta\frac{\partial V}{\partial\theta}+\frac{u^2+E^2\cos^2\theta}{\left(u^2+E^2\right)\sin^2\theta}\frac{\partial^2 V}{\partial\lambda^2}\right]=0 \qquad (2)$$

that is

$$\left(u^2+E^2\right)\frac{\partial^2 V}{\partial u^2}+2u\frac{\partial V}{\partial u}+\nabla_\sigma^2 V-\frac{E^2}{\left(u^2+E^2\right)}\frac{\partial^2 V}{\partial\lambda^2}=0 \qquad (3)$$

note that letting E→0 equation (2) becomes the usual Laplace equation in spherical coordinates.

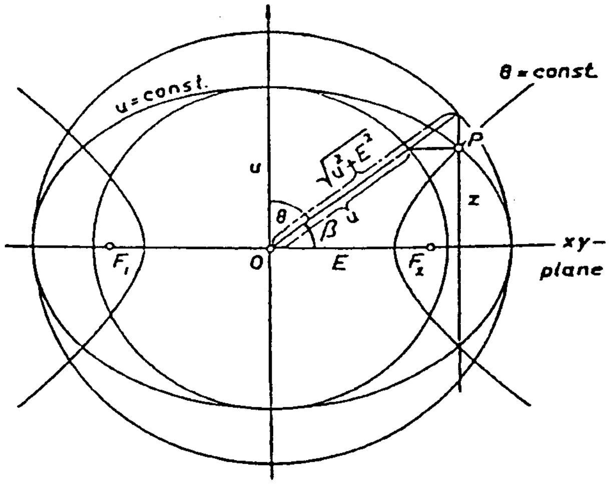

Fig.1.Ellipsoidal coordinates (Heiskanen and Moritz,1990).

Equation (2) can be solved by separation of variables, that is, looking for a solution of the form:

$$V(u,\theta,\lambda)=v(u)\Theta(\theta)\Lambda(\lambda) \qquad (4)$$

This method leads to solutions that can be combined into the general formula

$$V(u,\theta,\lambda)=\sum_{nm} v_{nm}(u)Y_{nm}(\theta,\lambda) \qquad (5)$$

where the functions $Y_{nm}(\theta,\lambda)$ are the well known spherical harmonics, but with a different argument: θ is in fact the ellipsoidal colatitude, the "reduced colatitude". The v_{nm}(u) functions are the solutions of the equation in the coordinate u:

$$(u^2+E^2)v''_{nm}+2uv'_{nm}-\left[n(n+1)-\frac{E^2m^2}{(u^2+E^2)}\right]v_{nm}=0 \qquad (6)$$

$$\left(v'=\frac{\partial v}{\partial u}, \qquad v''=\frac{\partial^2 v}{\partial u^2}\right)$$

that can be shown to be equivalent to a Legendre equation by using the pure imaginary variable t=iu/E; the Legendre associated functions of second kind Q_{nm} are then the suitable solutions of this equation (Heiskanen, Moritz, 1990).

The potential V can then be written as

$$V_{ext}(u,\theta,\lambda)=\sum_{nm}\frac{Q_{nm}\left(i\frac{u}{E}\right)}{Q_{nm}\left(i\frac{b}{E}\right)}Y_{nm}(\theta,\lambda) \qquad (7)$$

where b is the reference ellipsoid semiminor axis.
It can be proved that

$$\lim_{E\to 0}\frac{Q_{nm}\left(i\frac{u}{E}\right)}{Q_{nm}\left(i\frac{b}{E}\right)}=\left(\frac{R}{r}\right)^{n+1} \qquad (8)$$

which shows why Legendre functions of the second kind are involved.

3 General Characteristics of Q_{nm} functions

The change of variable t=iu/E in (6) leads to functions $Q_{nm}(t)$ that have non-real argument and that can be written as alternatively pure imaginary and real valued functions of u, e.g.

$$Q_{00}=-i\cot^{-1}\frac{u}{E} \qquad Q_{10}=\frac{u}{E}\cot^{-1}\frac{u}{E}-1$$

$$Q_{11}=-i\left(1+\left(\frac{u}{E}\right)^2\right)^{\frac{1}{2}}\left[\cot^{-1}\frac{u}{E}-\frac{\frac{u}{E}}{1+\left(\frac{u}{E}\right)^2}\right] \qquad (9)$$

$$Q_{20}=i\left[\left(\frac{3}{2}\left(\frac{u}{E}\right)^2+\frac{1}{2}\right)\cot^{-1}\frac{u}{E}-\frac{3}{2}\frac{u}{E}\right]$$

therefore, the functions $\overline{Q}_{nm}=\frac{Q_{nm}(iu/E)}{Q_{nm}(ib/E)}$, used in the expansion of V(u,θ,λ), are always real-valued.
For b≤u≤b(1+10^{-3}) that ratio is always less than 1 (∀n,m) and it is decreasing monotonically with increasing u. Moreover, at a fixed "height" u, the Q_{nm}(iu/E) functions have decreasing values (in modulus) with n; the same behaviour occurs for the $\overline{Q}_{nm}$ ratios.

