PART I: GRAVITY

0. Overview

The next few Sections cover some elementary matters with which many of you will be familiar: the idea of moment of inertia, which is usually first introduced in rotational dynamics. We use it for several additional purposes: first, as an exercise in the use of vectors and Einstein notation to tackle a problem in an arbitrary Cartesian coordinate system. Next we introduce the principal axes, which are part of an eigensystem of a symmetric linear operator — referring everything to the eigenvector axes simplifies the equations, and this idea will be seen again, but for a self-adjoint operator. There are many helpful analogies between the simple principal axis system and the more abstract one. We derive MacCullagh's formula, an approximation for the gravitational attraction of an arbitrary body in terms of the moments of inertia, our first contact with gravity.

The moments of inertia of a body give some idea of the internal distribution of matter, as well as being important in dynamics. We look at how they can be found for the Earth, from the gravitational field, which by itself is insufficient, together with astronomical measurements.

Next we have our first look at the shape of the Earth, how it is defined, and what it would be under a very simple, but reasonably accurate, approximation: Clairaut's approximation.

Then we return to the more technical matter of potential theory, the mathematical basis of a great deal of the theory for the Earth's gravitational and magnetic fields. We set the stage for a study of spherical harmonics, the most important functions in large-scale geophysics. We discuss the spherical harmonics as eigenfunctions of the surface Laplacian operator, which is self adjoint. The eigenfunctions are orthogonal and are in addition complete, making them a convenient basis for the representation of many geophysical properties on the surface of a sphere.

We show that the geoid height and the geopotential are essentially the same thing. We look at the long-wavelength geoid. Next we consider several questions in potential theory that involve thin layers at the surface, which leads naturally into Fourier theory on a flat Earth, and a brief discussion of isostasy and the bending of the lithosphere.

1. Moment of Inertia

We shall find it useful to employ *Einstein's summation convention:* in an expression with subscripts, a repeated subscript is summed over all its allowed values. Since the subscripts will always be drawn from the set {1, 2, 3} this means, for example

$$s_{ii} = \sum_{i=1}^{3} s_{ii} = s_{11} + s_{22} + s_{33} \tag{1}$$

and

$$A_{ij} x_j = \sum_{j=1}^{3} A_{ij} x_j \ . \tag{2}$$

The expression is taken to be valid for each of the un-summed subscripts, so the range of values $i = 1, 2, 3$ is implicit in (2). You may recall the interesting object, the *Kronecker delta,* defined by

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \tag{3}$$

It follows that

$$\delta_{ii} = 3 \tag{4}$$

and, perhaps more usefully, that

$$\delta_{ij} x_j = x_i \ . \tag{5}$$

Other notation we shall use includes: for a vector $\mathbf{r}$, we write its length as $|\mathbf{r}|$ or simply r. A unit vector will be written like this $\hat{\mathbf{u}}$, and of course $|\hat{\mathbf{u}}| = 1$. Components of vectors will be written r_i and the hat will be removed from unit vectors when they are written in components. For lots of examples of the use of the summation convention see Chapter 7 of *Foundations of Geomagnetism*, a copy of which has been placed on reserve.

Let us use this notation to obtain an expression for a familiar quantity, the *moment of inertia* of a body about an axis. A body B possesses an internal density distribution $\rho(\mathbf{s})$, where $\mathbf{s}$ is a position vector relative to the origin O. A unit vector $\hat{\mathbf{v}}$ through O defines an axis about which the moment of inertia is calculated; we imagine B being rotated about this axis.

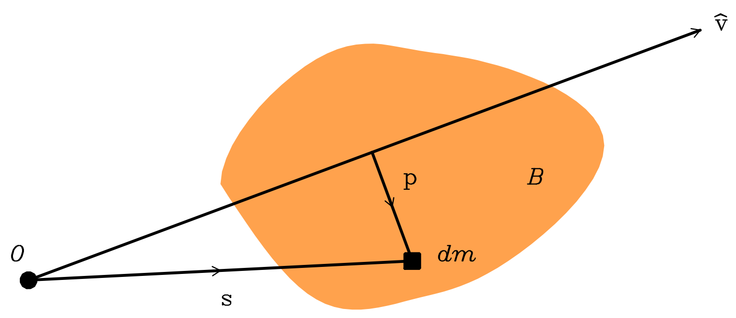


Figure 1

Then, the moment of inertia is defined by

$$I(\hat{\mathbf{v}}) = \int_B dm\ \mathbf{p} \cdot \mathbf{p} \tag{6}$$

$$= \int_B d^3\mathbf{s}\ \rho(\mathbf{s})\ \mathbf{p} \cdot \mathbf{p}\ . \tag{7}$$

where $\mathbf{p}$ is the vector from the axis to the mass element, as shown in the figure. Notice that the value of I depends on which axis we choose, and that the origin O is fixed and is not necessarily inside B.

We will now discover how I depends on the direction $\hat{\mathbf{v}}$. We take a Cartesian axis system with origin at O and express vectors in components relative to that system. We apply the Einstein summation convention:

$$\mathbf{p} \cdot \mathbf{p} = \mathbf{s} \cdot \mathbf{s} - (\mathbf{s} \cdot \hat{\mathbf{v}})^2 \tag{8}$$

$$= s^2\ \hat{\mathbf{v}} \cdot \hat{\mathbf{v}} - (\mathbf{s} \cdot \hat{\mathbf{v}})(\mathbf{s} \cdot \hat{\mathbf{v}}) \tag{9}$$

$$= s^2 \hat{v}_i \hat{v}_i - s_i \hat{v}_i\ s_j \hat{v}_j \tag{10}$$

$$= \hat{v}_i \hat{v}_j\ (s^2 \delta_{ij} - s_i s_j)\ . \tag{11}$$

Insert this into (7):

$$I(\hat{\mathbf{v}}) = \hat{v}_i \hat{v}_j \int_B d^3\mathbf{s}\ \rho(\mathbf{s})\ [s^2 \delta_{ij} - s_i s_j] \tag{12}$$

$$= \hat{v}_i \hat{v}_j M_{ij} \tag{13}$$

where M_{ij} is called the *inertia tensor*:

$$M_{ij} = \int_B d^3\mathbf{s}\ \rho(\mathbf{s})\ [s^2 \delta_{ij} - s_i s_j]\ . \tag{14}$$

Here are a couple of alternative ways of writing this expression. First, in vector notation, free of coordinates:

$$M = \int_B d^3\mathbf{s}\ \rho(\mathbf{s})[s^2\mathbf{I} - \mathbf{ss}]\ . \tag{14a}$$

where $\mathbf{I}$ is the *unit tensor*, which is just like the unit matrix: ones on the diagonal and zeros everywhere else, in any coordinate system. (Don't confuse the unit tensor $\mathbf{I}$ with the moment of inertia, I.) The notation $D = \mathbf{rs}$ defines a *dyad*; it is a second rank tensor (square matrix of coefficients) $D_{jk} = r_j s_k$. We won't use this notation much, however. Another way of writing (14) is to write out all the entries, which is sometimes helpful:

$$M = \int_B d^3\mathbf{s}\ \rho(\mathbf{s}) \begin{bmatrix} s_2^2 + s_3^2 & -s_1 s_2 & -s_1 s_3 \\ -s_2 s_1 & s_1^2 + s_3^2 & -s_2 s_3 \\ -s_3 s_1 & -s_3 s_2 & s_1^2 + s_2^2 \end{bmatrix}. \tag{14b}$$

And now for some useful properties of the inertia tensor. The inertia tensor is *symmetric*: this means $M_{ij} = M_{ji}$, which is obvious from (14). For future use we note the *trace*:

$$\mathrm{tr}\, M = M_{kk} = M_{11} + M_{22} + M_{33} \tag{15}$$

$$= \int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, [s^2 \delta_{kk} - s_k s_k] \tag{16}$$

$$= \int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, [3s^2 - s^2] \tag{17}$$

$$= 2 \int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, s^2 \,. \tag{18}$$

Thus

$$\int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, s^2 = ½ M_{kk} \,. \tag{19}$$

Obviously this number is independent of the orientation of the coordinate axes, though it does depend on the position of O in space. Another useful expression obtained from (14) and (19) is:

$$\int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, s_i s_j = ½ M_{kk} \delta_{ij} - M_{ij} \,. \tag{20}$$

These nine numbers are sometimes called the *products of inertia*.

2. Principal Axes

The behavior of I as $\hat{\mathbf{v}}$ varies can be understood more readily by imagining plotting values on the unit sphere, $S(1)$ (see next page). As a convenient shorthand, the points in the surface of a sphere of radius a will be called $S(a)$. Clearly the function I is fixed with respect to the body. Let us look for the stationary values of I as we move over $S(1)$. We seek extrema of $I(\hat{\mathbf{u}})$ subject to the condition that $\hat{\mathbf{u}}\cdot\hat{\mathbf{u}}=1$. To do this we introduce a *Lagrange multiplier* λ and define the unconstrained function

$$F(\hat{\mathbf{u}},\lambda) = I(\hat{\mathbf{u}}) - \lambda(\hat{\mathbf{u}}\cdot\hat{\mathbf{u}} - 1)\ . \tag{1}$$

The theory of Lagrange multipiers says that when F is stationary with respect to its arguments, the function I is stationary under the constraint. So we calculate the points where

$$\frac{\partial F}{\partial \lambda} = 0 \ \text{ and } \ \frac{\partial F}{\partial \hat{u}_k} = 0\ . \tag{2}$$

Obviously the first condition gives us merely that $\hat{\mathbf{u}}$ is unit:

$$\hat{\mathbf{u}}\cdot\hat{\mathbf{u}} - 1 \ = 0\ . \tag{3}$$

But the second is more interesting:

$$0 = \frac{\partial F}{\partial \hat{u}_k} = \frac{\partial}{\partial \hat{u}_k}\,[M_{ij}\hat{u}_i\hat{u}_j - \lambda(\hat{u}_i\hat{u}_i - 1)] \tag{4}$$

$$= M_{ij}\left[\frac{\partial \hat{u}_i}{\partial \hat{u}_k}\,\hat{u}_j + \hat{u}_i\,\frac{\partial \hat{u}_j}{\partial \hat{u}_k}\right] - \lambda\left[\frac{\partial \hat{u}_i}{\partial \hat{u}_k}\,\hat{u}_i + \hat{u}_i\,\frac{\partial \hat{u}_i}{\partial \hat{u}_k}\right]. \tag{5}$$

We can easily see that

$$\frac{\partial \hat{u}_i}{\partial \hat{u}_k} = \delta_{ik}\ . \tag{6}$$

So we find after tidying up (check this!):

$$M_{kj}\hat{u}_k = \lambda \hat{u}_j\ . \tag{7}$$

Since $\hat{\mathbf{u}}$ is not zero, (7) is a matrix eigenvalue problem. Because the matrix M_{ij} is symmetric, it must have real eigenvalues λ, each associated with an eigenvector. For the moment let us assume the problem is not *degenerate* which means the there are three distinct eigenvalues, call them $A<B<C$, with eigenvectors $\hat{\mathbf{u}}^A, \hat{\mathbf{u}}^B, \hat{\mathbf{u}}^C$. It is well known that for symmetric matrices the eigenvectors are then mutually orthogonal. Proof:

$$M_{ij}\hat{u}_j^A = A\hat{u}_i^A\ . \tag{8}$$

Dot both sides with $\hat{\mathbf{u}}^B$:

$$M_{ij}\hat{u}_i^B\hat{u}_j^A = A\hat{u}_i^A\hat{u}_i^B\ . \tag{9}$$

Because M_{ij} is symmetric and $\hat{\mathbf{u}}^B$ is an eigenvector

$$M_{ij}\hat{u}_i^B = M_{ji}\hat{u}_i^B = B\hat{u}_j^B\ . \tag{10}$$

Putting (10) in (9) and rearranging

$$(B - A)\,\hat{u}_i^A \hat{u}_i^B = 0 \tag{11}$$

and since $A \neq B$ by hypothesis, $\hat{u}_i^A \hat{u}_i^B = \hat{\mathbf{u}}^A \cdot \hat{\mathbf{u}}^B = 0$. So the three eigenvectors are mutually orthogonal and are called the *principal axes* of the body B when rotated about an axis through O. Obviously they are fixed with respect to the body. The maximum moment of inertia $I_{\max}$ is found about the axis $\hat{\mathbf{u}}^C$ and $I_{\max} = C$; if this isn't obvious just plug $\hat{\mathbf{u}}^C$ into (1.13). Similarly the minimum moment of inertia is A and is found about the axis $\hat{\mathbf{u}}^A$.

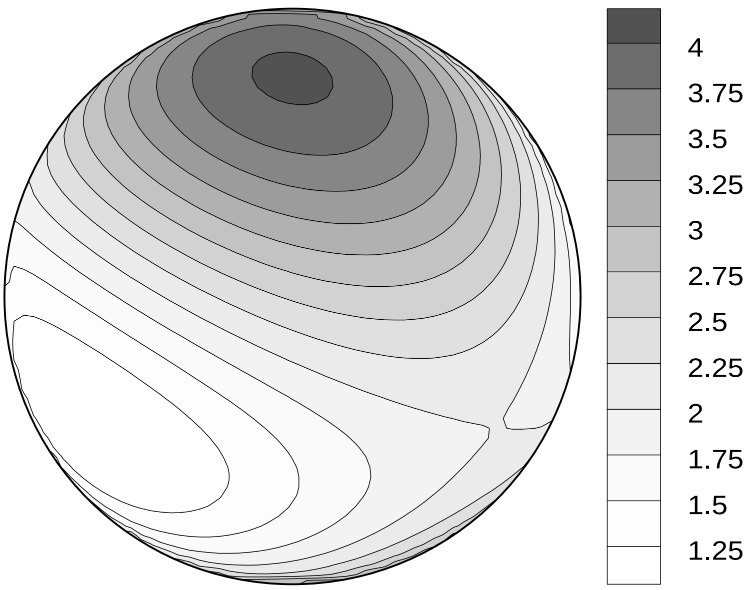
Suppose now that two of the eigenvalues are equal, say $A = B < C$, which approximates the case for the Earth. Then the system is degenerate, which means that any axis in the plane perpendicular to $\hat{\mathbf{u}}^C$ is an eigenvector with eigenvalue A. Of course one can choose two orthogonal directions in the plane to be the principal axes just as before.

It seems natural now to choose the coordinate axis system to make it coincide with the directions of the principal axes (the origin is still at O, however). Then (1.13) becomes very simple. Write the components of an arbitrary unit vector $\hat{\mathbf{v}}$ in the principal coordinate system as:

$$\hat{\mathbf{v}} = \hat{v}_1 \hat{\mathbf{u}}^A + \hat{v}_2 \hat{\mathbf{u}}^B + \hat{v}_3 \hat{\mathbf{u}}^C \ . \tag{12}$$

Substituting this into (1.13) and using the eigenvalue property we find:

$$I(\hat{\mathbf{v}}) = A\hat{v}_1^2 + B\hat{v}_2^2 + C\hat{v}_3^2 \ . \tag{13}$$



Another way of expressing this simplicity is to say that referred to the principal axis coordinate system the inertia tensor is diagonal; written out in full this means:

$$M = \begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{bmatrix} . \tag{14}$$

It is clear from this that $M_{kk} = A + B + C$, and because of (1.19) this is always true irrespective of the orientation of the axes.

We should briefly mention the importance of inertia tensor in rigid-body dynamics. Although not of an significance for this class, we note the following facts. The angular momentum about O of B moving with velocity $\mathbf{w}(\mathbf{s})$ is by definition:

$$\mathbf{L} = \int_B d^3\mathbf{s}\ \rho(\mathbf{s})\ \mathbf{s} \times \mathbf{w}(\mathbf{s}) \ . \tag{15}$$

This is true of a deformable or even fluid body. Now if B is rigid

$$\mathbf{w}(\mathbf{s}) = \mathbf{\Omega} \times \mathbf{s} + \alpha \mathbf{s} \tag{16}$$

where $\mathbf{\Omega}$ is the instantaneous *angular velocity* about O. Then we can show that setting (16) into (15) gives

$$L_i = M_{ij}\, \Omega_j \ . \tag{17}$$

Thus the angular momentum and spin vectors do not necessarily point in the same direction. If the Earth is approximated by a rigid body, and the angular momentum points in certain direction in inertial space (which will be invariant, to the extent there are no external torques acting), the spin vector moves around; this is the *Chandler wobble*, of which more later.

Exercises

2.1 Show that the eigenvalues of the inertia tensor are all positive.

2.2 Show rigorously (not by calculus) that C is the greatest value of $I(\hat{\mathbf{u}})$ and A is the least.

2.3 A rigid body has the following inertia tensor relative to a Cartesian axis system with origin on its center of mass (units are $\mathrm{kg\,m^2}$):

$$M_{ij} = \begin{bmatrix} 16.0 & -7.2 & -2.4 \\ -7.2 & 5.2 & -4.8 \\ -2.4 & -4.8 & 18.0 \end{bmatrix} .$$

Describe the body in as much detail as you can based on this information.

3. MacCullagh's Formula

Now we come to something to do with potential theory. We want to write down the first few terms in an expansion for the gravitational potential of an arbitrary body. This will give us an entry point to spherical harmonics and the result is interesting in its own right. Again we treat our friend B; the origin for vectors is at O, an arbitrary point not necessarily inside B. We know the potential of a point mass m observed at a distance d is $-Gm/d$, so we sum these contributions from all the points in B at an observer P situated at $\mathbf{r}$:

$$V(\mathbf{r}) = -G \int_B d^3\mathbf{s}\, \rho(\mathbf{s}) \frac{1}{|\mathbf{r}-\mathbf{s}|} \tag{1}$$

where $G = 6.67422 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$ is Newton's constant of gravitation.

A short digression: some workers in geophysics like $V = +Gm/d$; this has the disturbing effect that instead of being in a potential well when one is near a gravitating body, one is at a peak! Also notice this value for G is different from the one quoted in textbooks, and is more accurate, thought to be good to 14 parts in 10^6. See Schwarzschild, *Physics Today,* 53, p 21, 2000.