4 Why Q_{nm} cannot be computed recursively

It is known (Hobson,1965) that for the functions Q_n the same recurrent relations as for P_n hold, and in particular:

$$Q_n(t)=\frac{2n-1}{n}tQ_{n-1}(t)-\frac{n-1}{n}Q_{n-2}(t) \qquad (10)$$

(where t=iu/E). It can also be proved by a skillful application of results contained in Nikiforov and Uvarov (1988) that another useful recurrence relation holding for Q_{nm} functions is:

$$Q_{nm}(t)=\left(1-t^2\right)^{-\frac{1}{2}}\left[(m-n-1)tQ_{n,m-1}+(m+n-1)Q_{n-1,m-1}\right] \qquad (11)$$

These two formulae seem to allow a very straightforward computation of Q_{nm} ∀ n,m, column by column, from the starting values of Q_{00}, Q_{10}.(Fig.2)
Note that, to avoid computations with complex numbers, it is possible to write real recurrent formulae for :

$$Q^R_{nm}(x)=(-i)^{n+m+1}Q_{nm}(t) \qquad \left(t=ix=i\tfrac{u}{E}\right), \qquad (12)$$

which for m=0 give

$$Q^R_n(x)=(-1)^n\frac{2n-1}{n}xQ^R_{n-1}(x)-\frac{n-1}{n}Q^R_{n-2}(x) \qquad (13)$$

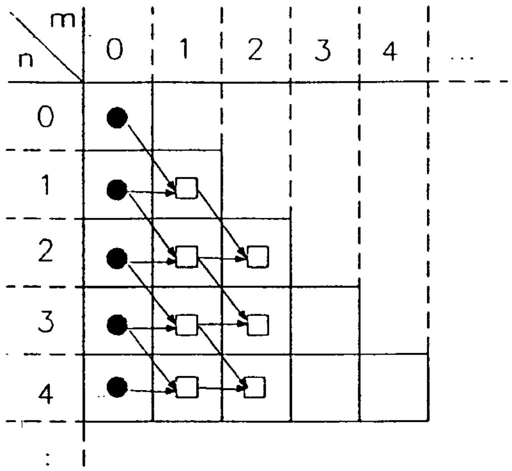

Fig.2.Recursive computation scheme using relation (11).

The first numerical problem faced in using relation (13) is that the values of $|Q_n^R|$ decrease very rapidly with increasing n, e.g. at u=b+4000m, we find:

Q_0^R= -8.203265•10^{-2} $\qquad Q_3^R$ =2.591545•10^{-6}

Q_6^R=1.336092•10^{-10} ,

so that, at high degrees, values out of computer range would stop the computation for underflow. To escape this problem, it would be necessary to choose a proper stabilising coefficient, dependent on degree n, as it was chosen for instance in Thong and Grafarend (1989); however, since it is the ratio $\overline{Q}_{nm}$ that has to be evaluated, and since this ratio is always ≤ 1, a recurrent relation has been deduced for $\overline{Q}_n$(u) from relation (13), and has been used in practical computations:

$$\overline{Q}_n(u) = \frac{(-1)^n(2n-1)\frac{u}{E}R_{n-1}\overline{Q}_{n-1}(u) - (n-1)\overline{Q}_{n-2}(u)}{(-1)^n(2n-1)\frac{b}{E}R_{n-1} - (n-1)} \quad (14)$$

where

$$R_n = \frac{Q_n^R(ib/E)}{Q_{n-1}^R(ib/E)} = (-1)^n(2n-1)\frac{b}{E} - \frac{n-1}{R_{n-1}} \quad (15)$$

Unfortunately, software implementation of recurrent formula (14) shows soon an intrinsic unstable behaviour, depending on the numerical computer representation and on the precision level used (real*4, real*8, real*16). At a critical degree n, which is higher the higher is the precision used, the functions $\overline{Q}_n$ deviate from the expected behaviour described in §3. As an example, on a VAX 3200, using FORTRAN REAL*8 representation, and for u=b+4000m, at degree n=6 the ratios (14) become greater than 1, and the functions start increasing with n; with a REAL*16 representation the same behaviour appears at n=13 (see Tab.1 and Tab.2.).

REAL*8

n	Q_n^R	$\overline{Q}_n$
0	-8.203265 10^{-2}	0.999374
1	-2.244126 10^{-3}	0.998748
2	+7.367441 10^{-5}	0.998122
3	+2.591545 10^{-6}	0.997496
4	-9.453652 10^{-8}	0.996873
5	-3.524908 10^{-9}	0.995252
6	+1.796759 10^{-10}	1.770888
7 ⇒	-1.037216 10^{-9}	-1.391526
8	-2.381141 10^{-8}	-1.409020

Tab.1. Values of Q_n^R functions and of $\overline{Q}_n$ ratios computed recursively with (13).

In order to understand this odd behaviour it is useful to study in more detail the recurrent relation for $Q_n^R(x)$ (13):

$$Q_n^R(x) = (-1)^n \frac{2n-1}{n} x Q_{n-1}^R(x) - \frac{n-1}{n} Q_{n-2}^R(x)$$

that, letting

$$\rho_n(x) = (-1)^n \frac{Q_n^R(x)}{Q_{n-1}^R(x)} \quad (16)$$

can be written as

$$\rho_n(x) = \frac{(2n-1)}{n}x + \frac{n-1}{n}\frac{1}{\rho_{n-1}(x)} = A_n x + \frac{B_n}{\rho_{n-1}(x)} \quad (17)$$

The asymptotic formula, for n→∞, is then

$$\rho(x) = 2x + \frac{1}{\rho(x)} \quad (18)$$

from which we obtain

$$\rho^2 - 2\rho x - 1 = 0$$

$$\rho_\infty = x \pm \sqrt{x^2+1} \; .$$

REAL*16

n	Q_n^R	$\overline{Q}_n$
.		
.		
4	-9.453643 10^{-8}	0.996871
5	-3.526970 10^{-9}	0.996246
6	+1.336092 10^{-10}	0.995622
7	+5.117651 10^{-12}	0.994998
8	-1.976696 10^{-13}	0.994374
9	-7.685238 10^{-15}	0.993751
10	+3.003770 10^{-16}	0.993128
11	+1.179554 10^{-17}	0.992802
12	-3.649881 10^{-19}	0.809643
13 ⇒	-2.351044 10^{-18}	5.474906 ⇐
14	-5.480961 10^{-17}	5.680913
15	+1.291040 10^{-15}	5.684118

Tab.2. Values of Q_n^R functions and of $\overline{Q}_n$ ratios computed recursively with (14).

These two solutions are graphically represented in Fig.3 , and a value of x=(b+4000m)/E≅12.16 yields:

$\rho_{\infty 1} = -0.041$, $\qquad \rho_{\infty 2} = 24.361$.

Since the $|Q_n|$ have to be decreasing with n, the right solution is $\rho_{\infty 1} = -0.041$.

The problem arises in performing numerical recurrent computations, with a starting value of $\rho_1 = -Q_1^R/Q_0^R$. The successive steps followed in recurrent computation are shown in Fig.4 for the asymptotic equation (18), and should lead to the stable solution point S. Actually, for small values of n, the hyperbolic horizontal asymptote and the hyperbole shape are always changing with n, so that the true behaviour is more complicated to show than the asymptotic one.

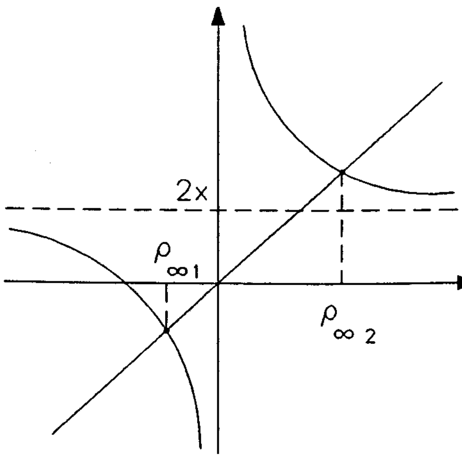

Fig.3.Solutions of asymptotic relation (18).

What is really obtained performing the computation is the asymptotic value of $\rho_{\infty 2}$, corresponding to the point P, and this means that, instead of following the procedure of Fig.4, we are rather in situation of Fig.5.

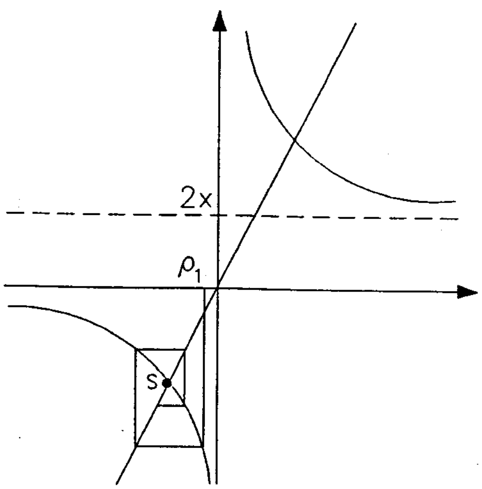

Fig.4.Convergence to $\rho_{\infty 1}$

A natural question arises at this point: why the recurrence relation (10) can be used to compute P_n and, on the contrary, it is unstable in computing Q_n?
A way to understand this difference is the following: for a complex variable t=ix, let us define $C_n(x)$ through

$$\rho_n(x) = (-1)^n \frac{Q_n^R(x)}{Q_{n-1}^R(x)} \equiv \frac{C_n(x)}{C_{n-1}(x)} \ , \quad C_0 = Q_0 \qquad (19)$$

so that $|C_n(x)| = |Q_n^R(x)|$.