Back to the body B: we want to find an expansion for V valid when P is far from B and O, that is $r \gg s$. We shall expand $1/|\mathbf{r}-\mathbf{s}|$ in powers of s_i. Recall the Taylor expansion of an analytic function, written in subscript notation:

$$f(s_i) = f(0) + s_i\left(\frac{\partial f}{\partial s_i}\right)_0 + \frac{s_i s_j}{2!}\left(\frac{\partial^2 f}{\partial s_i \partial s_j}\right)_0 + \cdots \tag{2}$$

Let

$$f(s_i) = \frac{1}{|\mathbf{r}-\mathbf{s}|} = \frac{1}{R(s_i)}\ . \tag{3}$$

Then

$$\frac{\partial f}{\partial s_i} = \frac{\partial}{\partial s_i}\frac{1}{R} = -\frac{1}{R^2}\frac{\partial R}{\partial s_i}\ . \tag{4}$$

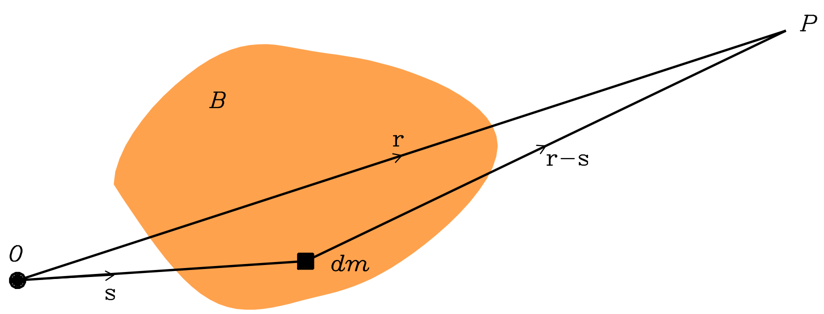


Figure 2

But by definition

$$R^2 = |\mathbf{r} - \mathbf{s}|^2 = r^2 + s^2 - 2\mathbf{r} \cdot \mathbf{s} \tag{5}$$

$$= r_i r_i + s_i s_i - 2 r_i s_i \ . \tag{6}$$

Differentiating (6):

$$2R \frac{\partial R}{\partial s_i} = 2s_i - 2r_i \tag{7}$$

or

$$\frac{\partial R}{\partial s_i} = \frac{s_i - r_i}{R} \ . \tag{8}$$

Now we substitute (8) into (4) and

$$\frac{\partial}{\partial s_i} \frac{1}{R} = \frac{r_i - s_i}{R^3} \ . \tag{9}$$

Looking back at (2) we need to evaluate (9) at $s_i = 0$, where $R = r$:

$$\left(\frac{\partial}{\partial s_i} \frac{1}{R} \right)_0 = \frac{r_i}{r^3} \ . \tag{10}$$

We spare the student the details and simply assert that in a similar way we can show:

$$\left(\frac{\partial^2}{\partial s_i \partial s_j} \frac{1}{R} \right)_0 = \frac{3 r_i r_j}{r^5} - \frac{\delta_{ij}}{r^3} \ . \tag{11}$$

We gather together (10), (11) and (3) into the formula for potential (39):

$$-\frac{V}{G} = \int_B d^3\mathbf{s}\, \rho(\mathbf{s}) \left[\frac{1}{r} + \frac{s_i r_i}{r^3} + \frac{s_i s_j}{2} \left(\frac{3 r_i r_j}{r^5} - \frac{\delta_{ij}}{r^3} \right) + \cdots \right] . \tag{12}$$

We rearrange the last term a bit to separate terms in s_j and r_j:

$$s_i s_j \left(\frac{3 r_i r_j}{r^5} - \frac{\delta_{ij}}{r^3} \right) = \frac{3 r_i r_j s_i s_j - r^2 \delta_{ij} s_i s_j}{r^5} \tag{13}$$

$$= \frac{3 r_i r_j s_i s_j - s^2 r^2}{r^5} \tag{14}$$

$$= \frac{r_i r_j}{r^5} (3 s_i s_j - s^2 \delta_{ij}) \ . \tag{15}$$

So, putting (15) into (12) and moving the integrals around:

$$-\frac{V(\mathbf{r})}{G} = \frac{1}{r} \int_B d^3\mathbf{s}\, \rho(\mathbf{s}) + \frac{r_i}{r^3} \int_B d^3\mathbf{s}\, \rho(\mathbf{s}) s_i + \frac{r_i r_j}{r^5} \int_B d^3\mathbf{s}\, \rho(\mathbf{s}) (\tfrac{3}{2}\, s_i s_j - \tfrac{1}{2} s^2 \delta_{ij}) + \cdots \tag{16}$$

$$= \frac{m_B}{r} + \frac{r_i\, c_i m_B}{r^3} + \frac{r_i r_j}{r^5} Q_{ij} + \cdots \tag{17}$$

where

$$m_B = \int_B d^3\mathbf{s}\,\rho(\mathbf{s})$$

$$c_i = \int_B d^3\mathbf{s}\,\rho(\mathbf{s})\,s_i/m_B \tag{17a}$$

$$Q_{ij} = ½\int_B d^3\mathbf{s}\,\rho(\mathbf{s})(3s_is_j - s^2\delta_{ij})$$

and m_B is the mass of B, $\mathbf{c}$ is a vector from O to the center of mass of B and Q_{ij} is called the *quadrupole tensor* of the mass distribution. In fact this is an expansion in inverse powers of r, which we can see more clearly if we write things in terms of the unit vector $\hat{\mathbf{r}}$:

$$-\frac{V}{G} = \frac{m_B}{r} + \frac{\hat{\mathbf{r}}\cdot\mathbf{c}m_B}{r^2} + \frac{\hat{r}_i\hat{r}_jQ_{ij}}{r^3} + \cdots \tag{18}$$

This equation gives the first three terms of a *multipole* expansion of the potential of the body B: the first term is the monopole, the second the dipole term, the third the quadrupole, etc in Greek.

To obtain the infamous MacCullagh's formula we restate the term in Q_{ij} in terms of the inertia tensor.

$$Q_{ij} = ½\int_B d^3\mathbf{s}\,\rho(\mathbf{s})(3s_is_j - s^2\delta_{ij})\ . \tag{19}$$

We now turn to (19) and (20) to compute the terms under the integral:

$$Q_{ij} = ½[3(½M_{kk}\delta_{ij} - M_{ij}) - (½M_{kk})\delta_{ij}] \tag{20}$$

$$= ½M_{kk}\delta_{ij} - \frac{3}{2}M_{ij}\ . \tag{21}$$

Thus

$$\hat{r}_i\hat{r}_jQ_{ij} = ½M_{kk} - \frac{3}{2}\hat{r}_i\hat{r}_jM_{ij} \tag{22}$$

$$= ½(A+B+C) - \frac{3}{2}I(\hat{\mathbf{r}})\ . \tag{23}$$

This is *MacCullagh's formula* in its general form:

$$-\frac{V}{G} = \frac{m_B}{r} + \frac{\mathbf{r}\cdot\mathbf{c}m_B}{r^3} + \frac{1}{r^3}[½(A+B+C) - \frac{3}{2}I(\hat{\mathbf{r}})] + \cdots \tag{24}$$

Exercises

3.1 Express the following vector formulas in terms of Einstein summation notation: $|\mathbf{r}|$, ∇f, $\nabla \cdot \mathbf{A}$, $\nabla^2 V$.

3.2 Rewrite the following expressions that employ Einstein notation in vector form: $3r_i s_i - s_k s_k$, $a_k c_k b_i - a_k b_k c_i$, $\partial_i \partial_j A_j$.
Here I have introduced the useful abbreviation $\partial_i = \partial/\partial x_i$.

3.3 A function V is said to be *harmonic* in a region if $\nabla^2 V = 0$ there; we can also say that V obeys Laplace's equation. Suppose V is harmonic in a region D, and $\mathbf{x}$ is a vector from a fixed origin O. Using the summation convention, or by some other means, calculate $\nabla^2(\mathbf{x} \cdot \nabla V)$.

4. Determination of the Moments of Inertia of the Earth

MacCullagh's formula, equation (3.24), can be written more simply if we make some obvious choices. First, choose O to be at the center of mass of B. This removes the dipole term from the expansion (3.24) because then $\mathbf{c}=0$. Second, let the axes of the coordinates be the principal axes of B. Then in the third term we use (2.12)

$$\hat{r}_i\hat{r}_j Q_{ij} = ½(A+B+C) - \frac{3}{2}(A\hat{r}_1^2 + B\hat{r}_2^2 + C\hat{r}_3^2) \tag{1}$$

Finally, let us look at a degenerate case with two equal eigenvalues: $A=B<C$ as it is fairly accurately for the Earth.

$$\hat{r}_i\hat{r}_j Q_{ij} = ½(2A+C) - \frac{3}{2}[A(\hat{r}_1^2 + \hat{r}_2^2) + C\hat{r}_3^2] \,. \tag{2}$$

$$= ½(C-A)(1-3\hat{r}_3^2) \tag{3}$$

because $|\hat{\mathbf{r}}|^2 = \hat{r}_1^2 + \hat{r}_2^2 + \hat{r}_3^2 = 1$. With all this in effect (3.24) becomes

$$-\frac{V}{G} = \frac{m_B}{r} + \frac{C-A}{r^3}\frac{1-3\hat{r}_3^2}{2} + \cdots \tag{4}$$

When we analyze the Earth's gravity field this is often written as follows for reasons soon to become apparent

$$-\frac{V}{G} = \frac{m_E}{r}\left[1 - J_2 P_2(\cos\theta)\frac{a^2}{r^2} + \cdots\right] \tag{5}$$

where m_E is the Earth's mass, a is its equatorial radius,

$$J_2 = \frac{C-A}{m_E a^2} \tag{6}$$

and $\cos\theta = \hat{r}_3 = \hat{\mathbf{u}}^C \cdot \hat{\mathbf{r}}$ where θ is the geocentric colatitude and P_2 is the *Legendre polynomial* of degree 2:

$$P_2(\mu) = ½(3\mu^2 - 1) \,. \tag{7}$$

Equation (5) is an example of a very simple spherical harmonic expansion of the potential V. The quantity J_2 is a dimensionless number, known from the analysis of satellite orbits to high accuracy for the Earth: $J_2 = .0010826265$. You should remember it is about 10^{-3}.

Now at last we are beginning to discuss the Earth. The constants of geodesy that I will use are all taken from one source: *Global Earth Physics: A Handbook of Physical Constants*, Editor: T. J. Ahrens, 1995. Chapters 1 and 2 are a rather dry compilation of the current best estimates of geodetic properties of the Earth and the other planets. Actually because geodesists and astronomers love to give as many significant figures as they can, values for various numbers are given to extravagant precision and don't always agree with each other in different parts of the book!

The accuracy of our measurements of J_2 means that we can actually see it varying in time! The expected direction is a decrease in J_2 as the Earth recovers from glacial loading in the polar regions — the Earth is becoming less oblate. But very recent work (C. M. Cox and B. F. Chao, *Science*, 297, pp 831-3, 2002) shows that the annual decline of about 3×10^{-11} suddenly reversed in about 1997. The reasons for this are completely mysterious.

Why should we be interested in these quantities? Originally the main interest came from the constraints they place on the radial density structure of the Earth. If we assume the main variation in density is with s the distance from the Earth's center (Yes, now O is at the center of mass!) (1.19) shows us that:

$$\frac{A+B+C}{8\pi} = \int_0^a \rho(s)\, s^4 \, ds \tag{8}$$

and of course we have

$$\frac{m_E}{4\pi} = \int_0^a \rho(s)\, s^2 \, ds \ . \tag{9}$$

Thus we have two integrals of the density profile to which we can fit models. For most planets in the solar system these are the only two constraints available! Another reason, which we come to later is that the main shape of the Earth and the moment of inertia are connected and simple models of the Earth's flattening due to rotation can predict J_2 (and they get it wrong).

First let's look at estimation of J_2, defined in (4.13). Here we give an oversimplified treatment; the real thing needs spherical harmonics and a lot of computer modeling. We look at the motion of an artificial satellite. If the Earth had the gravitational potential of only the first term in (3.31), and we ignore attractions of the sun, moon and other planets, the orbit of a near-Earth satellite would lie in a plane passing through the Earth's center of mass. But the equatorial bulges exert a torque on the satellite tending to pull the plane towards the equator, which has the effect of making the plane of the orbit precess in the opposite direction to that of the satellite, called *retrograde precession*. This motion is called *regression of the nodes*, the nodes being the points where the orbit intersects the equatorial plane. To a first approximation, the rate of precession depends on J_2. After a tedious calculation (see Section A in the Annex) in which one computes the average torque in one revolution due to the J_2 term, it can be shown that T_1, the period of precession is given by

$$\frac{T_1}{T_0} = \frac{2}{3}\frac{r^2}{a^2}\frac{1}{J_2 \cos\Theta} \tag{10}$$

where T_0 is the period of the satellite in its orbit, r is the mean radius of the orbit, Θ is the angle between the instantaneous plane of the orbit and $\hat{\mathbf{u}}^C$, the Earth's rotation axis. Clearly the fastest rate of precession is when $\Theta = 0$ (when the plane of the orbit lies in the Earth's equatorial plane and so the precession becomes invisible); since $J_2 \approx 0.001$ we can compute that $T_1 \geq$ 38 days.

Obviously having J_2 is not enough to determine C and A separately (We assume that to a good enough approximation $A = B$). The second piece of information comes from the rate of precession of the Earth's rotation axis because of tidal torques from the sun and moon acting on the equatorial bulges. The precession makes itself known by the movement of the fixed point in the night sky relative to the stars. It is called the *precession of the equinoxes* because the time of the equinox (when day and night have equal length) occurs when the spin axis of the Earth is orthogonal to the Earth-sun line. This precession is also retrograde in sense, that is, opposite to the spin direction of the Earth. The period of the precession is very long, 25,730 years, or 0.014 degrees per year, yet the ancient Greeks had noticed the motion! The period is calculated in a manner very similar to that of the satellite orbit precession; again it is retrograde, meaning in the opposite direction from the Earth's rotation.

Whereas the obliquity of the Earth is believed to be constant at about 23° for most of geologic time, the torques of the sun may have caused Mars to undergo big excursions in obliquity; see Bills, B. G., *JGR*, 95, pp 14137-53, 1990.

For the Earth we find (omitting the details):

$$T_2 = \frac{2\pi}{\Omega_M + \Omega_S} \tag{11}$$

where Ω_M and Ω_S are contributions from the moon and sun:

$$\Omega_M = \frac{3Gm_M \cos\Theta}{2\omega R_M^3} \, \frac{C - A}{C} \tag{12}$$

$$\Omega_S = \frac{3Gm_S \cos\Theta}{2\omega R_S^3} \, \frac{C - A}{C} \; . \tag{13}$$

Here m_M, m_S are the masses of the two bodies, R_M, R_S their distances, ω is the radian frequency of the Earth's rotation (2π/ sidereal day) and Θ is the Earth's obliquity, the tilt of the spin axis from the normal to the ecliptic plane. Since we can estimate all these quantities (see Exercise 4.4) and T_2 as well, we find

$$\frac{C - A}{C} = 0.0032738 \; . \tag{14}$$

This ratio is often called H in the geodetic literature. Combine this with

$$J_2 = \frac{C - A}{m_E a^2} = 0.0010826 \tag{15}$$

and we find

$$\frac{C}{m_E a^2} = 0.33070 \quad \text{and} \quad \frac{A}{m_E a^2} = 0.32961 \; . \tag{16}$$

Also for use in (1.26) we see the trace of the inertia tensor

$$\frac{M_{kk}}{2m_E a^2} = \frac{A + B + C}{2m_E a^2} = 0.49496 \simeq \frac{2A + C}{2m_E a^2} \; . \tag{17}$$

Before leaving this subject we might ask if there is some physical insight available to help see why the regression of the nodes gives $(C - A)/m_E a^2$ while the precession of the equinoxes gives $(C - A)/C$. The term giving the couple that drives the precession is proportional to $C - A$ in both cases, a gravitational force acting on, or due to, the equatorial bulges. But the angular momentum of the satellite is $ma^2\omega = Gm_E m/a\omega$ from Kepler, while it is $C\omega$ for the Earth. Since the rate of precession is the ratio of the couple to the angular momentum, this leads to the term in

$(C-A)/m_E a^2$ for the satellite and $(C-A)/C$ for the Earth.

Exercises

4.1 It is well known that gravity g on the Earth's surface is stronger at the poles than at the equator because of rotation. If the Earth stopped rotating, but retained its present shape and interior structure, would this still be true? Use MacCullagh to find out.

4.2 Calculate the values of the quantities in (8)-(10) for a uniform sphere and for an ellipsoid of uniform density with the same flattening as the Earth, $f=(a-c)/a$ $=0.00336$. What do you conclude from the comparison with the observed numbers from the Earth?

4.3 Assume the Earth has a uniform density mantle and a uniform density core; what densities must be taken to satisfy the estimates of (10) and the mass of the Earth? Despite a certain inconsistency, treat the density as if it were a function of s alone, as in (1) and (2). Compare your values with estimates of averages for current Earth models.

4.4 For equation (2) we need m_E. Explain what needs to be measured to find m_E. For (6) show how the quantity Gm_S/R_S^3 can be computed from a knowledge of the length of the year. But the quantity in (5) Gm_M/R_m^3 requires more information than the length of the lunar month; what other knowledge is needed and how is it determined? Which is bigger: Ω_S or Ω_M, and by how much?

4.5 In geodetic and geophysical calculations the product $G\,m_E=3.98600434\times10^{14}$ $\mathrm{m}^3/\mathrm{s}^2$ arises often. The value of this product is known to very high accuracy, more than 9 significant figures, even though $G=6.6742\times10^{-11}$ $\mathrm{N\,m^2/kg^2}$, Newton's gravitational constant, has been determined to an accuracy of at most five figures. Explain why this is, describing what kinds of measurements are employed to find the best values of $G\,m_E$, G and m_E.

5. The Geoid - Clairaut's Formula

The shape of the Earth is obviously governed by its gravity field to a large extent and indeed, geodesy has come to *define* the shape in terms of the gravitational potential. But first we must account for the effects of rotation.

The vector acceleration of gravity as experienced by an observer rotating with the Earth can be written

$$\mathbf{g} = -\nabla V + \omega^2 \mathbf{p} \tag{1}$$

$$= -\nabla V + \omega^2 (\mathbf{r} - \hat{\mathbf{z}}\hat{\mathbf{z}} \cdot \mathbf{r}) \tag{2}$$

where the term in ω^2 is the apparent centrifugal force. If we let

$$\mathbf{g} = -\nabla U \tag{3}$$

then we find

$$U = V - ½\omega^2 (r^2 - (\hat{\mathbf{z}} \cdot \mathbf{r})^2) \tag{4}$$

$$= V - ½\omega^2 r^2 \sin^2 \theta \tag{5}$$

where we call U the *geopotential*. Consider surfaces of constant U; since the vector $\mathbf{g}$ is normal to any such surface, it is the local horizontal and, apart from the effects of wind and currents, the ocean defines one such surface. The geopotential best coinciding with the ocean surface on average is called the *geoid*. Heights on the continent referring to sea level are distances vertically to the local geoid. We can compute the shape of the geoid to a rough approximation from MacCullagh's formula.