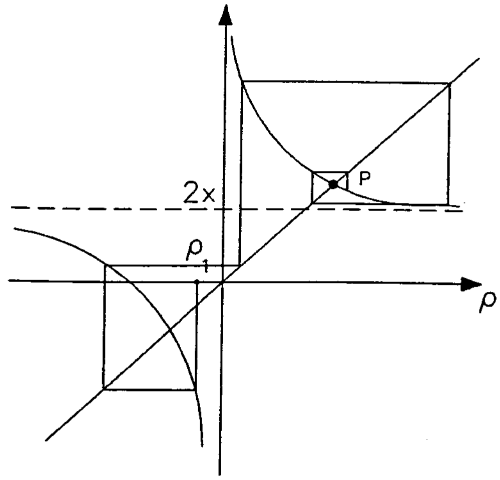

Fig.5.Convergence to $\rho_{\infty 2}$

The recurrence asymptotic relation for $C_n(x)$ differs from the recursive asymptotic relation for $P_n(x)$, in a sign:

$$C_n - 2xC_{n-1} - C_{n-2} = 0 \qquad (20)$$

$$P_n - 2xP_{n-1} + P_{n-2} = 0 \qquad (21)$$

These are linear, homogeneous, difference equations for which a general solution can be found letting P_n or C_n equal to $A\lambda^n$.
Solving in this way equation (21) we obtain:

$$P_n = A\lambda^n \quad \Rightarrow \quad \lambda^n - 2x\lambda^{n-1} + \lambda^{n-2} = 0 \qquad (22)$$
$$\lambda^2 - 2x\lambda + 1 = 0$$

that has solutions :

$$\lambda = x \pm i\sqrt{1-x^2} = e^{\pm i\theta} \qquad (x = \cos\theta \Rightarrow |x| \le 1)$$

This means a general integral of the form

$$P_n = Ae^{in\theta} + Be^{-in\theta} \qquad (23)$$

displaying a typical oscillating and bounded behaviour for $n \to \infty$.

As regards C_n:

$$C_n = A\lambda^n \quad \Rightarrow \quad \lambda^n - 2x\lambda^{n-1} - \lambda^{n-2} = 0 \qquad (24)$$
$$\lambda^2 - 2x\lambda - 1 = 0$$

the solutions are:

$$\lambda = x \pm \sqrt{1+x^2} = \left\langle \begin{array}{lll} \lambda_1 = -e^{-\alpha_1} & \lambda_1 < 0 & |\lambda_1| < 1 \\ \lambda_2 = +e^{+\alpha_2} & \lambda_2 > 0 & |\lambda_2| > 1 \end{array} \right.$$

that means a general solution of the form:

$$C_n = A\lambda_1^n + B\lambda_2^n = A(-1)^n e^{-n\alpha_1} + Be^{n\alpha_2} \quad (25)$$

These two general integrals (23 and 25) have to be interpreted as the shapes of the quantities P_n and C_n, as functions of n; both depend on two arbitrary constants A and B that can be fixed by imposing as initial conditions the values of P_0 and P_1, or C_0 and C_1. This means that computing P_n, Q_n through the iterative relations (20), (21) is the same thing as computing them from (23), or from (25) respectively, after having fixed A and B, as we said.
Since however the determination of A and B can be done only at the cost of committing some numerical error, e.g. the truncation error, we see that through (23) this error will generate a "small" oscillating component in the computation of P_n, while through (25) an error in B will give rise to an exploding component of the type $\delta Be^{\alpha n}$, which will sooner or later become overwhelming, no matter how small δB is.
We remark that similar, though slightly more complicated considerations can be drawn for Q_{nm} ($m\neq0$) and this explains why nobody has ever used these recursive relations in practice.

5 On the use of hypergeometric series

An integral expression that can be efficiently exploited to reach the target of Q_{nm} computation is the following. It is known (Hobson, 1965) that Q_n can be written as:

$$Q_n(z) = \sum_{k=n+1}^{+\infty} \frac{1}{z^k} \int_0^\pi P_n(\cos\theta)\sin k\theta d\theta \quad (26)$$

with $z = t + \sqrt{t^2-1}$, $t = ix$, and that the integral term is equal to zero for $k \le n$, and even values of k+n, and is equal to

$$2\frac{(k-n+1)(k-n+3)......(k+n-1)}{(k-n)(k-n+2)......(k+n)}$$

when k+n is odd. Using this representation, an expansion of the form:

$$Q_n(z) = \frac{A_1}{z^{n+1}} + \frac{A_3}{z^{n+3}} + \frac{A_5}{z^{n+5}} + \frac{A_7}{z^{n+7}} + ... =$$

$$= \frac{A_1}{z^{n+1}}\left(1 + \frac{A_3/A_1}{z^2} + \frac{A_5/A_1}{z^4} + \frac{A_7/A_1}{z^6} + ...\right) \quad (27)$$

is obtained.
This coincides with the expansion of Q_n in $1/z^2$ power, written through the hypergeometric function $F(\alpha,\beta,\gamma,1/z^2)$ (Hobson,1965):

$$F(\alpha,\beta,\gamma;x) = 1 + \frac{\alpha\cdot\beta}{1\cdot\gamma}x + \frac{\alpha(\alpha+1)\cdot\beta(\beta+1)}{1\cdot2\cdot\gamma(\gamma+1)}x^2 + ... \quad (28)$$