Assuming an axisymmetric Earth, we use the axisymmetric form of MacCullagh's formula, equation (4.5) to write an expression for the geopotential at polar coordinates r and θ:

$$U(r, \theta) = -\frac{Gm_E}{r} + Gm_E J_2 P_2(\cos\theta) \frac{a^2}{r^3} - ½\omega^2 r^2 \sin^2 \theta \ . \tag{6}$$

We seek the function $r_0(\theta)$ such that $U(r_0(\theta)) = U_0 =$ constant. Then, from (6) we see

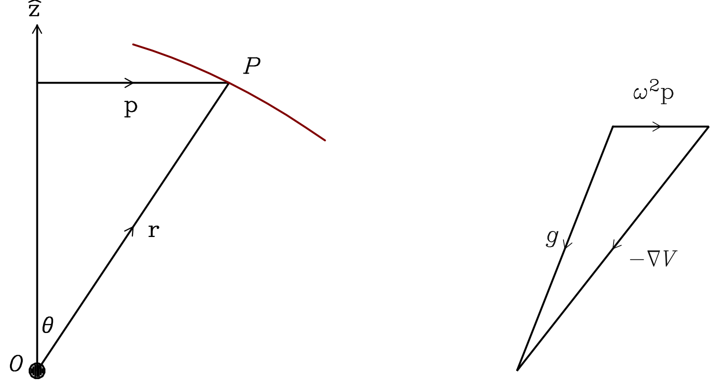


Figure 3

$$-\frac{aU_0}{Gm_E} = \frac{a}{r_0} - J_2\frac{a^3}{r_0^3}(\tfrac{3}{2}\cos^2\theta - ½) + ½\frac{\omega^2 a^3}{Gm_E}\frac{r_0^2}{a^2}\sin^2\theta\ . \tag{7}$$

We know for the geoid that a/r_0 must be nearly unity; but $J_2 \approx 10^{-3}$ and $n = \omega^2 a^3/Gm_E \approx \omega^2 a/g \approx 3\times 10^{-3}$, both small enough that we can neglect their squares compared with unity. So we write

$$\frac{r_0(\theta)}{a} = 1 - \varepsilon(\theta) \tag{8}$$

where $\varepsilon \ll 1$ and substitute this into (7) and use the first term in the binomial expansion where required:

$$-\frac{aU_0}{Gm_E} = 1 + \varepsilon - J_2(1+3\varepsilon)(\tfrac{3}{2}\cos^2\theta - ½) + ½\frac{\omega^2 a^3}{Gm_E}(1-2\varepsilon)\sin^2\theta\ . \tag{9}$$

Rearranging and dropping such products as εJ_2 we get:

$$\varepsilon(\theta) = -1 - \frac{aU_0}{Gm_E} + J_2(\tfrac{3}{2}\cos^2\theta - ½) - ½\frac{\omega^2 a^3}{Gm_E}\sin^2\theta\ . \tag{10}$$

At the equator where $\theta = \pi/2$, $r = a$, the equatorial radius; equivalently $\varepsilon(\pi/2) = 0$. Adjust U_0 in (9) to make this so:

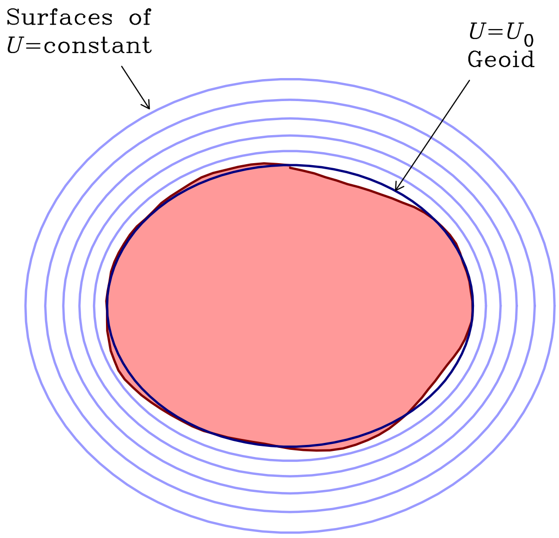
$$-\frac{aU_0}{Gm_E} = 1 + ½J_2 + ½\frac{\omega^2 a^3}{Gm_E} \tag{11}$$

and then (10) simplifies to

$$\varepsilon(\theta) = \left(\frac{3}{2}J_2 + ½\frac{\omega^2 a^3}{Gm_E}\right)\cos^2\theta \tag{12}$$

or from (8)

$$r_0(\theta) = a\,(1 - f\cos^2\theta),\ \text{where } f = \frac{3}{2}J_2 + ½\frac{\omega^2 a^3}{Gm_E}\ . \tag{13}$$



When $\theta = 0$ we are at the pole and $r_0 = c$, the polar radius. Thus $f = (a-c)/a$, the *flattening* of the geoid. Using the numbers for J_2 and other quantities we compute $f = 1/297.8$ compared with the actual observed value of 298.256; the difference, 1.5 parts in 10^3, is what we expect from dropping the higher order terms in J_2 and n.

Fig 3a

6. Some Potential Theory

Consider the potential of a point mass at the origin:

$$V = -\frac{Gm}{r} \ . \tag{1}$$

For any point not at the origin we can compute $\nabla^2 V$:

$$-\frac{1}{Gm}\nabla^2 V = \frac{\partial^2}{\partial r_i \partial r_i}\frac{1}{r} \ . \tag{2}$$

From (3.11) we see

$$\frac{\partial^2}{\partial r_i \partial r_i}\frac{1}{r} = \frac{3r_i r_i}{r^5} - \frac{\delta_{ii}}{r^3} \tag{3}$$

$$= \frac{3r^2}{r^5} - \frac{3}{r^3} = 0 \ . \tag{4}$$

So the potential of a point mass obeys *Laplace's equation*:

$$\nabla^2 V = 0 \tag{5}$$

provided that $r \neq 0$. Functions obeying (5) in a region are said to be *harmonic* functions. The potential obtained by summing all the contributions from the points of an extended body like B is also therefore harmonic in the region not containing any of the mass.

There is a great deal of theory concerning harmonic functions; for example, if V is harmonic in a bounded region $D \subset \mathbb{R}^3$, then V is infinitely differentiable at all points strictly within D; V never attains a maximum or minimum value inside D – these are always on the boundaries of D. Or this: consider a spherical surface within D; the average value over the sphere equals the value at the center of the sphere. A result we might find useful is the uniqueness theorem for exterior Neumann boundary value problems: suppose B is a smooth compact body with an outward normal at each point of ∂B, its surface. Suppose further that the value of the normal derivative of a harmonic function V is known everywhere on ∂B and that V tends to zero like $1/r$ at infinity. Then knowledge of $\partial V/\partial n$ on ∂B is sufficient to define V uniquely everywhere outside B. An aside on notation: the quantity $\partial V/\partial n$ is shorthand for $\hat{\mathbf{n}} \cdot \nabla V$, the derivative projected onto the normal $\hat{\mathbf{n}}$, and is known as the *normal derivative*. On the geoid the normal derivative of U is $-|\nabla U| = -g$; this should be obvious to you.

The uniqueness theorem appears to be useful because it says that we need only measure g everywhere on an equipotential surface and that's enough to determine the potential and g everywhere above the Earth. The reason for the qualifications is that we measure the geopotential U, not V, and a short calculation shows that

$$\nabla^2 U = -2\omega^2 \tag{6}$$

which is a version of *Poisson's equation*:

$$\nabla^2 V = 4\pi G\rho \tag{7}$$

for a constant but negative density distribution. However, we shall prove uniqueness for (6) also. Now we suppose the normal gradient of U is measured on ∂B and that at infinity solutions behave like

$$-\tfrac{1}{2}\omega^2 r^2 \sin^2\theta + O(r^{-1}) \,. \tag{8}$$

We assume a solution U_1 exists, satisfying (6) and the $\partial U_1/\partial n = f$ on ∂B. Now, suppose there is a second solution, U_2, different from U_1, but obeying the same boundary conditions and (6): we shall show this leads to a contradiction. Consider the function $W = U_1 - U_2$. Clearly

$$\nabla^2 W = \nabla^2 U_1 - \nabla^2 U_2 = 0 \tag{9}$$

so W is harmonic. Also $\partial W/\partial n = 0$ on ∂B and at infinity $W = O(r^{-1})$. Let us compute

$$J = \int_{\bar{B}} d^3\mathbf{s}\; |\nabla W|^2 \tag{10}$$

where $\bar{B}$ means all of space outside B. Recall the vector identity number 4 (it's really useful to remember this one!):

$$\nabla\cdot(f\,\nabla g) = \nabla f\cdot\nabla g + f\nabla\cdot\nabla g \,. \tag{11}$$

With $f = g = W$ we conclude

$$\nabla W\cdot\nabla W = \nabla\cdot(W\,\nabla W) - W\nabla^2 W \,. \tag{12}$$

But the since W is harmonic, from (9), the second term vanishes and so

$$J = \int_{\bar{B}} d^3\mathbf{s}\; \nabla\cdot(W\,\nabla W) \,. \tag{13}$$

We now apply Gauss's Divergence Theorem, but to be careful we do so inside a bounded region, a large sphere radius R:

$$J = \int_{S(R)} d^2\mathbf{s}\;\, W\,\nabla W\cdot\hat{\mathbf{r}} - \int_{\partial B} d^2\mathbf{s}\;\, W\,\nabla W\cdot\hat{\mathbf{n}} \tag{14}$$

where $\hat{\mathbf{n}}$ is the outward normal on ∂B. From the definition of the problem

$$\nabla W\cdot\hat{\mathbf{n}} = \frac{\partial W}{\partial n} = 0 \tag{15}$$

so the second term vanishes. As R tends to infinity, $W\,\nabla W\cdot\hat{\mathbf{r}} = O(R^{-3})$ while the surface of $S(R)$ increases like R^2; so we conclude the first term also vanishes over the whole space. Thus J vanishes, which means from (10) that W is a constant, which must be zero from the behavior at infinity. Thus, contrary to assumption, U_1 and U_2 are not different – there is only one solution to (6) obeying the stated boundary conditions.

In fact large parts of the geoid are inaccessible since most continental land is above sea level, so that strictly we cannot measure g on the geoid. Can the theorem be strengthened? Yes, Backus (*Quarterly Journal of Mechanics and Applied Mathematics,* 21, pp 195-221, 1968) showed that if $|\mathbf{g}|$ is measured on ∂B the solution is unique for V within a sign. Strangely, and quite importantly, the result is *not* true for magnetic fields because the proof depends on $\mathbf{g} \cdot \hat{\mathbf{n}}$ not changing sign anywhere on ∂B, which holds for gravity, but not for magnetic fields.

7. Spherical Harmonics

Now we come to some of the most ubiquitous functions in geophysics, used in gravity, geomagnetism and seismology. There are several ways to look at spherical harmonics and there is a confusing use of them to mean two different, but related things. Specifically, spherical harmonics are functions defined on the unit sphere, which are the equivalent to Fourier series for the circle. But these functions can also be used to build solutions to Laplace's equation and those solutions are called spherical harmonics too. It is helpful sometimes to refer to the functions on the sphere as surface harmonics if a distinction must be made.

We shall treat spherical harmonics as eigensolutions of the surface Laplacian. This would be like developing Fourier series as eigensolutions of the operator $(d/dx)^2$ on a finite line, but with boundary conditions that y and dy/dx match at the two ends. We sometimes get some mileage from representing a thing in two ways, one within a fixed coordinate system, the other in coordinate-free form. First we need a spherical polar coordinate system: see the figure. The origin O is always fixed to be the center of the unit sphere, and all coordinates are referred to that origin. Let us define a *surface differential operator* for the sphere in two ways:

$$\nabla_1 = \hat{\theta}\,\frac{\partial}{\partial\theta} + \frac{\hat{\phi}}{\sin\theta}\,\frac{\partial}{\partial\phi} \tag{1}$$

$$= r\nabla - \mathbf{r}\,\frac{\partial}{\partial r}\ . \tag{2}$$

The subscript one is to remind us the operator acts over the unit sphere, $S(1)$. The first definition shows how to compute the surface gradient in a spherical polar coordinate system; the second assumes the function is defined in all of space and just subtracts out the radial part. The second definition shows the operator is independent of coordinate orientation and also that nothing funny happens at the poles. The ordinary Laplacian operator in $\mathbb{R}^3$ is

$$\nabla^2 = \frac{\partial}{\partial x_i}\,\frac{\partial}{\partial x_i} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{3}$$

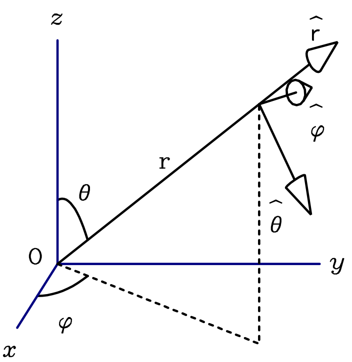


Figure 4

$$= \frac{1}{r}\frac{\partial^2}{\partial r^2}\,r + \frac{1}{r^2}\,\nabla_1^2 \tag{4}$$

where ∇_1^2 is the *surface Laplacian*, sometimes also called the *Beltrami operator*; relative to a coordinate system

$$\nabla_1^2 = \frac{\partial^2}{\partial\theta^2} + \cot\theta\,\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\,\frac{\partial^2}{\partial\phi^2} \;=\; \frac{1}{\sin\theta}\,\frac{\partial}{\partial\theta}\,\sin\theta\,\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\,\frac{\partial^2}{\partial\phi^2}\;. \tag{5}$$

We can think of ∇_1^2 as the ordinary Laplacian, with the radial part subtracted and scaled by r^2 to make it unitless: from (4)

$$\nabla_1^2 = r^2\nabla^2 - r\,\frac{\partial^2}{\partial r^2}\,r\;. \tag{6}$$

As with (2), this equation shows the operator is independent of any coordinate system and the singularity at the poles is not intrinsic to ∇_1^2, but is an artifact of the coordinates. We remark there is another surface operator $\nabla_s = r^{-1}\nabla_1$; this one has dimensions of $1/L$ like d/dx or the regular gradient operator ∇.

To develop spherical harmonics we ask for the eigenvalues and eigenfunctions of the surface Laplacian. We do this in part because, just as in $\mathbb{R}^3$ the eigenvectors of a symmetric matrix provide an orthogonal basis for the space, so here the self adjoint operator has a collection of orthogonal functions that span the function space; see Annex notes B. It is helpful to introduce a bit of modern notation. We regard complex-valued functions on the unit sphere as elements in the Hilbert space $L_2(S(1))$, which means they are functions that all obey

$$\| f \| = \sqrt{\int_{S(1)} d^2\hat{\mathbf{s}}\, | f(\hat{\mathbf{s}}) |^2} < \infty \tag{7}$$

where the unusual notation $d^2\hat{\mathbf{s}}$ is a surface element on $S(1)$, which would be $\sin\theta\, d\theta\, d\phi$ in a particular polar coordinate system. We find it handy to work complex functions on $S(1)$.

As you will know $\| f \|$ is the *norm* of the function f, and measures its size. The space L_2 comes with an *inner product* given by

$$(f, g) = \int_{S(1)} d^2\hat{\mathbf{s}}\; f(\hat{\mathbf{s}})\, g(\hat{\mathbf{s}})^* \tag{8}$$

which plays the role of the dot product between ordinary vectors. Notice that $\| f \| = (f, f)^{1/2}$. Most important is the idea of *orthogonality*: two functions are orthogonal if their inner product vanishes

$$(f, g) = 0 \tag{9}$$

just as two vectors are at right angles when their dot product is zero. With this notation in hand we can define a *self adjoint* operator. Such an operator O satisfies

$$(Of, g) = (f, Og) \tag{10}$$

for every pair of elements for which $Of, Og \in L_2$. It is shown in the Annex, Section B, that the surface Laplacian ∇_1^2 is self adjoint. Furthermore, it is shown there that the eigenfunctions of self adjoint operators are orthogonal.

Suppose $u = u(\hat{\mathbf{r}}) = u(\theta, \phi)$ satisfies

$$\nabla_1^2 u = \lambda\, u \tag{11}$$

and u is continuous everywhere on $S(1)$ but is not identically zero, then u is an eigenfunction for ∇_1^2. Not every value of λ can support such solutions; indeed, only when λ is one of the eigenvalues

$$\lambda = 0, -2, -6, -12, -20, \cdots - l(l+1) \cdots \tag{12}$$

does (11) have nontrivial solutions. How does one prove this? The traditional way is by a technique called *separation of variables*, assuming that u can be written as a product of two single-argument functions: $u(\theta, \phi) = \Theta(\theta)\,\Phi(\phi)$, substituting in and getting two one-dimensional eigenvalue problems, one each for Θ and Φ. See Morse and Feshbach, *Methods of Mathematical Physics*, Vol II, 1953, for example. Chapter 3 of Backus, Parker and Constable (*Foundations of Geomagnetism*, 1996) does this entirely differently, by looking at homogeneous harmonic polynomials.

We call l the *degree* of the spherical harmonic. The eigenfunctions of ∇_1^2 associated with the eigenvalues are called *spherical harmonics*; we write them

$$u(\theta, \phi) = Y_l^m(\theta, \phi)\,. \tag{13}$$

We will give an explicit formula for these functions later; they are complex-valued on the sphere. The eigenvalues in (12) are not simple (except $\lambda = 0$), but $2l+1$-fold degenerate: thus associated with $\lambda = -6 = -2\times(2+1)$, say, there are $5 = 2\times 2+1$ different (that is, linearly independent) eigenfunctions. This is where the index m comes in: for each degree l the *order* m can be any one of $-l, -l+1, \cdots 0, 1, \cdots l$, giving the required number of different functions for any l. These are arranged to be mutually orthogonal, just as we choose mutually perpendicular principal axes in the case of an inertia tensor with two equal eigenvalues. Thus the spherical harmonics form an orthogonal family:

$$(Y_l^m, Y_n^k) = \int_{S(1)} d^2\hat{\mathbf{s}}\; Y_l^m(\hat{\mathbf{s}})\, Y_n^k(\hat{\mathbf{s}})^* = 0, \quad m \neq k \text{ or } l \neq n\,. \tag{14}$$

We can scale the spherical harmonics to be of unit norm:

$$\| Y_l^m \| = 1 \tag{15}$$

then the spherical harmonics are said to be *fully normalized*, although not everyone does this. With fully normalized harmonics (14) and (15) combine to give

$$(Y_l^m, Y_n^k) = \delta_{ln}\, \delta_{mk} \tag{16}$$

Notice that any linear combination of eigenfunctions of degree l is also an eigenfunction with eigenvalue $-l(l+1)$.

It is time to write out an explicit form for Y_l^m. These solutions are the ones obtained by the separation of variables mentioned earlier – they are each a product of a function of θ (colatitude) and one of ϕ (longitude). Here we go:

$$Y_l^m(\theta, \phi) = N_{lm}\, e^{im\phi}\, P_l^m(\cos\theta) \tag{17}$$

where N_{lm} is a normalization constant to adjust the size of the functions; as I mentioned earlier, I usually choose it to enforce (15); then

$$N_{lm} = (-1)^m \left(\frac{2l+1}{4\pi}\right)^{1/2} \left(\frac{(l-m)!}{(l+m)!}\right)^{1/2} . \tag{18}$$

There are however several different conventions regarding N_{lm}. For example, leaving off the alternating sign and removing the factor $((2l+1)/4\pi)^{1/2}$, results in functions that are *Schmidt normalized*. Our convention, (18), is most convenient for theoretical work, because of (15), which the others fail to comply with. The factor $\exp(im\phi)$ is self explanatory; it contains the complex behavior and is obviously pretty simple. The last factor is called an *Associated Legendre function* and is defined by

$$P_l^m(\mu) = \frac{1}{2^l l!} (1-\mu^2)^{m/2} \frac{\partial^{l+m}}{\partial \mu^{l+m}} (\mu^2 - 1)^l . \tag{19}$$

When the order $m=0$, the Associated Legendre function becomes a polynomial in μ and instead being written $P_l^0(\mu)$ it is designated $P_l(\mu)$, the *Legendre polynomial*. But it is still possible to get it from (19). We provide a very short table. Here $s = \sin\theta = (1-\mu^2)^{1/2}$.

$$P_0(\mu) = 1$$

$$P_1(\mu) = \mu \qquad P_1^1(\mu) = s$$

$$P_2(\mu) = (3\mu^2 - 1)/2 \qquad P_2^1(\mu) = 3\mu s \qquad P_2^2(\mu) = 3s^2$$

$$P_3(\mu) = \mu(5\mu^2 - 3)/2$$

$$P_3^1(\mu) = 3s(5\mu^2 - 1)/2$$

$$P_3^2(\mu) = 15s^2\mu \qquad P_3^3(\mu) = 15s^3$$

$$P_4(\mu) = (35\mu^4 - 30\mu^2 + 3)/8$$

$$P_4^1(\mu) = 5s\mu(7\mu^2 - 3)/2$$

$$P_4^2(\mu) = 15s^2(7\mu^2 - 1)/2$$

$$P_4^3(\mu) = 105s^3\mu \qquad P_4^4(\mu) = 105s^4$$

$$P_5(\mu) = \mu(63\mu^4 - 70\mu^2 + 15)/8$$

$$P_5^1(\mu) = 15s(21\mu^4 - 14\mu^2 + 1)/8$$

$$P_5^2(\mu) = 105s^2\mu(3\mu^2 - 1)/2$$

$$P_5^3(\mu) = 105s^3(9\mu^2 - 1)/2$$

$$P_5^4(\mu) = 945 s^4 \mu \qquad P_5^5(\mu) = 945 s^5 .$$

We have listed only positive m since there is a nice symmetry that allows us to do without:

$$Y_l^{-m} = (-1)^m (Y_l^m)^* \; . \tag{20}$$

We also note these special values – they are worth remembering:

$$P_l(1) = 1, \quad P_l^m(1) = 0, \; m \neq 0, \quad P_l^l(\cos\theta) = c_l \sin^l \theta \; . \tag{21}$$

Next we come to one of the most important properties of the spherical harmonics: they are a *complete* set for expanding a wide class of functions on the unit sphere. This means for example that an arbitrary continuously differentiable function $f(\theta, \phi)$ can always be written as an infinite converging series like this:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} \, Y_l^m(\theta, \phi) \; . \tag{22}$$

This is where the analog to the Fourier series emerges. To get the expansion coefficients we just take the inner product of f against the basis function:

$$c_{lm} = (f, Y_l^m) = \int_{S(1)} d^2\hat{\mathbf{s}} \; f(\hat{\mathbf{s}}) \, Y_l^m(\hat{\mathbf{s}})^* \; . \tag{23}$$

It is this property that makes spherical harmonics so useful. Orthogonality is a property that follows from the self-adjointness of ∇_1^2. Completeness follows from a more subtle property, that the inverse operator of ∇_1^2 is compact, a property that would take us too far afield to explore.

Two powerful properties of the spherical harmonics that do not follow so immediately from their connection with ∇_1^2 are the *Addition Theorem* and the *generating function* series for Legendre polynomials. We state the second one first:

$$\frac{1}{(1 - 2\mu x + x^2)^{1/2}} = \sum_{l=0}^{\infty} x^l \; P_l(\mu) \; . \tag{24}$$

We can prove this later from the development of the potential of a point mass.

The Spherical Harmonic Addition Theorem says

$$P_l(\hat{\mathbf{u}} \cdot \hat{\mathbf{v}}) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_l^m(\hat{\mathbf{u}}) \, Y_l^m(\hat{\mathbf{v}})^* \; . \tag{25}$$

The following proof won't be found in any of the books; for another see *Foundations*, Chapter 3. Consider a delta function on $S(1)$ which peaks at the north pole; we would like its surface harmonic series in the form of (22). So deduce from (23) that

$$c_{lm} = \int_{S(1)} d^2\hat{\mathbf{s}} \; \delta(\hat{\mathbf{s}} - \hat{\mathbf{z}}) \, Y_l^m(\hat{\mathbf{s}})^* \tag{26}$$

$$= Y_l^m(\hat{\mathbf{z}})^* \; . \tag{27}$$

It follows from (21) that all the Y_l^m vanish at the pole ($\theta = 0$) except those of order zero, that is $m = 0$; applying (21) and (18) we calculate

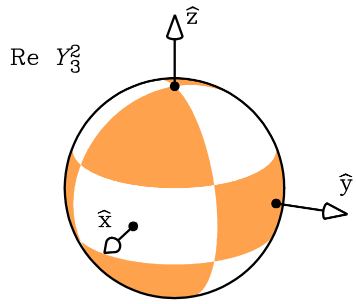


Figure 5

$$c_{l0} = \left(\frac{2l+1}{4\pi}\right)^{1/2} \tag{28}$$

and so, using (17) and (18) again

$$\delta(\hat{\mathbf{s}} - \hat{\mathbf{z}}) = \sum_{l=0}^{\infty} \left(\frac{2l+1}{4\pi}\right)^{1/2} Y_l^0(\hat{\mathbf{s}}) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} P_l(\cos\theta) \tag{29}$$

$$= \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} P_l(\hat{\mathbf{z}} \cdot \hat{\mathbf{s}}) \, . \tag{30}$$

Next consider the delta function peak to be moved to some other point on the unit sphere, say $\hat{\mathbf{u}}$; then because (30) does not depend on the fact that $\hat{\mathbf{z}}$ is a coordinate axis, we can replace $\hat{\mathbf{z}}$ by $\hat{\mathbf{u}}$ and the equation is still true:

$$\delta(\hat{\mathbf{s}} - \hat{\mathbf{u}}) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} P_l(\hat{\mathbf{u}} \cdot \hat{\mathbf{s}}) \, . \tag{31}$$

With the delta function in the new position, make a spherical harmonic expansion (22) again. Following the same steps, we quickly arrive at:

$$\delta(\hat{\mathbf{s}} - \hat{\mathbf{u}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m(\hat{\mathbf{s}}) \, Y_l^m(\hat{\mathbf{u}})^* \, . \tag{32}$$

Now we compare these two expressions degree by degree. Since these are both expansions in the eigenfunctions of ∇_1^2 and the eigenfunctions of different degrees are orthogonal to each other, the degree-l terms in each sum must be equal to each other: hence (25).

Unless you have seen them before, you probably have no idea what the spherical harmonics look like at this point. One further property of the Associated Legendre functions helps: for $-1 < \mu < 1$ the function $P_l^m(\mu)$ crosses zero exactly $l - m$ times. With this information we can picture the Y_l^m on a sphere, by graphing the places where $\mathrm{Re}\, Y_l^m$ is zero; $\mathrm{Im}\, Y_l^m$ has the same pattern, but rotated about the $\hat{\mathbf{z}}$ axis, as can be seen at once from (17). So here is the recipe: for Y_l^m there are two sets of lines on the sphere where $\mathrm{Re}\, Y_l^m$ vanishes: (1) a set of $2m$ equally-spaced halves of great circles through the poles (meridians) coming from the exponential; (2) a set of $l - m$ small circles with planes normal to the z coordinate axis. After we have drawn a few we get the idea that the higher degree harmonics are shorter wavelength than the lower ones. In fact one can consider the surface harmonics to be standing waves in the surface of the sphere. It can be shown that the wavelength of the waves in degree l is $2\pi/(l+½)$ independently of m and the position on the sphere; in seismological circles this is called *Jean's formula*. In analogy with Fourier filtering, people often make a spherical harmonic expansion, then remove the high-degree terms in order to emphasize the long-wavelength features; or conversely, the low degree-terms are removed to show the high-wavenumber behavior, of the geoid, for example.

One very useful expansion is obtained by putting the generating function and the Addition Theorem together, to find an expansion for the potential of point mass, not at the origin. With $s<r$ we write

$$\frac{1}{R} = \frac{1}{|\mathbf{r}-\mathbf{s}|} = \frac{1}{\sqrt{r^2+s^2-2rs\,\hat{\mathbf{r}}\cdot\hat{\mathbf{s}}}} \tag{33}$$

$$= \frac{1}{r}\,\frac{1}{[1+(s/r)^2-2(s/r)\,\hat{\mathbf{r}}\cdot\hat{\mathbf{s}}]^{½}} \,. \tag{34}$$

Now we recognize the generating function:

$$\frac{1}{R} = \frac{1}{r}\sum_{l=0}^{\infty}\left(\frac{s}{r}\right)^l P_l(\hat{\mathbf{r}}\cdot\hat{\mathbf{s}}) \tag{35}$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\,\frac{s^l}{r^{l+1}}\,Y_l^m(\hat{\mathbf{r}})\,Y_l^m(\hat{\mathbf{s}})^* \,. \tag{36}$$

The last line follows from the Addition Theorem.

We mention the fact that the surface gradients of spherical harmonics are orthogonal in a certain sense and complete. They can be used to expand surface vector fields, that is vectors defined in $S(1)$ that are always tangent to the sphere. The relevant relation is

$$(\nabla_1 Y_l^m,\, \nabla_1 Y_n^k) = \int_{S(1)} d^2\hat{\mathbf{s}}\; \nabla_1 Y_l^m(\hat{\mathbf{s}})\cdot\nabla_1 Y_n^k(\hat{\mathbf{s}})^* = l(l+1)\,\delta_{ln}\delta_{mk} \,. \tag{37}$$

Next, some philosophy and another expansion. Spherical harmonics occur everywhere in geophysics because they are the natural functions in which to expand functions on a sphere. In general, if you want to see how something affects a spherically symmetric object, you should expand the something is spherical harmonics and then allow each of the terms to interact with the sphere separately, which is usually easy to do. An example, not used in this course, but maybe useful to you elsewhere is scattering theory. Suppose a plane wave comes along and hits a sphere; if a single spherical harmonic excited the sphere we could get the response from the normal modes. So if we write the following expansion we are essentially done:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{4\pi}{kr}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} e^{i\pi l/2}\sqrt{\frac{\pi kr}{2}}\,J_{l+½}(kr)\;Y_l^m(\hat{\mathbf{r}})\,Y_l^m(\hat{\mathbf{k}})^* \,. \tag{38}$$

Many of the differential equations of mathematical physics contain the operator ∇^2 and an expansion in spherical harmonics of the solution decouples the radial and azimuthal parts of the equation. We shall see this in the next Section, where we solve Laplace's equation that way. But the method is much more general and is used in many branches of geophysics and physics. For example, these days, the equations for convection the mantle are solved on the computer by a spherical harmonic expansion, and so are those for the geomagnetic dynamo.

Finally, in this tail end of briefly mentioned things, we come to the 3-j symbols. Occasionally one runs into the need to perform integrals over the sphere of products of three spherical harmonics. These can always be done simply, and there are symmetry relationships that make many such products vanish. The problem was first studied systematically in quantum mechanics of angular momentum, and so the reference everyone uses is: Edmonds, *Angular Momentum in Quantum Mechanics*, Princeton, 1960. There is a table made worthless because if its lack of supporting explanation in Chap 27 Abramowitz and Stegun, *Handbook of Mathematical Functions*, Dover, 1970. This book is worth knowing about in general, and has many results on Associated Legendre functions in Chaps 8 and 22.

A Table of Spherical Harmonic Lore

	Property	Formula	Comments
1.	Laplacian in polar coordinates	$\nabla^2 = \frac{1}{r^2}\nabla_1^2 + \frac{1}{r}\frac{\partial^2 r}{\partial r^2}$	∇_1^2 is angular part of familiar Laplacian
2.	Eigenvalue	$\nabla_1^2 Y_l^m = -l(l+1)Y_l^m$, $l=0,1,2,\cdots$	There are $2l+1$ linearly independent eigenfunctions per l
3.	Orthogonality	$\int d^2\hat{\mathbf{s}}\, Y_l^m(\hat{\mathbf{s}})\, Y_n^k(\hat{\mathbf{s}})^* = 0$, unless $l=n$ and $m=k$	True for every normalization
4.	Theoretician's normalization	$\int d^2\hat{\mathbf{s}}\, \lvert Y_l^m(\hat{\mathbf{s}})\rvert^2 = 1$	Other choices: 4π or $4\pi/(2l+1)$
5.	Completeness	$f(\hat{\mathbf{s}}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} c_{lm} Y_l^m(\hat{\mathbf{s}})$	Works for any reasonable function f on $S(1)$
6.	Expansion coefficients	$c_{lm} = \int d^2\hat{\mathbf{s}}\, f(\hat{\mathbf{s}})\, Y_l^m(\hat{\mathbf{s}})^*$	Requires property 4
7.	Addition Theorem	$\frac{2l+1}{4\pi} P_l(\hat{\mathbf{s}}\cdot\hat{\mathbf{r}}) = \sum_{m=-l}^{l} Y_l^m(\hat{\mathbf{s}})\, Y_l^m(\hat{\mathbf{r}})^*$	Requires property 4
8.	Wavelength of Y_l^m	$\frac{2\pi}{l+½}$	Depends only on degree l, not on order m or $\hat{\mathbf{s}}$
9.	Appearance	Re Y_l^m vanishes on $2m$ meridians and $l-m$ parallels	Im Y_l^m the same but rotated about $\hat{\mathbf{z}}$
10.	Parseval's Theorem	$\int d^2\hat{\mathbf{s}}\, \lvert f(\hat{\mathbf{s}})\rvert^2 = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \lvert c_{lm}\rvert^2$	Requires property 4. Get RMS value of f by dividing by 4π and taking square root
11.	Generating function	$\frac{1}{(1-2\mu r + r^2)^{½}} = \sum_{l=0}^{\infty} r^l P_l(\mu)$	Often used in conjunction with property 7
12.	Another orthogonality	$\int d^2\hat{\mathbf{s}}\, \nabla_1 Y_l^m(\hat{\mathbf{s}})\cdot\nabla_1 Y_n^k(\hat{\mathbf{s}})^* = l(l+1)\,\delta_{ln}\,\delta_{mk}$	Very useful but little known! Requires property 4.

8. Solution of Laplace's Equation in Spherical Harmonics

This is the application of spherical harmonics that we need for potential theory. We consider the region inside a spherical shell $R_1 \le r \le R_2$ inside which Laplace's equation (6.5) is obeyed by V, a potential. Sometimes we will let R_2 tend to infinity, but it is useful to keep it finite for now. Let us write V as a function of spherical polar coordinates $V(r, \theta, \phi)$ relative to O the origin at the center of the concentric spheres. Provided V is reasonably well behaved we can write it thus:

$$V(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} V_{lm}(r) Y_l^m(\theta, \phi) \; . \tag{1}$$

This follows because in any spherical surface the function V can be expanded in surface harmonics, and (1) simply says there is a different set of expansion coefficient for each radius r. An expansion like (1) is perfectly general, and does not depend on V being harmonic. Let us now add the constraint that V obeys Laplace's equation, which we write using (7.4):

$$\frac{1}{r} \frac{\partial^2 rV}{\partial r^2} + \frac{1}{r^2} \nabla_1^2 V = 0 \; . \tag{2}$$

Inserting (1) into (2) and recalling the eigenfunction property (7.11) that

$$\nabla_1^2 Y_l^m = -l(l+1) Y_l^m \tag{3}$$

we obtain

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left[\frac{1}{r} \frac{d^2 rV_{lm}}{dr^2} - \frac{l(l+1)}{r^2} V_{lm} \right] Y_l^m(\theta, \phi) = 0 \; . \tag{4}$$

Because of the orthogonality of the surface harmonics the only way that this sum can be zero is if the factor in square brackets vanishes identically for every r and each l and m separately. Thus we find the ordinary differential equations:

$$\frac{d^2 rV_{lm}}{dr^2} - \frac{l(l+1)}{r} V_{lm} = 0 \; . \tag{5}$$

Notice the $2l+1$ different functions with the same l but differing m obey the same differential equation. The standard way of solving such equations is by substituting a power series. Omitting the details, we get

$$V_{lm}(r) = A\, r^l + \frac{B}{r^{l+1}} \tag{6}$$

as the most general solution. Clearly each degree and order in (4) will generally be associated with different constants A and B; thus the general solution to (2) is given by

$$V(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left[A_{lm} r^l + \frac{B_{lm}}{r^{l+1}} \right] Y_l^m(\theta, \phi) \tag{7}$$

where the constants A_{lm} and B_{lm} are coefficients in the spherical harmonic expansion of the potential; they must be determined by experiment or analysis for each different potential V.

There is a physical interpretation of the two kinds of coefficients: We see that the contribution to V (and the associated field vector field $\mathbf{g}$) from the A_{lm} grows with increasing radius r. This implies that the sources, that is the matter in the gravitational case, lie in the region outside R_2, since V increases with proximity to the sources. Similarly B_{lm} is associated with matter inside the inner sphere, radius R_1. The sum of the A_{lm} terms is called the *external part* of the potential and the B_{lm} sum the *internal part*; and the coefficients A_{lm} are the external coefficients, and B_{lm} are internal ones. They are identified with the location of the sources that generate the potentials, inside the inner sphere, $S(R_1)$ for interior, and outside the outer sphere $S(R_2)$ for the external. If the expansion (7) describes the gravitational field of the Earth, with O at its center, there can be no exterior part, as there is no matter outside the sphere radius R_1, which can be chosen to be the equatorial radius a. Then it is conventional to rewrite (7) thus

$$V = -\frac{Gm_E}{r}\left[1 + \sum_{l=2}^{\infty}\sum_{m=-l}^{l} c_{lm}\left(\frac{a}{r}\right)^l Y_l^m(\theta,\phi)\right] \tag{8}$$

Notice the absence of $l = 1$ terms because O is at the center of mass (recall (4.4)) so there is no dipole term. Also note how the coefficients c_{lm} are dimensionless, as the radial terms are scaled by the so-called reference radius a. As we mentioned earlier, a is often the equatorial radius, because then the inner sphere encloses all the matter; sometimes the Earth's mean radius is used instead (often people forget to say!). Of course, practical models of the Earth's potential do not sum to infinity, but very large degree model have been found with $l_{\max} = 360$; how many coefficients is that?