$$Q_n(z) = \frac{n!2^{n+1}}{3\cdot5\cdot7\cdot......\cdot(2n+1)}\cdot\frac{1}{z^{n+1}}F\left(\frac{1}{2}, n+1; n+3; \frac{1}{z^2}\right) \quad (29)$$

holding for $n\neq0$, $m=0$
The analogous general expression for $m\neq0$ is (Hobson,1965)

$$Q_{nm}(z) = (-i)^m \frac{n!(n+m)!2^{2n+m+1}}{(2n+1)!}\cdot\frac{1}{z^{n+m+1}}(x^2+1)^{+\frac{m}{2}}\cdot$$

$$\cdot F\left(\frac{1}{2}+m, n+m+1; n+\frac{3}{2}; \frac{1}{z^2}\right) \quad (30)$$

that can be written again in the form :

$$Q_{nm}(z) = (-i)^m \frac{(x^2+1)^{m/2}}{z^{m+n+1}} A_1\cdot$$

$$\left(1 + \frac{A_3/A_1}{z^2} + \frac{A_5/A_1}{z^4} + \frac{A_7/A_1}{z^6} + \frac{A_9/A_1}{z^8} + ...\right) \quad (31)$$

with coefficients:

$$A_1 = \frac{2^{2n+m+1}n!(n+m)!}{(2n+1)!}$$

$$A_3/A_1 = \frac{(2m+1)(n+m+1)}{2n+3} = C_1 \quad (32)$$

$$A_5/A_1 = \frac{(2m+1)(n+m+1)}{2n+3}\frac{(2m+3)(n+m+2)}{2n+5} = C_1\cdot C_2$$

$$A_7/A_1 = ... = C_1\cdot C_2\cdot C_3$$

that can be computed by the general rule :

$$A_k/A_1 = \prod_{l=1}^{k} C_l \qquad C_l = \frac{(2m+2l-1)(n+m+l)}{l(2n+2l+1)} \quad (33)$$

The successive terms are very easy to compute, as the series is quickly convergent, so that it is sufficient to sum few (5 or 6) terms to reach a good approximation of Q_{nm}.
In fact for high degree n, and for m=n:

$$Q_{nn}(z) = (-i)^n \frac{(x^2+1)^{n/2}}{z^{2n+1}}\frac{2^{3n+1}n!}{2n+1}\cdot$$

$$\left(1 + \frac{A_3/A_1}{z^2} + \frac{A_5/A_1}{z^4} + \frac{A_7/A_1}{z^6} + \frac{A_9/A_1}{z^8}\right) \quad (34)$$

the next terms, corresponding to $1/z^{10}$ and to $1/z^{12}$ are:

$$S_5 = \frac{A_{11}}{A_1}\frac{1}{z^{10}} \sim \frac{4}{15}n^5\frac{1}{z^{10}}$$

$$S_6 = \frac{A_{13}}{A_1}\frac{1}{z^{12}} \sim \frac{4}{45}n^6\frac{1}{z^{12}} \tag{35}$$

Remembering that $|z| \cong 2u/E$ and $u \geq b = 6356911.9$m, these two terms are small enough to be omitted even for high degree, as can be seen in Tab.3.
(It is clear that these coefficients increase with n, so that in order to obtain resulting functions with the same precision it would be necessary to keep an increasing number of terms, dependent on order and degree.)

	n= 100	n= 200	n= 360
S_5	$-3.6\cdot10^{-5}$	$-1.2\cdot10^{-3}$	$-2.2\cdot10^{-2}$
S_6	$2.0\cdot10^{-6}$	$1.3\cdot10^{-4}$	$4.4\cdot10^{-3}$

	n= 100	n= 200	n= 360
S_5 %	0.005 %	0.23 %	6.5 %
S_6%	0.00029%	0.025 %	1.39 %

Tab.3. Absolute and relative values of terms proportional to $1/z^{10}$ and $1/z^{12}$ (eqs.35) in expansion (34) (n=m).

It is worthwhile to observe that computing from (31,32) the ratio $\rho_n = (-1)^n Q^R_{n0}/Q^R_{n-1,0}$ it is easy to see that, for increasing n, ρ_n approaches asymptotically the correct value, that in our numerical example (u=b+4000m) is $\rho_\infty = 0.041$. This confirms, in a totally independent way, the reasoning reported in §4 on the recursive procedure.
For software implementation purposes it is still convenient to compute $\overline{Q}_{nm} = Q^R_{nm}(u)/Q^R_{nm}(b)$ so that the main terms out of brackets cancel out.
One has to observe that by performing the inverse transformation $t=(z^2+1)/2z$ one gets the hypergeometric expression of Q in terms of t, which is perfectly coinciding with what has been found in Jekeli (1988), Thong and Grafarend (1989) and Moritz (1980). In particular the series studied in Thong and Grafarend (1989) is essentially equivalent to ours, in terms of analytical behaviour, although the series (30) has superior speed of convergence since the variable z is approximately equal to 2t in the range of interest, so that the powers $1/z^{2n}$ go to zero faster than $1/t^{2n}$.
Once the proper values (correct, up to a sufficient precision) of Q_{nm} have been computed, at two different "heights" u_1 and u_2, (a third value is fixed: for u=b, Q=1) it is easy to find by quadratic interpolation the Q_{nm} values at different heights and the values of their first and second derivatives.

6 The perturbative approach

A classical procedure (Heck, 1991) that leads to approximate expressions of Q_{nm} up to first or higher orders in e^2 starts from searching perturbative solutions of Legendre equation (6), namely, letting

$$v(u) \cong v_0(u) + e^2 v_1(u) \tag{36}$$

or higher order expressions in e^2. It seems useful to repeat the first order computation also to point out its drawbacks.
Using (36) in (6), letting s=u/b and letting $e'^2 \approx e^2$ the Legendre equation becomes:

$$s^2 v_0'' + 2s v_0' - n(n+1)v_0 + e^2\left(v_0'' + \frac{m^2}{s^2+e^2}v_0 + s^2 v_1'' + \right.$$

$$\left. + 2s v_1' - n(n+1)v_1\right) + e^4\left(v_1'' + \frac{m^2}{s^2+e^2}v_1\right) = 0 \tag{37}$$

When the $O(e^4)$ terms are neglected, and the two coefficients of e^0 and e^2 terms are separately put equal to zero, the following solutions are found:

$$v_0(s) = \frac{1}{s^{n+1}}$$

$$v_1(s) = -\frac{(n+1)(n+2) + m^2/\left(s^2+e^2\right)}{4n+6} \tag{38}$$

The solutions of (37) are then:

$$v_{nm} = \left(\frac{b}{u}\right)^{n+1}\left(1 - e^2\left(\frac{b}{u}\right)^2 \frac{(n+1)(n+2)+m^2}{4n+6}\right) \tag{39}$$

This approximation can be used as far as

$$e^2\frac{(n+1)(n+2)+m^2}{4n+6} << 1 \tag{40}$$

that means n<<300.
Actually it is possible to see that under condition (40), and normalising these functions to the reference ellipsoid u=b:

$$\tilde{Q}^{(1)}_{nm} = \frac{v_{nm}(u)}{v_{nm}(b)} = \left(\frac{b}{u}\right)^{n+1}\left(\frac{1 - e^2\left(\frac{b}{u}\right)^2\frac{(n+1)(n+2)+m^2}{4n+6}}{1 - e^2\frac{(n+1)(n+2)+m^2}{4n+6}}\right) \tag{41}$$

it is possible to obtain a further approximated expression:

$$\tilde{Q}^{(1)}_{nm} \cong \left(\frac{b}{u}\right)^{n+1}\left[1+\left(1-\left(\frac{b}{u}\right)^2\right)e^2\frac{(n+1)(n+2)+m^2}{4n+6}\right] \cong$$

$$\cong \left(\frac{b}{u}\right)^{n+1}\left[1+e^2\left(\frac{u}{b}-1\right)\frac{(n+1)(n+2)+m^2}{2n+3}\right] \qquad (42)$$

It will be proved with a different approach that expression (42) can be used efficiently also for n>300, whilst the expression (41) has to satisfy condition (40), and moreover it has a critical denominator for some (n,m) with n>300.

The already mentioned different approach consists in considering that in all ground computations the variable s=u/b is ranging in a limited layer above the ellipsoidal surface (topographic layer), $1< s <1+10^{-3}$, namely $b \le u \le b+6400$m, that includes most of the actual topographic surface of the earth. In this way we can put in this layer

$$\frac{e^2}{s^2} \cong e^2, \qquad \frac{1}{s^2+e^2} \cong \frac{1}{1+e^2}$$

and the equation (6) can be transformed into

$$(1+e^2)s^2v'' + 2sv' - n(n+1)v + \frac{m^2e^2}{1+e^2}v = 0 \qquad (43)$$

Letting $v(s)=A/s^{\alpha}$ two solutions are obtained for α; the one with $\alpha \sim n+1$ is

$$\tilde{Q}^{(2)}_{nm} = v_{nm}(s) = \frac{1}{s^{n+1}} s^{e^2\frac{(n+1)(n+2)+m^2}{2n+1}} \qquad (44)$$

where A has been fixed so as to satisfy the condition that v(1)=1.
It is possible to expand the second factor, thus obtaining the expression

$$v_{nm}(s) \cong \frac{1}{s^{n+1}}\left(1+e^2(s-1)\frac{(n+1)(n+2)+m^2}{2n+1}\right) \qquad (45)$$

very similar to (42).
This expression holds as far as n and m satisfy

$$e^2(s-1)\frac{(n+1)(n+2)+m^2}{2n+1} << 1 \qquad (46)$$

that is, letting m=n, and for large n

$$\frac{2n^2+3n}{2n+1} \approx n << \frac{1}{e^2(s-1)} \cong \frac{1}{\frac{1}{150}\cdot 10^{-3}} = 150\cdot 10^3 \qquad (47)$$

meaning that, for actual purposes, there is no practical limit for n.
The same reasoning can be identically applied to (42), explaining why expression (42) can be used also for the case n>300.
The expression (44) is already normalised on the surface s=b, and is very simple to compute, moreover it never displays any singularity and it preserves the qualitative behaviour of Q_{nm} even for very high degrees and orders.
We remark that the change of variable from u to s and the subsequent manipulations of the differential equation, strongly resemble the limit layer method commonly used in hydrology (Wunsch, 1993).