The interior field in (8), that is the field with interior sources, is given a multipole expansion, each degree corresponding to a particular multipole potential. If we compare (8) to (3.18) we see that MacCullagh's expansion is just the first few terms of the full multipole series. The J_2 term corresponds to $l = 2$, $m = 0$ term. Of course the natural z axis of the Earth is the spin axis, which is on average the principal axis of greatest moment of inertia. I say on average, because the Earth wobbles, and the spin axis, the angular momentum vector and the main principal axis do not in fact coincide – they lie in a plane, with separation around 3 meters rms at the north pole and plane precesses. This is the *Chandler wobble,* period 430 days. See the Annex notes, G.

Let us briefly apply the theory to complete the work began by MacCullagh. Refer to the MacCullagh figure. Then we write

$$-\frac{V(\mathbf{r})}{G} = \int_B d^3\mathbf{s}\,\rho(\mathbf{s})\,\frac{1}{|\mathbf{r}-\mathbf{s}|}\,. \tag{9}$$

Recall the spherical harmonic expansion of the potential due to a point mass (7.36); insert this into (1):

$$-\frac{V}{G} = \int_B d^3\mathbf{s}\,\rho(\mathbf{s})\sum_{l=0}^{\infty}\frac{4\pi}{2l+1}\,\frac{s^l}{r^{l+1}}\sum_{m=-l}^{l} Y_l^m(\hat{\mathbf{r}})\,Y_l^m(\hat{\mathbf{s}})^* \tag{10}$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{r^{l+1}} Y_l^m(\hat{\mathbf{r}}) \left[\int_B d^3\mathbf{s}\, \rho(\mathbf{s}) s^l \frac{4\pi}{2l+1} Y_l^m(\hat{\mathbf{s}})^* \right]. \tag{11}$$

We recognize (11) as a form of (7) with only an exterior part, and now we have an explicit expression for the coefficients in terms of integrals over the density: in the notation of (8)

$$c_{lm} = \frac{4\pi}{(2l+1)m_E} \int_B d^3\mathbf{s}\, \rho(\mathbf{s}) \left(\frac{s}{a} \right)^l Y_l^m(\hat{\mathbf{s}})^* . \tag{12}$$

Exercise

8.1 From the small spherical harmonic table given later and other information in the notes, calculate the three principal moments of inertia of the Earth. Also find the geographic coordinates of principal axes. Do not assume $A = B$, but you may assume (4.8) gives the correct trace.

8.2 If the Earth were a fluid of uniform density, 5,500 $\mathrm{kg\,m^{-3}}$, rotating at its present rate, what would be its flattening f, and its J_2?
Hint: Add a thin layer of material to a sphere and use Clairaut's analysis.

9. The Geoid continued

We have already seen that the shape of the Earth is defined as an equi-geopotential surface. Since this is a difficult property of the Earth to determine it is convenient to have a defined *reference ellipsoid* which closely approximates the geoid. The choice of a oblate ellipsoid is no accident: a uniform gravitating fluid body assumes an ellipsoidal shape if the body is not rotating too fast. The standard body has an equatorial radius $a = 6,378,136$ m, a polar radius $c = 6{,}356{,}751$ m and a flattening $f = 1/298.257$; see page 8 in Yoder's Chapter of Ahrens' book. The difference in geoid elevation from the reference ellipsoid is called the *geoid anomaly* or *geoid height*. It ranges from −105 m near the southern tip of India to 73 m near Indonesia. Since $f \approx 1/300$, the departure from a sphere is about 20 km. It is easy to see that geoid height is directly related to the local geopotential as follows: local gravity is given by

$$\mathbf{g} = -\nabla U = -\hat{\mathbf{n}}\,\frac{\partial U}{\partial n} \tag{1}$$

and clearly $|\mathbf{g}| = g = \partial U/\partial n$. Note though U is negative, it increases outwards; g, a magnitude, is always positive, even though $\mathbf{g}$ points downwards. A Taylor expansion of U for an observer at $\mathbf{s}$ on the reference ellipsoid begins:

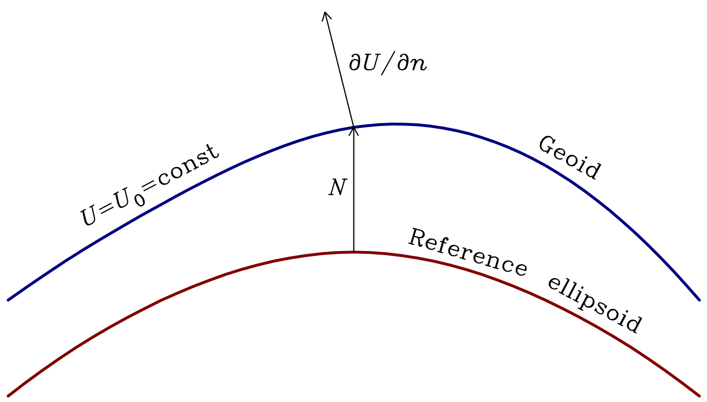
$$U_0 = U(\mathbf{s}) + N\,\frac{\partial U}{\partial n} + \cdots \tag{2}$$

where N is the local geoid height (also called the *geoid anomaly*) and U_0 is the value of the geopotential on the geoid. Solving for N in (2) we find

$$N = -\frac{U(\mathbf{s}) - U_0}{g(\mathbf{s})} = -\frac{\Delta U}{g} \tag{3}$$

and since g varies by less than a percent it is an excellent approximation to say geoid height is simply proportional to minus local geopotential deviation from the standard U_0; this difference is conventionally denoted by T but we will not use that notation. Notice in some books (including Blakely's) the sign is reversed in (3) because of the choice of a different sign for the gravitational potential!

Figure 5a



All the books will tell you that U is constant on the geoid, but it is hard to find out how that constant value is determined. In our earlier discussion of Clairaut's formula for the shape of the Earth we took the equi-geopotential intersecting the Earth at the equator. The geoid's potential is the mean value of U over the reference ellipsoid.

Suppose we cover the Earth with a thin sheet of matter with surface density in the form of a single spherical harmonic:

$$\sigma(\hat{\mathbf{s}}) = \beta\, Y_l^m(\hat{\mathbf{s}})\ . \tag{4}$$

What would be the resultant gravity change in g, traditionally called the *gravity anomaly*? The answer is not quite as simple as you might think because gravity anomalies are always referred to local sea level (= local geoid) and the presence of the new matter shifts the geoid via (3). Let us ignore the ellipticity as a good approximation; then $\hat{\mathbf{r}} = \hat{\mathbf{n}}$. The two factors causing a change in gravity are (a) the new attraction, (b) the movement of the observer through the gradient of the main field as the geoid height N shifts:

$$\Delta g = \frac{\partial \Delta U}{\partial r} + N\,\frac{\partial g}{\partial r}\ . \tag{5}$$

Since $g \approx g(a)(a/r)^2$, we see that $\partial g/\partial r = -\,2g/r$. Substituting this and (3) into (5) we get

$$\Delta g = \frac{\partial \Delta U}{\partial r} + \frac{-\Delta U}{g}\cdot\frac{-2g}{r} \tag{6}$$

$$= \frac{\partial \Delta U}{\partial r} + \frac{2\Delta U}{r}\ . \tag{7}$$

From (3) and the fact that $\rho(\mathbf{s}) = \delta(s-a)\,\sigma(\hat{\mathbf{s}})$, the change in geopotential is

$$-\frac{\Delta U}{G} = \frac{4\pi a\beta}{2l+1}\left(\frac{a}{r}\right)^{l+1} Y_l^m \tag{8}$$

and so finally differentiating and substituting into (5) gives

$$\Delta g = -\frac{4\pi a^{l+2} G\beta}{2l+1}\left[-\frac{l+1}{r^{l+2}} + \frac{2}{r^{l+2}}\right] Y_l^m \tag{9}$$

$$= \frac{4\pi G(l-1)}{2l+1}\,\beta\, Y_l^m \quad \text{on } r = a\ . \tag{10}$$

You can see as l becomes very large this is the familiar $2\pi G\sigma$ for a flat sheet, so the effect of the geoid shift is largest at the long wavelengths. What is happening for $l = 0$, and $l = 1$?

This is a good place to discuss Earth coordinate systems – the three kinds of latitude. Mathematically, we work with a spherical polar coordinate system, the one in Figure 4. The angle θ is often called the colatitude, and the corresponding angle to the radius vector measured from the x-y plane would be the latitude; but strictly this is the *geocentric latitude*, for the obvious reason that the angle is measured at the center of the Earth. In geophysics, the position of a satellite is always given in geocentric coordinates.

But if you are standing on the Earth, because the geoid is not a sphere, the local vertical or normal to the geoid, does not pass through the Earth's center. Until the advent of GPS, latitudes were determined by astronomical means, for example by measuring the angle of the pole star from local vertical. There is no easy way to determine the direction to the center of the Earth, but local vertical is readily determined. Thus *astronomical latitude* is the angle made between the equatorial plane and the local vertical.

However, this system of latitude also has a drawback: the corresponding coordinate grid is very complicated, because it is tied to the shape of the true geoid with its almost random highs and lows – there is no analytic formula to relate a position given by astronomical latitude and longitude to a Cartesian frame. So finally, we come to *geographic latitude*. The site is moved vertically onto the reference ellipsoid, and the normal to ellipsoid is used to define the geographic latitude in the analogous way. This is the latitude you will find on maps, and obviously the coordinate system has a fairly simple shape. Before GPS, to determine the geographic latitude you do have to know the actual shape of the local geoid, that is, the geoid anomaly, to correct the astronomical latitude, which was the only directly observable kind of latitude.

The difference between geocentric and geographic latitude for a given point can be very large and one must be very careful not to use the wrong one. The differences between astronomical and geographic latitudes are very small because, as we will shortly see, the geoid anomaly is typically only some tens of meters, less than one part in 10^5.

Exercise

9.1 Derive an expression that converts the geographic latitude to geocentric. What is the maximum error in distance on the ground you would commit if you improperly used geocentric latitude, when the geographic value was given? At what (geographic) latitude does the largest error arise?

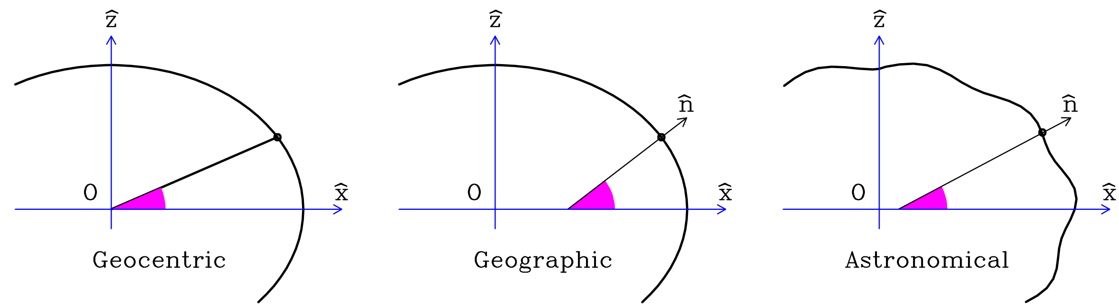


Figure 5b

10. Determination of the Geoid - I

The long wavelength geoid is found by the generalization of the process we sketched earlier for J_2. We use (8.7), the expansion of the gravitational potential V to calculate the shape of the geoid via (9.3). The expansion (8.7) is called *Stokes' expansion* and the c_{lm} are *Stokes' coefficients* in the geodetic literature. To get the c_{lm} careful records are kept of artificial satellite orbits, whose positions in space are triangulated from the ground via their radio time-of-flight delays, based on very accurate on-board clocks. Then the equations of motion of a small body in the field represented as (8.7) can be solved and a massive least-squares fitting program is undertaken for fitting the unknown coefficients. There are complications, including corrections for the perturbations to the orbits of the fields from the sun and moon. The table below is a small one abstracted from the huge degree-360 model of Richard Rapp. It will be obvious that the potential V is real yet the coefficients c_{lm} are complex. This has the effect that $c_{l,-m} = (-1)^m (c_{lm})^*$. Thus the negative-order coefficients need not be listed. The numbers given here are different from those you will normally see because we give the complex coefficients. In addition to the table there is Figure 6, a contour plot of geoid anomaly in meters from a model that goes out to $l = 20$. We note the geoid has no particular relation to the presence of oceans or continents old crust or young – it is a major disappointment to geologists that nothing stands out and that is partly because of isostasy. Somewhat surprisingly the numerical values reveal something interesting.

Complex Spherical Harmonic Coefficients for V (Table gives $c_{lm} \times 10^7$)

l	m	Re	Im	l	m	Re	Im	l	m	Re	Im
2	0	–17163.20	0	5	2	16.33	8.06	7	3	–6.26	–5.40
2	1	0	0	5	3	11.33	–5.39	7	4	–6.92	3.11
2	2	61.12	35.09	5	4	–7.42	–1.24	7	5	0.03	0.44
3	0	33.93	0	5	5	–4.26	–16.78	7	6	–8.99	–3.81
3	1	–50.88	6.25	6	0	–5.33	0	7	7	0.00	0.59
3	2	22.68	15.55	6	1	1.98	0.68	8	0	1.76	0
3	3	–18.06	35.46	6	2	1.15	9.38	8	1	–0.67	1.44
4	0	19.16	0	6	3	–1.44	0.19	8	2	1.99	–1.66
4	1	13.42	–11.88	6	4	–2.23	11.80	8	3	0.45	–2.25
4	2	8.79	–16.64	6	5	6.70	–13.47	8	4	–6.19	–1.76
4	3	–24.84	–5.07	6	6	0.23	5.94	8	5	0.64	2.19
4	4	–4.78	–7.76	7	0	3.22	0	8	6	–1.63	–7.77
5	0	2.43	0	7	1	–6.94	2.44	8	7	–1.68	1.88
5	1	1.46	–2.40	7	2	8.22	–2.26	8	8	–3.09	–2.95

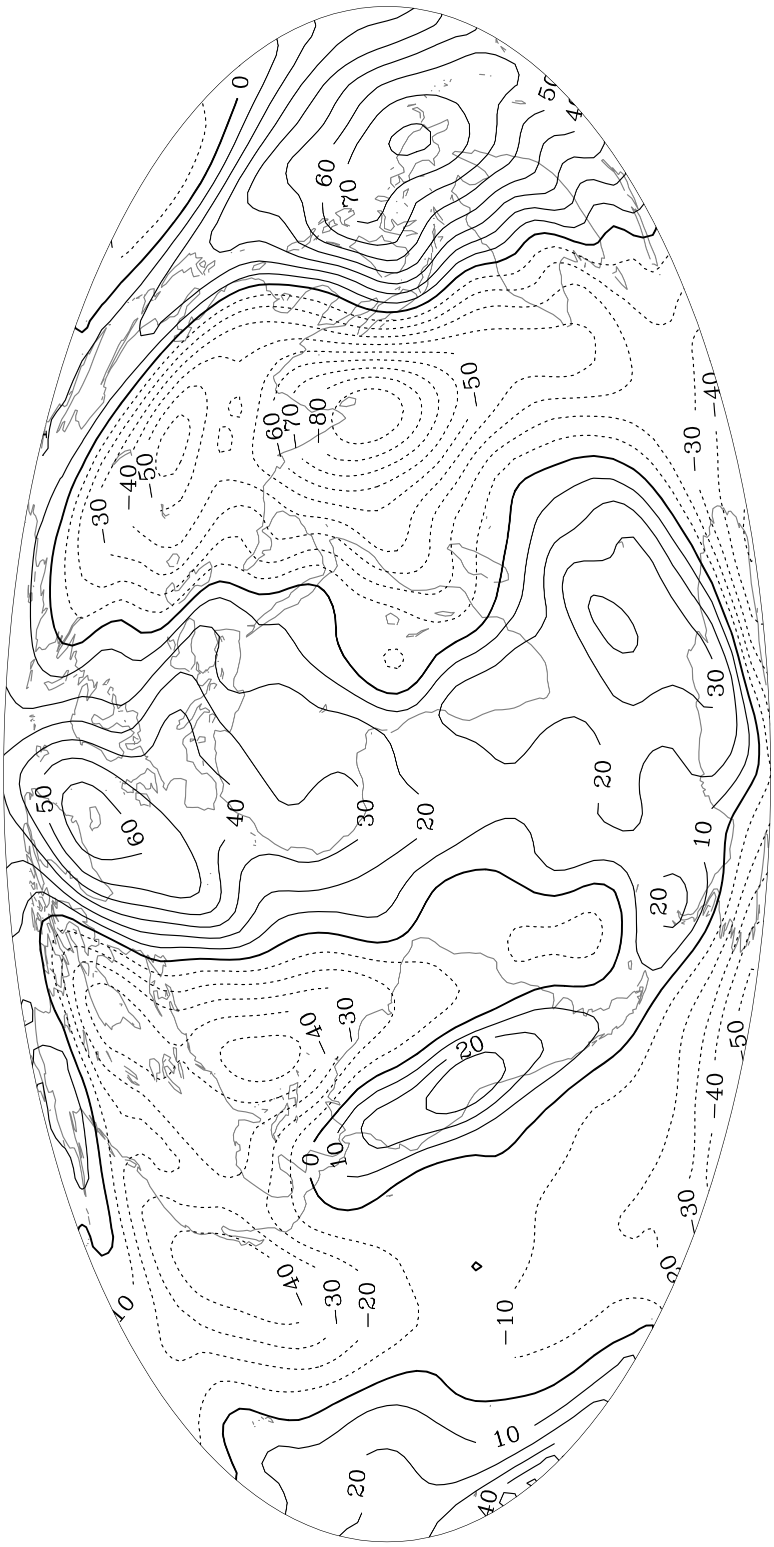


Figure 6: Geoid anomaly in meters

Let us concentrate on the $l=2$ coefficients. We note two things at once. First the $m=0$ term is largest by a factor of 400. Second, the $l=2, m=1$ entry is zero. The large $m=0$ value arises because the Earth rotates about its major principal axis of inertia and is flattened. The greatest departure from a spherical shape (with everything in $l=0$) is from the flattening. We return to this in a moment. Next there are no $m=1$ terms. This is also because the Earth rotates about its principal axis of inertia, and we have aligned the $\hat{\mathbf{z}}$ with that axis. Recall from (8.12) that

$$m_E a^2 c_{21} = \int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, s^2\, \frac{4\pi}{5}\, Y_2^1(\hat{\mathbf{s}})^* \tag{1}$$

$$= \int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, s^2\, \frac{4\pi}{5}\, N_{21}\, 3\cos\theta\, \sin\theta\, \mathrm{e}^{-\mathrm{i}\phi} \tag{2}$$

$$= \int_B d^3\mathbf{s}\, \rho(\mathbf{s})\, \frac{12 N_{21}\pi}{5}\, (s_1 s_3 - \mathrm{i} s_2 s_3) \tag{3}$$

where we note that $s_1 = x = s\sin\theta\cos\phi$, $s_2 = y = s\sin\theta\sin\phi$ and $s_3 = z = s\cos\theta$, and we have used the explicit formula on p 17 for P_2^1. Aside from constant factors, these are the Q_{13} and Q_{23} terms of the quadrupole term of MacCullagh's formula; see (3.18) and (3.19). Now the Greenwich meridian is *not* a principal axis of inertia, so referred to the conventional axis system the inertia tensor is not diagonal but in the form

$$M_{ij} = \begin{bmatrix} M_{11} & M_{12} & 0 \\ M_{21} & M_{22} & 0 \\ 0 & 0 & C \end{bmatrix}. \tag{4}$$

Verify this! Hence by (3.21)

$$Q_{13} = ½ M_{kk}\delta_{13} - \frac{3}{2} M_{13} = 0\ . \tag{5}$$

and similarly with Q_{23}. Thus $c_{21} = 0$, as we had wished to show.