7 Comparisons

In order to understand which computation method is the most convenient to compute ellipsoidal harmonics, numerical comparisons among truncated hypergeometric expansion (relation (31) truncated), classical perturbative approach (relation(41)) and "limit layer" approach (relation(44)) is needed.
A comparison between the two perturbative approaches can tell us if they give or give not similar results, if there is any kind of problem in computation procedures, in order to get a first idea about which of them could be preferred for practical computation. However, more significant comparisons are obviously performed between the truncated hypergeometric series (i.e., as we have seen, a very good approximation of the exact values of Q_{nm} functions), and each approximated relation ((41) and (44)), enabling us to investigate if one or both of them can closely reproduce the exact values of Q_{nm}, and to choose the best approximation of these functions, computable through an expression simpler than the hypergeometric expansion.
These comparisons have been performed by computing for each degree n, the mean quadratic difference:

$$\sigma_n = \sqrt{\frac{1}{2n+1}\sum_m \left(Q^{(a)}_{nm} - Q^{(b)}_{nm}\right)^2} \qquad (48)$$

at the altitude u=b+4000m.
Let us remark that in this case all the three objects we want to compare, namely $\tilde{Q}^{(1)}_{nm}$, $\tilde{Q}^{(2)}_{nm}$ and $\overline{Q}_{nm}$, are very close to 1 so that a renormalization of (48) is not necessary.
The numerical comparison between $\tilde{Q}^{(1)}_{nm}$ and $\tilde{Q}^{(2)}_{nm}$ (rel.(41) and (44)) gives results plotted in Fig.6, showing that for n<300 the two perturbative approaches yield very close values, but for larger n a critical behaviour for some couple (n,m) is evidenced by the sharp peaks occurring in σ_n.

A further comparison between the same $\tilde{Q}^{(1)}_{nm}$ (41) and $\overline{Q}_{nm}$ computed through the hypergeometric expansion (34) confirms that it is relation (41) that cannot be used for n>300 (Fig.7).
On the contrary, a satisfactory result has been obtained by

comparing between $\tilde{Q}^{(2)}_{nm}$ and $\overline{Q}_{nm}$ (Fig.8, full line): a very good agreement between the two methods is clearly proved and, even if the difference is obviously increasing with n, it attains very low values also for n=360.

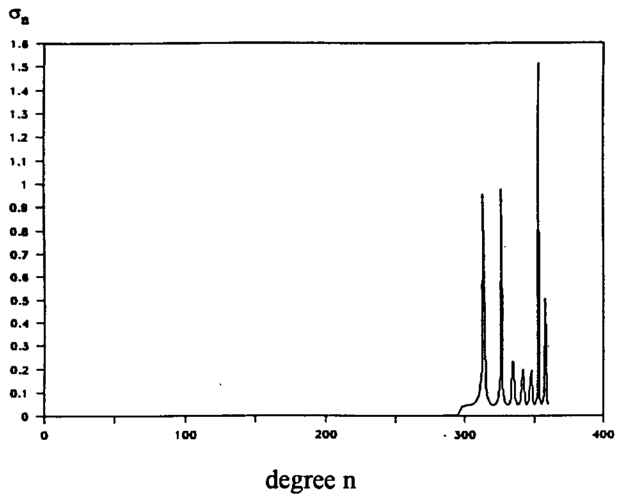

Fig.6.Mean quadratic difference between perturbative ($\tilde{Q}^{(1)}_{nm}$) and limit layer ($\tilde{Q}^{(2)}_{nm}$) approaches.

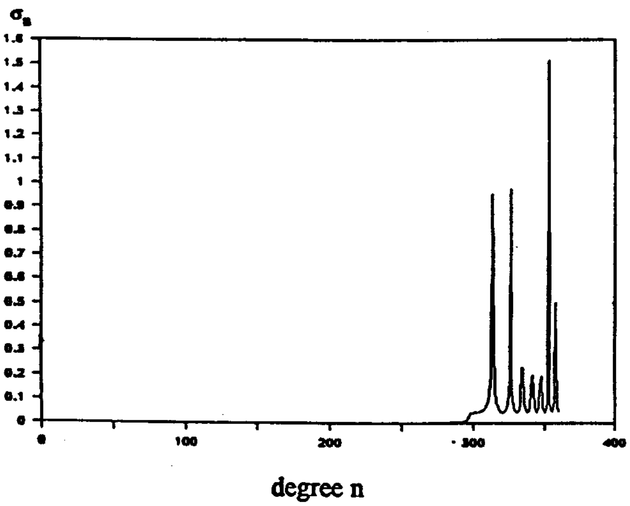

Fig.7.Mean quadratic difference between classical perturbative approach ($\tilde{Q}^{(1)}_{nm}$) and hypergeometric representation of Q_{nm} ($\overline{Q}_{nm}$ ratios)

Moreover, no computational irregularities are met in implementing relation (44). In Fig.8 for the purpose of a quick look comparison, σ_n up to degree 100 is also plotted, to show that $\tilde{Q}^{(2)}_{nm}$ is much closer to $\overline{Q}_{nm}$, also at n<<300, and not only at higher degrees, where $\tilde{Q}^{(1)}_{nm}$ is not reliable, as discussed previously.