We return to the question of the size of the c_{20} term. Suppose we say the Earth on the longest time scales has no rigidity, it behaves like a density stratified liquid: then we can compute the shape it would assume under rotation. (How is this done, exactly? We don't have time and it's not part of this course, but you might like to think about it.) Then the J_2 coefficient turns out to be 0.0010727 which translates to $-17,006 \times 10^{-7}$ for c_{20}. We see that the real Earth is more oblate than the hydrostatic model by an amount 157.2 in the units of the table. Thus the excess flattening in c_{20} is bigger than the other coefficients, which also represent out of equilibrium gravitational fields. Why is this? An idea of Walter Munk concerns tidal forces. We may discuss tides in more detail later, but for now it is sufficient to note that the moon exerts a braking torque on the Earth, and in addition to theoretical calculations, the length of the day is increasing by on the order of a fraction of a millisecond per year. If we imagine the Earth to be a viscous fluid on the time scales of this effect, it is clear that the flattening we observe today, corresponds to a slightly faster spin rate of the past, as the viscous mantle is not in equilibrium with the present rotation rate – it lags behind. This sounds good until one puts in the numbers: it turns out the viscosity would be so high to achieve the present excess flattening that mantle convection would be prohibited. The explanation favored today was given by Goldreich and Toomre (*JGR*, 74, p 2555, 1969) and it is quite subtle. The Earth rotates about a major principal axis because off-axis rotation involves wobble, which causes shear in the mantle and frictional energy losses that gradually reduce wobble amplitude to zero. But the density structure within the Earth is not constant in time, but changes through convection. If these changes are slow compared with the wobble decay rate (and they are, because the connective overturn is about 10^8 years, while the wobble decay time is merely tens of years) it can be shown that the non rigid Earth will deform so as to maintain the rotation axis on the major principal axis.

Exercises

10.1 Calculate an accurate numerical value for the geopotential on the geoid from the table of spherical harmonics and other data in the notes. The answer is not simply $U_0 = -Gm_E/a$. Give a physical interpretation of U_0.

10.2 The product Gm_E is given as 398,600.440 $\mathrm{km}^3\,\mathrm{s}^{-2}$ in the *Global Earth Physics*, but G is known to an accuracy of 1 part in 10^4 at best. Assuming the given number of figures is meaningful (and it is) explain how it is possible to find Gm_E to 9 significant figures, without knowing G to this accuracy.

11. A Few Properties from Spherical Harmonics

A table of the spherical harmonic coefficients can be used to find a number of interesting properties. We calculate the root-mean-square (RMS) geoid anomaly, the amount the true geoid departs from the reference ellipsoid. We have already seen from Figure 6 it must be some tens of meters. The calculation is straightforward if one is willing (as we are) to accept (9.3), the proportionality of geoid anomaly to gravitational potential T, as a good enough approximation and we hold g constant.

First let us find a spherical harmonic expansion for the geoid height, an important result in itself. Let $\mathbf{s} \in S(a)$, and $\mathbf{r}$ be in the reference ellipsoid; and $\mathbf{r} = \mathbf{s} + h\hat{\mathbf{s}}$; we want an expression for N in terms of $\mathbf{s}$ in order to apply spherical harmonic analysis. From (9.3)

$$-gN = U(\mathbf{r}) - U_0 = U(\mathbf{s}) + [U(\mathbf{r}) - U(\mathbf{s})] - U_0. \quad (1)$$

Now we use the definition of the geopotential (5.9), that $U = V - ½r^2\omega^2$ and the normal gradient approximation from (9.2)

$$-gN = V(\mathbf{s}) - ½a^2 \sin^2\theta + h\,\frac{\partial U}{\partial n} - U_0 \quad (2)$$

$$= V(\mathbf{s}) - ½a^2 \sin^2\theta + hg - U_0\,. \quad (3)$$

$$= V(\mathbf{s}) - ½a^2 \sin^2\theta - afg\cos^2\theta - U_0\,. \quad (4)$$

Note that we have also used the fact that $h = -af\cos^2\theta$ with f as the flattening from Clairaut. Next we separate out the Y_2^0 spherical harmonic, or axial quadrupole, from the gravitational potential V, taking advantage of the factor that the radius is constant on $S(a)$:

$$-gN = \tilde{V}(\mathbf{s}) + c_{20}Gm_E Y_2^0(\hat{\mathbf{s}})/a - ½\omega^2a^2\sin^2\theta - agf\cos^2\theta - U_0 \quad (5)$$

$$= \tilde{V}(\mathbf{s}) + [-(Gm_E/a)J_2P_2(\cos\theta) - ½\omega^2a^2\sin^2\theta - agf\cos^2\theta] - U_0 \quad (6)$$

$$= \tilde{V}(\mathbf{s}) + [-(Gm_E/a)J_2(3\cos^2\theta - 1)/2 - ½\omega^2a^2(1-\cos^2\theta) - agf\cos^2\theta] - U_0 \quad (7)$$

$$= \tilde{V}(\mathbf{s}) - \cos^2\theta\left[\frac{3Gm_EJ_2}{2a} - \frac{\omega^2a^2}{2} + agf\right] + \text{constant}\,. \quad (8)$$

Because of Clairaut's formula, and the fact that $g = Gm_E/a^2$ the term in brackets vanishes. So we have that gN has essentially the same spherical harmonic expansion as V except that we drop the $l = 0$ constant and the axial quadrupole terms.

$$N(\hat{\mathbf{s}}) = \frac{Gm_E}{ag}\sum{}' c_{lm}\,Y_l^m(\hat{\mathbf{s}}) = a\sum{}' c_{lm}\,Y_l^m(\hat{\mathbf{s}}) \quad (9)$$

where the prime means omit the monopole and axial quadrupole terms. Thus, except for the very longest wavelengths, the coefficients c_{lm} in the expansion for V, are also coefficients of the geoid height expansion. To get geoid height one multiplies the c_{lm} by a.

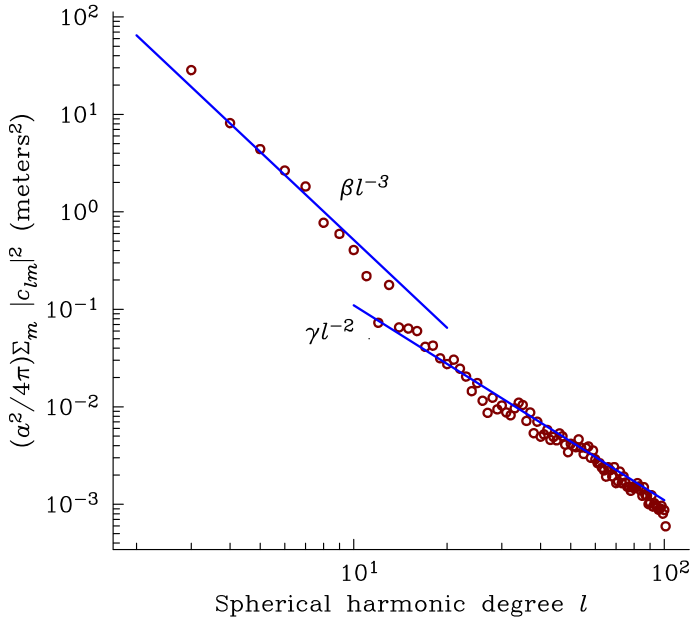


Figure 7

Let us now look at the mean-square geoid height

$$< N^2 > = \frac{1}{4\pi} \int_{S(1)} d^2\hat{\mathbf{r}}\ N(\hat{\mathbf{r}})^2 = \| N(\hat{\mathbf{r}}) \|^2 \tag{10}$$

By Parseval's Theorem, property (15) in the table of SH Lore:

$$< N^2 > = \frac{a^2}{4\pi} \sum_l \left[\sum_{m=-l}^{l} | c_{lm} |^2 \right] . \tag{11}$$

So we see from (10) that the sum of absolute value squared of the SH coefficients for the gravitational potential is proportional to the squared amplitude of the geoid height. When I did the sum from a table that went out to $l = 20$, I obtained 30.42 meters for N_{RMS}.

I have written the sums over the order m separately, because each degree l represents a length scale, or wavelength $2\pi a/(l + ½)$ according to Jean's formula. Thus when we talk about the amount of power (squared amplitude) in each degree, we are effectively describing a *power spectrum* of geoid height. This is plotted in Figure 7. Also drawn in the figure is a line proportional to l^{-3}, representing *Kaula's Rule of Thumb* (Kaula, W. M., *Theory of Satellite Geodesy*, 1966) which Bill Kaula of UCLA gave when there were very few measured coefficients. As you can see the rule is not very accurate, but at least it gives the correct impression that the geoid height spectrum is dropping off in a roughly reciprocal polynomial way. In fact in the Annex we show we might expect the spectrum to fall off more like l^{-2}, which it does rather well for higher values of the degree. A model for this is a random collection of point sources; I prove this in the Annex. You will see later this behavior is completely different from a similar kind of spectrum for the Earth's magnetic field, which has an exponential fall-off at the low degree end. The reason for this difference is the distribution throughout the mantle of the sources of the gravity field, while currents generating the main field start abruptly at the core. More of this in the magnetism part of the course, where the spherical harmonic power spectrum is referred to a good deal.

12. The Equivalent Source Theorem

Let us think again about the question of estimating the interior density of the Earth from these gravity data. Recall we had two integrals (4.8) and (4.9) when the mass and moment of inertia are known. From the satellite data we have thousands of observations of c_{lm}, each giving an integral over density; see (8.12). Surely from all this information we can get a great estimate of $\rho(\mathbf{s})$. Unfortunately this is not so. We can see this from a famous result of potential theory, which we now come to. Knowledge of the c_{lm} is equivalent to knowledge of V and hence $\mathbf{g}$ outside the Earth. So let us give ourselves complete knowledge of V in $\bar{B}$. First let's derive *Green's Theorem*. Recall (6.11), which we write out again for convenience:

$$\nabla \cdot (f \nabla g) = \nabla f \cdot \nabla g + f \nabla^2 g \; . \tag{1}$$

Obviously by interchanging f and g we have:

$$\nabla \cdot (g \nabla f) = \nabla g \cdot \nabla f + g \nabla^2 f \; . \tag{2}$$

Subtracting these two:

$$\nabla \cdot (f \nabla g) - \nabla \cdot (g \nabla f) = f \nabla^2 g - g \nabla^2 f \; . \tag{3}$$

Now integrate over B and apply Gauss's Divergence Theorem and we have our result, Green's Theorem:

$$\int_B d^3\mathbf{s}\, (f \nabla^2 g - g \nabla^2 f) = \int_{\partial B} d^2\mathbf{s}\, \hat{\mathbf{n}} \cdot (f \nabla g - g \nabla f) \; . \tag{4}$$

Recall (6.7), the generalization of Laplace's equation to a region containing matter is Poisson's equation:

$$\nabla^2 V = 4\pi G \rho \; . \tag{5}$$

Let $\mathbf{r}$ lie outside B and $\mathbf{s}$ inside. In (4) let

$$f = V, \quad \text{and} \quad g = \frac{1}{|\mathbf{r} - \mathbf{s}|} \; . \tag{6}$$

Then

$$\int_B d^3\mathbf{s} \left(V \nabla^2 \frac{1}{|\mathbf{r}-\mathbf{s}|} - \frac{1}{|\mathbf{r}-\mathbf{s}|} \nabla^2 V \right) = \int_{\partial B} d^2\mathbf{s}\, \hat{\mathbf{n}} \cdot \left(V \nabla \frac{1}{|\mathbf{r}-\mathbf{s}|} - \frac{1}{|\mathbf{r}-\mathbf{s}|} \nabla V \right). \tag{7}$$

Remember the differential operators here are with respect to the source point $\mathbf{s}$. Then from Section 6

$$\nabla^2 \frac{1}{|\mathbf{r}-\mathbf{s}|} = \nabla^2 \frac{1}{|\mathbf{s}-\mathbf{r}|} = 0, \quad \mathbf{r}\,! - \mathbf{s} \tag{8}$$

because $1/r$ is harmonic except at the origin, and since the observer is outside B the function is never singular; we can substitute Poisson's equation (5) on the left too and rewrite the right side:

$$\int_B d^3\mathbf{s}\, \frac{-4\pi G \rho(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} = \int_{\partial B} d^2\mathbf{s} \left(V \frac{\partial}{\partial n} \frac{1}{|\mathbf{r}-\mathbf{s}|} - \frac{1}{|\mathbf{r}-\mathbf{s}|} \frac{\partial V}{\partial n} \right). \tag{9}$$

But the left side is just 4π times the potential at $\mathbf{r}$. Hence:

$$V(\mathbf{r}) = \frac{1}{4\pi} \int_{\partial B} d^2\mathbf{s} \left(V \frac{\partial}{\partial n} \frac{1}{|\mathbf{r}-\mathbf{s}|} - \frac{1}{|\mathbf{r}-\mathbf{s}|} \frac{\partial V}{\partial n} \right). \tag{10}$$

This result is called the *Equivalent Source theorem*. The two terms on the right have simple physical interpretations. The second is just the potential due to a surface layer of matter with surface density $\partial V/\partial n = g$. To understand the first, imagine putting a local coordinate system with $\hat{\mathbf{z}}$ in the $\hat{\mathbf{n}}$ direction, then $\partial/\partial n$ is the z derivative; then we see that at each point we have a contribution identical to the second term in the multipole expansion of the potential – a *dipole* normal to the surface at each point with strength V. So (10) says we can get the observed potential by coating the surface with a surface density and a surface dipole layer, however the true density is distributed within B; these are equivalent sources as far as the potential is concerned. Thus no matter how accurately we know V outside, the internal density is not determined by it, because we can replace the true density by this artificial surface distribution of sources. And of course we can get other distributions that do the job as well: imagine drawing a closed surface inside B and getting the potential from inside that by a set of surface sources, and from outside by the ordinary density. Obviously the solution is not unique in a very bad way. Equation (10) is known as the *equivalent source* representation of the gravity field.

Exercises

12.1 Newton showed a uniform spherical shell of matter has the same external gravitational field as point mass at its center, so all radially stratified density distributions have exactly the same potential, $-Gm/r$. Find a density distribution that is not radially stratified, but has this same external potential.

12.2(a) Consider a uniform spherical mass, radius a, density ρ. Calculate the gravitational potential V, inside and outside the sphere; plot the result. Repeat for the gravitational acceleration g.

(b) Now consider the density distribution generated by the Equivalent Source Theorem for this body. Calculate V and g for the equivalent sources, inside and outside the sphere.

13. Green's Function for the Exterior Dirichlet Problem on a Sphere

Equation (12.10) seems to require two kinds of surface sources, but as we saw earlier in Section 6, it is sufficient to supply $\partial V/\partial n$ over B to be able to get V (within a constant of course). It can also be proved that it is also enough to know just V over ∂B, which is known as the *exterior Dirichlet* boundary value problem. So (12.10), which requires both V and its gradient, is redundant. In fact if ∂B is smooth enough, it is always possible to write the potential outside the body in terms of an integral of the form:

$$V(\mathbf{r}) = \int_{\partial B} d^2\mathbf{s}\, V(\mathbf{s})\, \mathcal{G}(\mathbf{s},\mathbf{r}) \tag{1}$$

where $\mathcal{G}$ is called *Green's function* for the boundary value problem.

While the functions in (12.10) are quite straightforward, Green's functions are not easy to discover for bodies of general shape. Let us now apply spherical harmonics to derive Green's function for the V on ∂B problem, when $\partial B = S(a)$ the surface of a sphere, radius a. This question is also directly related to the idea of *upward continuation* of a harmonic function. The approach is simple. Given V on $S(a)$, we find its surface spherical harmonic expansion, then with all the coefficients, we can evaluate the field anywhere else via the exterior field series. So because the field has no exterior part we write (8.7) without A_{lm} terms:

$$V(r,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} B_{lm}\left(\frac{a}{r}\right)^{l+1} Y_l^m(\theta,\phi)\ . \tag{2}$$

On the surface $r = a$ and we can obtain the coefficients by the orthogonality, (7.23):

$$B_{lm} = \int_{S(1)} d^2\hat{\mathbf{s}}\ V(a\hat{\mathbf{s}})\, Y_l^m(\hat{\mathbf{s}})^* = \int_{S(a)} \frac{d^2\mathbf{s}}{a^2}\, V(\mathbf{s})\, Y_l^m(\hat{\mathbf{s}})^*\ . \tag{3}$$

At first sight it looks as though we are through, since we have a recipe: evaluate all the B_{lm} from (3), then plug them into (2) for other positions. But we want to reduce all this work to one simple integral and so find Green's function for the problem. Substitute (3) into (2) and rearrange a bit:

$$V(\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\int_{S(a)} \frac{d^2\mathbf{s}}{a^2}\, V(\mathbf{s})\, Y_l^m(\hat{\mathbf{s}})^* \left(\frac{a}{r}\right)^{l+1} Y_l^m(\hat{\mathbf{r}}) \tag{4}$$

$$= \int_{S(a)} \frac{d^2\mathbf{s}}{a^2}\, V(\mathbf{s}) \sum_{l=0}^{\infty}\left(\frac{a}{r}\right)^{l+1} \sum_{m=-l}^{l} Y_l^m(\hat{\mathbf{r}}) Y_l^m(\hat{\mathbf{s}})^*\ . \tag{5}$$

But now we recognize the sum over m from the Addition Theorem (7.24)

$$V(\mathbf{r}) = \int_{S(a)} \frac{d^2\mathbf{s}}{a^2}\, V(\mathbf{s}) \sum_{l=0}^{\infty}\left(\frac{a}{r}\right)^{l+1} \frac{2l+1}{4\pi}\, P_l(\hat{\mathbf{s}}\cdot\hat{\mathbf{r}}) \tag{6}$$

$$= \int_{S(a)} \frac{d^2\mathbf{s}}{a^2}\, V(\mathbf{s})\, \mathcal{G}(\mathbf{s},\mathbf{r}) \tag{7}$$

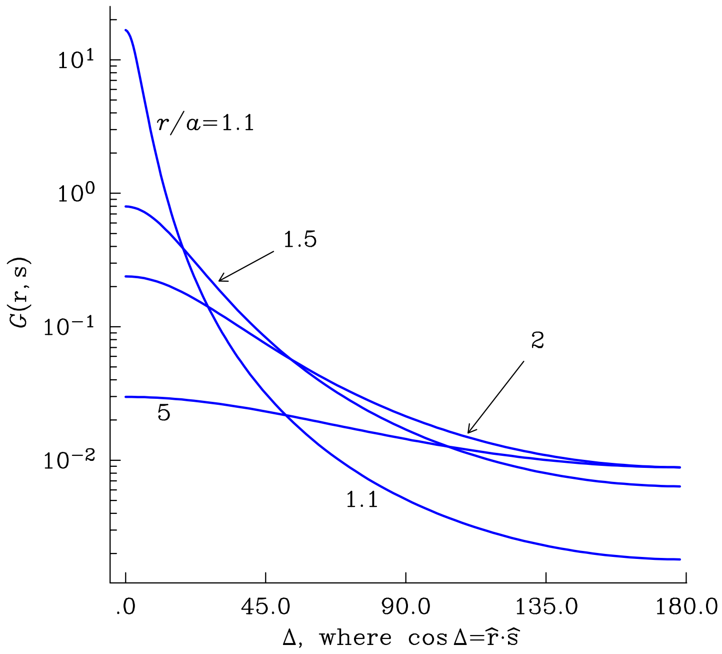


Figure 7a

where Green's function is:

$$\mathcal{G}(\mathbf{s}, \mathbf{r}) = \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) \left(\frac{a}{r} \right)^{l+1} P_l(\hat{\mathbf{s}} \cdot \hat{\mathbf{r}}) \, . \tag{8}$$

To evaluate the infinite sum in closed form requires the generating function (7.24). If we differentiate (7.24) with respect to x, we find

$$\sum_{l=0}^{\infty} l \, x^{l-1} P_l(\mu) = \frac{d}{dx} \frac{1}{(1 - 2\mu x + x^2)^{1/2}} \tag{9}$$

$$= - \frac{x - \mu}{(1 - 2\mu x + x^2)^{3/2}} \, . \tag{10}$$

Now all we need do to evaluate (8) is to add $2x^2$ times (10) to x times (7.24) and after a little algebra we get

$$\mathcal{G}(\mathbf{s}, \mathbf{r}) = \frac{1}{4\pi} \frac{s(r^2 - s^2)}{(r^2 - 2\mathbf{r} \cdot \mathbf{s} + s^2)^{3/2}} \, . \tag{11}$$

It is evident from this expression that Green's function varies over $S(a)$ as a function of angle away from the line joining the origin to the observer at $\mathbf{r}$, in other words, $\mathcal{G}$ is axisymmetric about this line. We plot its behavior in the figure.

Where (7) and (11) give a clean way of calculating the answer, the original (2) helps see the effect in a more physical way. The contribution from degree l spherical harmonic in the potential, which you will recall has a wavelength of $2\pi a/(l+½)$, is attenuated by the factor $(a/r)^{l+1}$ for the observer at radial distance r. Thus the shorter wavelength energy is more strongly attenuated relative to the long wavelength fields.

14. Thin Layers of Sources

We take this opportunity to look at the two types of surface source distribution in (12.10), the statement of the Equivalent Source Theorem: a thin sheet of mass, and a surface dipole layer, sometimes called a *doublet*. We have already briefly studied a thin layer of matter in Section 9. One way of writing the gravitational potential of such a layer is just to integrate:

$$-\frac{V(\mathbf{r})}{G} = \int_{\partial B} d^2\mathbf{s}\, \frac{\sigma(\mathbf{s})}{R} \tag{1}$$

where $R = |\mathbf{r} - \mathbf{s}|$ and σ is the surface density; this is the second term in (12.10). Here instead we solve Laplace's equation directly.

Imagine an infinite plane ($z = 0$) covered with a thin uniform layer of mass, with surface density σ, meaning the mass in any area S is σS. By symmetry the gravitational attraction above the plane, $z > 0$, has the opposite sign from that below, and the attraction at $-z$ must be equal and opposite to that at $+z$. As the plane of material is infinite in extent we may conclude that nothing varies in the x or y directions, so $\partial/\partial x = \partial/\partial y = 0$. Above the plane the potential V is harmonic:

$$0 = \nabla^2 V = -\nabla\cdot\mathbf{g} = -\left[\frac{\partial g_x}{\partial x} + \frac{\partial g_y}{\partial y} + \frac{\partial g_z}{\partial z}\right] \tag{2}$$

and therefore

$$\frac{\partial g_z}{\partial z} = 0\,. \tag{3}$$

Thus g_z is constant in space, since it doesn't vary with x or y. To find the size of the constant attractive force draw a cylinder C of any height, with cross-sectional area A, that intersects the plane, and apply the Divergence Theorem:

$$\int_C \nabla\cdot\mathbf{g}\, d^3\mathbf{s} = \int_{\partial C} \hat{\mathbf{n}}\cdot\mathbf{g}\, d^2\mathbf{s} = 2Ag_z. \tag{4}$$

There is no contribution to the surface integral from the wall of the cylinder because $\mathbf{g}$ acts only vertically; recall $g_x = -\partial V/\partial x = 0$, and similarly for g_y. Starting again at (4), we have

$$\int_C \nabla\cdot\mathbf{g}\, d^3\mathbf{s} = \int_C -\nabla^2 V\, d^3\mathbf{s} = \int_C -4\pi G\rho\, d^3\mathbf{s} = -4\pi G\sigma A. \tag{5}$$

Comparing (5) and (4) we find:

$$g_z = -2\pi G\sigma, \quad z > 0. \tag{6}$$

Integrating a stack of such thin layers gives the attraction of an infinite slab of thickness h and density ρ, the well-known attraction of a uniform slab:

$$g_z = -2\pi G\rho h, \quad z > 0. \tag{7}$$

We will encounter this simple formula as the Bouguer correction of gravity surveys.

When the matter density is concentrated onto a plane, we see that the gravitational attraction g undergoes a discontinuous jump at $z = 0$ of $4\pi G\sigma$. By using a short cylinder in the argument above, we can obtain the same value for the jump in g across a sheet of matter, when the layer is of finite extend; also (6) remains true provided the observation point is just above the layer. Even though the value of g is discontinuous at the matter layer, the potential V, which we can obtain by integrating g_z, remains continuous there.

Next we study a uniform layer of vertical dipoles. Consider two infinite parallel matter layers with opposite sign mass (obviously an idealized system), brought close together to form a dipole layer, with moment density μ. Above or below the layer, the sum of the forces is zero, that is gravitational attraction vanishes above, or below the dipole layer. But if we integrate g_z through the layer we get a *jump in the potential*:

$$\Delta V = \int_0^h \frac{\partial V}{\partial z}\, dz = \int_0^h - g_z\, dz = -4\pi G\sigma h \tag{8}$$

As h tends to zero, with σ increasing, their product is the surface dipole density μ and we find $\Delta V = -4\pi G\mu$. And this result holds also for dipole layers of finite extent.

So we have this curious set of boundary conditions:

$$\text{across a thin mass layer} \begin{cases} \Delta g_z = -4\pi G\sigma \\ \Delta V = 0 \end{cases} \tag{9}$$

$$\text{across a dipole layer} \begin{cases} \Delta g = 0 \\ \Delta V = -4\pi G\mu. \end{cases} \tag{10}$$

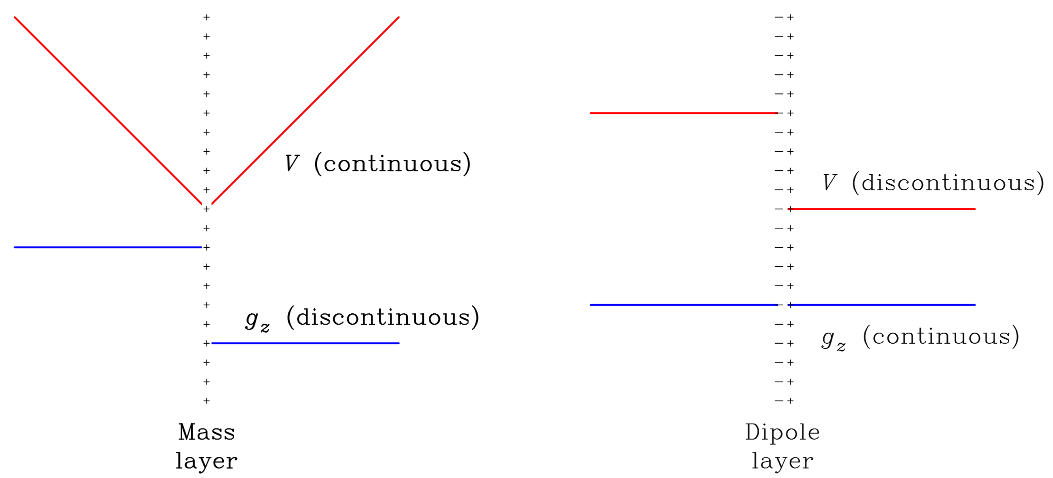


Figure 7b

We can use these boundary conditions in an alternative method for finding the gravitational potential of a thin mass layer covering a sphere. Instead of integrating (1), as we did in Section 9, we will solve Laplace's equation. On the sphere $S(a)$ we have the mass density $\sigma Y_l^m(\hat{\mathbf{r}})$, with $\nabla^2 V = 0$ both when $r < a$ and $r > a$. Let V_- be the potential inside $S(a)$ and V_+ that outside. Then

$$V_- = \frac{Ar^l}{a^l} Y_l^m(\hat{\mathbf{r}}) \;\text{ and }\; V_+ = \frac{Ba^{l+1}}{r^{l+1}} Y_l^m(\hat{\mathbf{r}}) \tag{11}$$

because V_- has exterior sources, and V_+ has only interior sources. To find the unknown constants A and B we apply the condition (9) at $r = a$. First we look at the jump in the gradient:

$$\Delta g_r = -4\pi G\sigma Y_l^m = -\left[\frac{\partial V_+}{\partial r} - \frac{\partial V_-}{\partial r}\right] \tag{12}$$

$$= \left[\frac{(l+1)B}{a} + \frac{lA}{a}\right] Y_l^m \tag{13}$$

Then we apply continuity of V:

$$0 = V_+ - V_- = [B - A] Y_l^m \tag{14}$$

This equation implies $A = B$ and substituting that into (13) gives us

$$A = B = -\frac{4\pi G\sigma a}{2l+1} \tag{15}$$

and then

$$V_+(\mathbf{r}) = -\frac{4\pi G\sigma a}{2l+1}\left(\frac{a}{r}\right)^{l+1} Y_l^m(\hat{\mathbf{r}}) \tag{16}$$

which is just what we found in (9.8).

Exercises

14.1a Consider a uniform spherical mass, radius a, density ρ. Calculate the gravitational potential V, inside and outside the sphere; plot the result. Repeat for the gravitational acceleration g.

14.1b Now consider the density distribution generated by the Equivalent Source Theorem for this body. Calculate V and g for the equivalent sources, inside and outside the sphere.

15. Determination of the Geoid - II

On shorter length scales, it is not practical or even possible to find the spherical harmonic coefficients by analyzing satellite orbits. But we still would like to know where the geoid is locally. Over the land we can go back to the idea that the geoid height is proportional to negative potential, and again we can appeal to Laplace's equation to relate potential to its derivative **g** which can be measured. So we can ask the question, Given a knowledge of g on a known surface, what is the inferred shape of the geoid? The basic idea is to use (9.7), to go from Δg to ΔU; gravity values are calculated on the standard reference ellipsoid by a formula not too different from MacCullagh's. One corrects the measured values for height then the difference is the *anomaly* Δg. The mathematical idea is similar to the one we used to find Green's function: develop both sides of (9.7) in spherical harmonics and then equate coefficients. We make the approximation that the ellipsoid is a sphere locally. Alternatively one could simply adopt a spherical reference surface for the calculations, even though that is a poor approximation, the theory is perfectly valid. Then vertical and $\hat{\mathbf{r}}$ are the same. Let us repeat (9.7):

$$\Delta g = \frac{\partial \Delta U}{\partial r} + \frac{2\Delta U}{r} \,. \tag{1}$$

As usual we propose to expand Δg, which we considered an observed quantity on the sphere, in surface harmonics:

$$\Delta g = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} g_{lm} \, Y_l^m(\hat{\mathbf{s}}) \,. \tag{2}$$

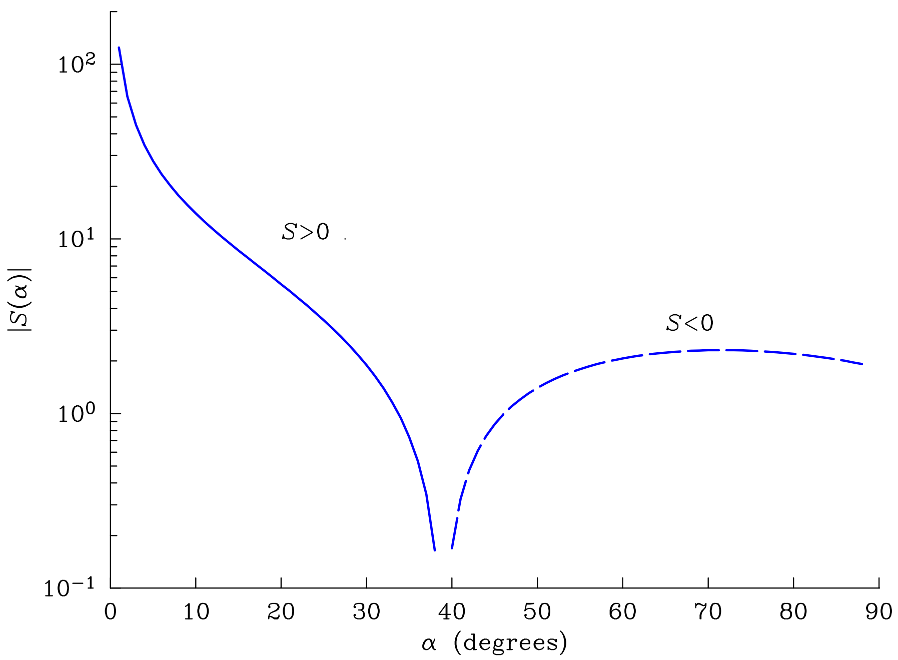


Figure 8

We start at $l=2$ because the $l=0$ term has been corrected out and the $l=1$ terms vanish identically because of the choice of origin. But on the other side if

$$\Delta U(\mathbf{r}) = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} u_{lm} \left(\frac{a}{r} \right)^{l+1} Y_l^m(\hat{\mathbf{r}}) \tag{3}$$

differentiating (3) and setting $r=a$

$$\Delta g = \left[\frac{\partial \Delta U}{\partial r} + \frac{2\Delta U}{r} \right]_{r=a} = -\frac{1}{a} \sum_{l=2}^{\infty} \sum_{m=-l}^{l} (l-1)\, u_{lm}\, Y_l^m(\hat{\mathbf{s}}) \,. \tag{4}$$

So we can equate coefficients in (2) and (4):

$$u_{lm} = -\frac{a\, g_{lm}}{l-1} \,. \tag{5}$$

Now we can insert this into (3), and the write the integral for g_{lm} and insert that too. Then the Addition Theorem applies just as in the previous Section and we are left with an integral called *Stokes'* formula

$$\Delta U(\mathbf{s}') = \int_{S(a)} d^2\mathbf{s}\; \Delta g(\mathbf{s})\; S(\mathbf{s}, \mathbf{s}') \tag{6}$$

where

$$4\pi a S = \operatorname{cosec} \tfrac{1}{2}\alpha + 1 - 6 \sin \tfrac{1}{2}\alpha - 5 \cos \alpha - 3 \cos \alpha \,\ln(\sin \tfrac{1}{2}\alpha + \sin^2 \tfrac{1}{2}\alpha) \tag{7}$$

with $\cos \alpha = \hat{\mathbf{s}} \cdot \hat{\mathbf{s}}'$. Of course we won't go in to the gory details of this one. The basic idea is that one can integrate a function against the local Δg values to obtain a geoid height. If the anomaly is very localized, it turns out the leading term is all one needs in (7) and S reduces to $1/(2\pi a\alpha)$; the flat-Earth approximation correspondingly would be $1/2\pi(x^2+y^2)^{1/2}$.

We shall see in a short while that in certain modern applications, it is possible to estimate N the geoid height itself directly, without having to go through g. In fact we then want to calculate the local Δg variations from N, reversing the Stokes analysis. If you attempt to do this as we have done here, you run into a sum over P_l that simply does not converge! The reason for this is that Δg is a differential of N. One way to fix up the analysis is to say that we choose to estimate gravity not from ΔN, but from $\nabla_1^2 N$:

$$\Delta g(\mathbf{s}') = \int_{S(a)} d^2\mathbf{s}\; \nabla_1^2 N(\mathbf{s})\; \tilde{S}(\mathbf{s}, \mathbf{s}')$$

where I leave it to you to evaluate the function $\tilde{S}$. In practice this analysis is carried out in the flat Earth approximation as we shall see, and it is performed in the spectral (that is Fourier) domain, where the bad behavior is brought under control in a different way.

16. Determination of the Geoid - III

If the ocean were still, its surface would have to be an equi-geopotential surface. In recent years the satellites Geosat and Seasat have been equipped with radar altimeters that are capable of measuring the satellite altitude above a mean ocean (in a several square kilometer 'foot print') to an accuracy of a few centimeters. The orbit of Geosat was arranged to repeat paths on the Earth exactly so that heights could be averaged in time to remove the effect of time-varying winds and currents. Given that the satellite position in space can be fixed to a meter or so, we have a very direct way of getting the geoid height over the ocean – measure it directly. The short wavelength information (scales less than 500 km) is mostly about submarine topography, since the geoid is lifted over topographic highs. At longer wavelengths the strength of the crust is such that the elevated areas are always supported at depth by low density material (this is *isostasy*) and so the features in the geoid no longer reflect surface processes in any obvious way. We will discuss isostasy again later. The gravity anomaly Δg is more directly connected with the presence of topography, so we would like a method that reverses the process described in the previous Section: converts geoid height into gravity anomaly. While this is almost straightforward from a spherical harmonic point of view, it is not very practical and we will now spend some time developing some theory for treating potential fields over small areas: the flat-Earth equivalent of spherical harmonics. We'll return to the actual geoid calculations when we have the theory under control.