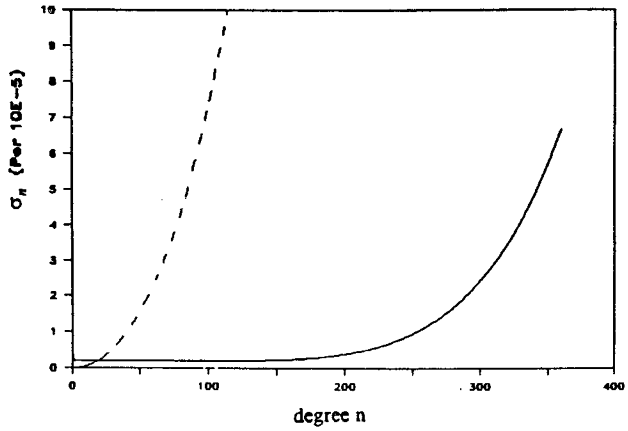

Fig.8. Mean quadratic difference between limit layer approach ($\tilde{Q}^{(2)}_{nm}$) and hypergeometric representation of Q_{nm} ($\overline{Q}_{nm}$ ratios) (full line). (For comparison σ_n of Fig 7 is reported as dashed line.)

Concluding we can say that although the item of the numerical computation and manipulation of ellipsoidal harmonics has already been solved, both theoretically and numerically, to a sufficient degree, yet we have shown that a systematization of the matter was possible and desirable. In particular, even a negative result like the statement that recursive relations cannot be used as such, because they generate a necessary instability in the numerical behaviour, was nice for us to learn: maybe a natural question is whether system theory cannot help to construct a stabilizer for these recursive relations. Furthermore the field of hypergeometric representation of Q_{nm} has been explored and beyond what was already known (Jekeli, 1988; Thong and Grafarend, 1989) a new convenient series expansion has been found and tested.

Finally, even in the most classical field of perturbation theory, by exploiting analogies with the hydrological limit layer method, it has been possible to find an extremely simple representation of Q_{nm} which seems to display extremely good approximation properties up to degree 360, so that we believe that probably this last solution is the most convenient for practical implementation at least when a relative error larger than 10^{-4} is sufficient.

References

Cruz, J.Y. (1986) Ellipsoidal corrections to potential coefficients obtained from gravity anomaly data on the ellipsoid, Report N.371, Department of Geodetic Science and Surveying, Ohio State University.

Gleason, D.M. (1988) Comparing ellipsoidal corrections to the transformation between the geopotential's spherical and ellipsoidal spectrum, Man.Geod.N.13, pp 114-129.

Gleason, D.M. (1989) Some notes on the evaluation of ellipsoidal and spherical harmonic coefficients from surface data, Man.Geod.N.14, pp 110-116.

Heck, B. (1991) On the linearized boundary value problems of physical geodesy, Report N.407, Department of Geodetic Science and Surveying, Ohio State University.

Heiskanen, W.A. and Moritz, H. (1990) Physical Geodesy, Reprint, Institute of Physical Geodesy, Technical University, Graz, Austria.

Hobson, E.W. (1965) The Theory of Spherical and Ellipsoidal Harmonics, Chelsea Publishing Company, New York.

Jekeli, C. (1988) The exact transformation between ellipsoidal and spherical harmonic expansions, Man.Geod.N.13, pp 106-113.

Moritz, H. (1980) Advanced Physical Geodesy, Herbert Wichmann Verlag Karlsruhe, Abacus Press, Tunbridge Wells Kent.

Nikiforov, A.F. and Uvarov, V.B. (1988) Special functions of mathematical physics, Birkhauser, Basel-Boston.

Rapp, H.R., and Pavlis, N.K. (1990) The Development and Analysis of Geopotential Coefficient Models to Spherical Harmonic Degree 360, Journal of Geophysical Research, Vol.95, N.B13, pp 21,885-21,911.

Sansò, F. and Sona G. (1993) The challenge of computing the geoid in the nineties, Surveys in Geophysics, Vol.14, N.4-5, Kluwer Academic Publ., Netherlands.

Thong, N.C. (1989) Simulation of gradiometry using spheroidal harmonic model of the gravitational field, Man. Geod. N.14, pp 404-417.

Thong, N.C. and Grafarend, E.W. (1989) A spheroidal harmonic model of the terrestrial gravitational field, Man. Geod. N.14, pp 285-304.

Wunsch, C. (1993) Physics of the Ocean Circulation, in Satellite Altimetry in Geodesy and Oceanography, Lecture Notes in Earth Science N.50, R.Rummel, F.Sansò (Eds.), Springer Verlag.