But first let us compute the sorts of geoid distortion we might expect from marine topography. Let us look at a seamount: this is a submarine volcano of which there are literally tens of thousands all over the oceans, ranging in size from monsters that are 40 km at their base to 500 m hillocks. Let us approximate the seamount by a segment of a sphere. This gives us a chance to examine another handy trick of spherical harmonics – the calculation of the potentials of axially symmetric bodies. The idea is this: you can usually compute V on the axis of symmetry quite easily, so write this as power series in inverse powers of z the distance from some origin:

$$V(z) = \sum_{l=0}^{\infty} \frac{b_l}{z^{l+1}} \; . \tag{1}$$

But we know the axisymmetric spherical harmonic expansion has only $m=0$ terms in it. Choose the normalization that makes $N_{lm}=1$ and this means

$$V(R,\theta) = \sum_{l=0}^{\infty} \frac{c_{l0}}{R^{l+1}} P_l(\cos\theta) \; . \tag{2}$$

Along the axis $r=z$ and $\theta=0$, so that $P_l=1$. Thus we can identify the expansion coefficients in (2) with those in (1). This means we need only perform the calculation on the axis to get an expansion which works off axis. I performed this kind of calculation for a model seamount in the shape of a spherical cap, 2 km high, 20 km base diameter; the body stands in water 4 km deep and has a density of $2,400 \text{ kg m}^{-3}$. For details see Parker and Shure (*Geophys. J. Roy. Astr. Soc,* 80, p 631, 1988). Figure 9 shows the geoid and the gravity anomaly over the body, which is assumed to be uncompensated. The common units of gravity are milligals or mGal, where 1 mGal

$= 10^{-5}\,\mathrm{m\,s^{-2}}$. This gravity anomaly is quite large, and the elevation of the geoid is 50 cm. The contribution to gravity anomaly from the change in geoid height is $N\,\partial g/\partial r$ or $0.5 \times 3.08 \times 10^{-6}\,\mathrm{m\,s^{-2}}$ or only 0.15 mGal, confirming the observation we made earlier the at short wavelengths the geoid elevation effect is small compared with the direct attraction. Notice how the gravity anomaly mirrors the shape of the topography much more closely than the geoid height, which is a very smoothed version of the original and much lower in amplitude. We can see in this example how the Stokes' formula transforms a gravity anomaly into a geoid-height anomaly, although that is not how I computed N, of course.

Exercise

16.1 Consider the potential of a point mass on the $\hat{\mathbf{z}}$ axis at a distance r from the origin O. Find the spherical harmonic expansion relative to O for the potential for points (a) further away from O than the mass; (b) closer to O. Hence prove the generating formula for Legendre polynomials.

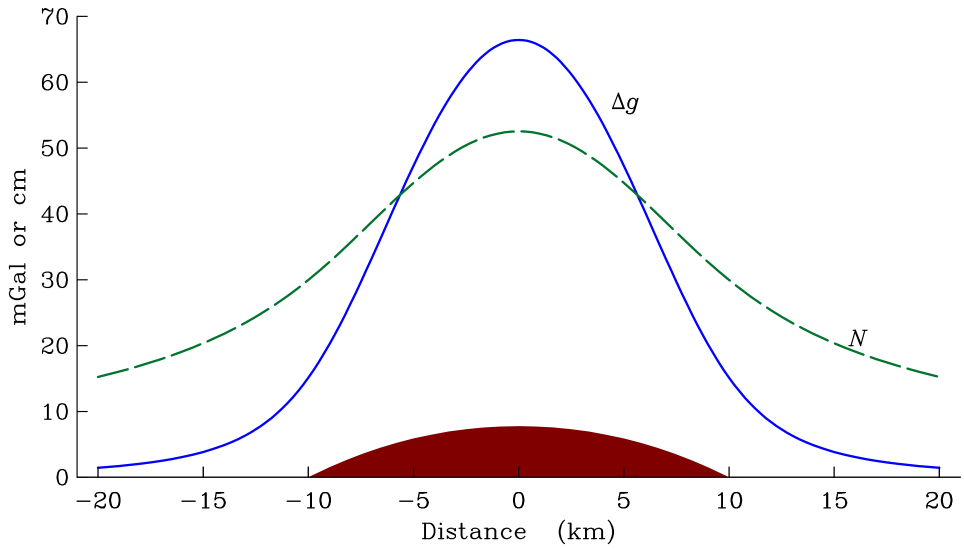


Figure 9

17. The Fourier Transform and Potential Theory

Suppose that the area we wish to study is sufficiently small that the Earth's curvature is unimportant – then we can use a flat-Earth approximation. Surprisingly, this is what the ocean-geoid people do as we shall see later. We could derive what we need from an analysis of high-degree spherical harmonics, but this requires asymptotics for differential equations and is rather messy. It is easier to start from scratch. We begin by looking at the analysis of a functions over the plane $\mathbb{R}^2$, essentially expanding them in eigenfunctions of the plane Laplacian. But in fact we are already familiar with this kind of decomposition: it is the 2-dimensional Fourier transform. We review the essential properties here. We shall assume that the functions $f(\mathbf{x})$ we wish to transform are smooth and die away rapidly (at least like $1/r$) at infinity. Notice this means some functions we might want to include are disallowed, like the constant or $x^2 - y^2$. Then the 2-dimensional *Fourier transform* is defined by

$$\hat{f}(\mathbf{k}) = \mathcal{F}[f](\mathbf{k}) = \int_{\mathbb{R}^2} d^2\mathbf{x}\; e^{-2\pi i \mathbf{k}\cdot\mathbf{x}}\, f(\mathbf{x}) \tag{1}$$

where $\mathbf{x}, \mathbf{k} \in \mathbb{R}^2$. The vector $\mathbf{k}$ is called the *wavevector*. The complex exponential $\exp(-2\pi i \mathbf{k}\cdot\mathbf{x})$ is a real and imaginary wave of unit amplitude varying along the direction of the wavevector (and constant perpendicular to that direction) and the wavelength is $1/|\mathbf{k}|$. The integral in (1) can be thought of as decomposing the function f into a spectrum of pure unidirectional waves, because we can reconstruct f with the *inverse transform*:

$$f(\mathbf{x}) = \mathcal{F}^{-1}[\hat{f}](\mathbf{x}) = \int_{\mathbb{R}^2} d^2\mathbf{k}\; e^{+2\pi i \mathbf{x}\cdot\mathbf{k}}\, \hat{f}(\mathbf{k})\,. \tag{2}$$

Notice the only difference between the Fourier transform and its inverse is a sign! There are a whole series of properties of the Fourier transform that make it very powerful. Here a few.

Parseval's Theorem:

$$\int_{\mathbb{R}^2} d^2\mathbf{x}\; f(\mathbf{x})\, g(\mathbf{x})^* = \int_{\mathbb{R}^2} d^2\mathbf{k}\; \hat{f}(\mathbf{k})\, \hat{g}(\mathbf{k})^* \tag{3}$$

from which follows the:

Power Theorem or *Plancherel's Theorem*

$$\int_{\mathbb{R}^2} d^2\mathbf{x}\; |f(\mathbf{x})|^2 = \int_{\mathbb{R}^2} d^2\mathbf{k}\; |\hat{f}(\mathbf{k})|^2\,. \tag{4}$$

The action of differentiation of a function has a very simple effect on its Fourier transform; in vector notation:

$$\mathcal{F}[\nabla f] = 2\pi i \mathbf{k}\, \mathcal{F}[f] \tag{5}$$

or, in Einstein summation notation:

$$\mathcal{F}[\partial f/\partial x_j] = 2\pi i k_j\, \mathcal{F}[f]\,. \tag{6}$$

Easily the most important property for numerical work is the *Convolution Theorem*. We define the *convolution* of two functions to be an integral

$$(f * g)(\mathbf{x}) = \int_{\mathbb{R}^2} d^2\mathbf{y}\, f(\mathbf{x}-\mathbf{y})\, g(\mathbf{y}) \,. \tag{7}$$

Convolution is symmetric with respect to the two functions: $f * g = g * f$; (why?). Although this looks complicated it is a very common operation; for example, application of Green's function on a plane is often in this form. Equation (7) is just the definition; here is the Convolution Theorem:

$$\mathcal{F}[f * g] = \hat{f}\, \hat{g} \,. \tag{8}$$

In words: the Fourier transform of a convolution is simply the product of the Fourier transforms of the two functions.

I have been very sloppy about the conditions under which these formulas apply; for those interested in seeing a mathematically rigorous statement of these results see D. C. Champeney, *A Handbook of Fourier Theorems,* Cambridge Univ. Press, 1987.

The reason why the Convolution Theorem is so important in practical calculations is the existence the *Fast Fourier Transform* or FFT. Suppose you have values of f within a rectangle sampled evenly on a lattice; one can imagine finding the transform (1) or (2) approximately by replacing the integral with a sum. If you want the same number of output points as you values, it is easy to see that you have to do on the order of N^2 calculations: N computations for each output point. If the number of nodes in each lattice dimension has a lot of small factors, and preferably is an exact power of 2, the FFT will do exactly same calculation in only $N \log_2 N$ calculations. This is huge gain in speed if N is large (say a million). Obviously we don't have time to describe this in detail – see Bracewell, *The Fourier Transform and its Applications*. Now suppose you have a convolution of numerical values to do: it will take on the order of N^2 calculations for that too. But by (7) we see that we can find the Fourier transform of the answer, by taking two FFTs. Then applying the inverse (2) is just another FFT. So three FFTs will do a convolution and that turns out to be an immense saving of computer time for large problems.

Let return to potential theory. We saw how by definition the surface harmonics were eigenfunctions of the surface Laplacian on a sphere $S(1)$. It is easy to verify that the complex exponential is an eigenfunction of the Laplacian on $\mathbb{R}^2$:

$$\nabla_2^2\, e^{2\pi i \mathbf{k}\cdot\mathbf{x}} = \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i} \exp(2\pi i\, k_j x_j) \tag{9}$$

$$= \frac{\partial}{\partial x_i}\, 2\pi i k_j \frac{\partial x_j}{\partial x_i} \exp(2\pi i\, k_j x_j) = 2\pi i k_j \delta_{ij} \frac{\partial}{\partial x_i} \exp(2\pi i\, k_j x_j) \tag{10}$$

$$= 2\pi i k_i \frac{\partial}{\partial x_i} \exp(2\pi i\, k_j x_j) = -\,4\pi^2 k_i k_i \exp(2\pi i\, k_j x_j) \tag{11}$$

$$= -\,4\pi^2\, |\mathbf{k}|^2\, e^{2\pi i \mathbf{k}\cdot\mathbf{x}} \,. \tag{12}$$

The odd thing here is that any negative number will do as an eigenvalue! Furthermore, you can see for each eigenvalue there are infinitely many eigenfunctions – the eigenvalues are not only continuous, but infinitely degenerate. Thus when we make an expansion we must use the integrals (1) for the coefficients and (2) for the expansion.

We shall proceed as we did for spherical harmonics. We look at a harmonic function in the domain $0 < z < H$, a cavity between two horizontal planes. We write Laplace's equation as

$$\nabla^2 = \frac{\partial^2}{\partial z^2} + \nabla_2^2 \,. \tag{13}$$

On each level within the cavity we write the potential as an inverse Fourier transform:

$$V(\mathbf{x}, z) = \int_{\mathbb{R}^2} d^2\mathbf{k}\; \hat{V}(\mathbf{k}, z)\, \mathrm{e}^{2\pi i \mathbf{x}\cdot\mathbf{k}} \,. \tag{14}$$

This is the same idea as in the spherical harmonic case but now it is a lot less general. Whereas *every* sufficiently smooth function over $S(1)$ has a spherical harmonic expansion, the class of functions with 2-dimensional Fourier transforms is much smaller: they have to decay at infinity and be reasonably well behaved in between. Classically, even the constant function is inadmissible. We should think of this solution as applying mainly to small perturbations in V. We will call these *benign potentials* without attempting to define them precisely. We now apply the 3-dimensional Laplacian to (14) and set the result to zero. Interchanging integration and differentiation and using (12) gives us:

$$0 = \int_{\mathbb{R}^2} d^2\mathbf{k} \left[\frac{d^2\hat{V}}{dz^2} - 4\pi^2 |\mathbf{k}|^2 \hat{V} \right] \mathrm{e}^{2\pi i \mathbf{x}\cdot\mathbf{k}} \,. \tag{15}$$

But the Fourier transform vanishes only if the function is itself zero, and this must be true on each level z. We conclude

$$\frac{d^2\hat{V}}{dz^2} - 4\pi^2 |\mathbf{k}|^2 \hat{V} = 0 \,. \tag{16}$$

This is an easy equation to solve: by inspection the most general solution of (16) is

$$\hat{V}(\mathbf{k}, z) = A(\mathbf{k})\mathrm{e}^{2\pi|\mathbf{k}|z} + B(\mathbf{k})\mathrm{e}^{-2\pi|\mathbf{k}|z} \,. \tag{17}$$

Placing this into (14) we get a Fourier representation in the cavity of the benign potentials:

$$V(\mathbf{x}, z) = \int_{\mathbb{R}^2} d^2\mathbf{k}\; [A(\mathbf{k})\mathrm{e}^{2\pi|\mathbf{k}|z} + B(\mathbf{k})\mathrm{e}^{-2\pi|\mathbf{k}|z}]\, \mathrm{e}^{2\pi i \mathbf{x}\cdot\mathbf{k}} \,. \tag{18}$$

The two terms A and B have the same significance as in the spherical harmonic discussion: A is generated by sources above the cavity in $z > H$, while B belongs to potentials whose sources lie below the cavity in $z < 0$. It is most common of course for local sources to be underground, caused by density variations and topography etc; then A is absent. Normally therefore, (18) appears as

$$V(\mathbf{x}, z) = \int_{\mathbb{R}^2} d^2\mathbf{k}\ B(\mathbf{k}) e^{-2\pi|\mathbf{k}|z}\ e^{2\pi i \mathbf{x}\cdot\mathbf{k}}\ . \tag{19}$$

Finding the function B in this case is easy if V is known on any horizontal plane. For example, if we have V on $z = 0$ we put this value into (19) and inverse transform:

$$\hat{V}(\mathbf{k}, 0) = B(\mathbf{k}) = \int_{\mathbb{R}^2} d^2\mathbf{x}\ V(\mathbf{x}, 0)\ e^{-2\pi i \mathbf{x}\cdot\mathbf{k}} \tag{20}$$

Referring to (17) we see that the transforms at different levels are simply related:

$$\hat{V}(\mathbf{k}, z) = \hat{V}(\mathbf{k}, 0)\, e^{-2\pi|\mathbf{k}|z} \tag{21}$$

Equation (19) has a very simple interpretation. An interior gravity potential is thought of as composed of a set of sine waves at $z = 0$, which are added together to make the signal. Each sinusoidal potential with wavelength $\lambda = 1/|\mathbf{k}|$ is attenuated by the factor $\exp(-2\pi z/\lambda)$ at the height z above the ground. That is a big factor: at one wavelength altitude, the signal is reduced by a factor of 535. Thus satellites flying at altitudes of 200 km above the Earth hardly feel geoid variations of wavelength as short as than 200 km, which is why the shorter wavelength geoid is not found by analysis of the orbits. Such exponential attenuation is a feature of upward continuation in a flat-Earth geometry.

Exercise
17.1 A point mass m is buried at a depth h below the origin. Find $B(\mathbf{k})$ in (19) for the potential V observed on the plane $z = 0$.

18. Free-Air Gravity and Bouguer Corrections

On land one still makes gravity surveys the old-fashioned way: usually with a LaCoste-Romberg meter at a series of surveyed and leveled sights. The gravity field locally gives some clue about buried density contrasts and is used in mineral exploration. The biggest signal in many areas is simply topographic, from the fact that the meter is confined to be on the ground, which has varying elevation. To remove this effect, one simply subtracts the so-called *free air* term, the correction resulting from the main vertical gradient, roughly 0.3 $\mathrm{mGal\,m^{-1}}$. Then as we saw there is a fairly large variation of gravity with latitude from the J_2 term and the rotational term; this is removed in a 'standard' gravity formula. The resultant gravity is called the free-air gravity, reduced to sea level (or equivalently, the geoid). For ocean surveys, the elevation correction is of course unnecessary.

Even after these corrections, land gravity correlates with topography for the obvious reason that on a mountain there is more mass attracting downwards between you and sea level than on a plane at a lower elevation. The same is true at sea: rock is more dense than water, and of course gravity is increased over high topography just as we saw with our seamount calculation. To get a clue about underground density variations the next correction is to try to remove the attraction of the topographic features, assuming some kind of simple density model, usually a constant number. Because slopes are usually gentle, a first very simple approximation is to remove the attraction of horizontal slab lying between the observation site and sea level: the correction is

$$\Delta g_S = 2\pi G \rho h \tag{1}$$

is called the *Bouguer* slab correction and the resultant gravity field is a *Bouguer anomaly*. This result was proved in Section 14. If one does a better job of approximating the shape of the topography, this is called *terrain correction*. What is discovered, as I have hinted before, is that the Bouguer anomaly is consistently negative in mountainous regions. This is isostasy at work: light material at depth buoys up the elevated regions. By correlating Bouguer anomaly with elevation it is possible to get estimates of the elastic strength of the crust over the continents.

Exercises

18.1 Show in the flat Earth approximation that with z positive downward

$$\Delta\hat{g}(\mathbf{k}) = 2\pi G \int_0^\infty dz\, \Delta\hat{\rho}(\mathbf{k}, z)\, \mathrm{e}^{-2\pi|\mathbf{k}|z}$$

where $\Delta\hat{g}$ is the Fourier transform of the gravity anomaly at the surface $z=0$ due to density anomalies $\Delta\rho$; $\Delta\hat{\rho}$ is the horizontal Fourier transform of the density contrast at depth z. layer, thickness z_0 and they are locally isostatically

18.2a If an uplifted region is isostatically compensated, we can approximate its gravitational effect by that of a surface layer of vertical mass dipoles. Suppose the Earth is covered with such a layer of moment density $\mu(\hat{\mathbf{s}})$. By decomposing the density into its spherical harmonic series, calculate ΔU, the change in the geopotential immediately above the layer, and hence find N the change in geoid height caused by layer. You may leave your answer in terms of a sum and treat the Earth as a sphere for these purposes.

18.2b From your result show that the change in geopotential for a layer with μ = constant is zero. Show that in the limit of short wavelength, your result agrees with that predicted by the theory for an infinite flat layer in Section 14.

18.3a For an isostatically compensated layer of thickness h, show that the mass dipole density is given by

$$\mu = \int_0^h \Delta\rho(z)\, z\, dz$$

where $\Delta\rho(z)$ is the density contrast with a standard section and z is positive upwards. According to the flat-Earth approximation of Section 14, how much is the geoid level raised by a broad layer of this kind?

18.3b A wide plateau stands 2000 m higher than an adjacent continental shield; local compensation takes place at the moho, which is 40 km deep under the plateau and 30 km under the shield. If the crustal density of both provinces is 2700 $\mathrm{kg\,m^{-3}}$ calculate the density contrast between crust and mantle. Find the equivalent mass dipole density and hence calculate the geoid elevation caused by the plateau. Use the flat-Earth theory, and you may assume the crust and mantle are both uniform in density.

19. A Fourier Terrain Corrector

As another application of Fourier techniques, I want to show you a method of computing the terrain correction used for marine work (invented by me in 1973). Suppose the topographic height above some mean seafloor level at $z=0$ is given by the function $h(\mathbf{s})$ with $\mathbf{s} \in \mathbb{R}^2$, and the observer is at the constant level $z=z_0$. We want to compute the 2-D Fourier transform of the gravitational attraction on the observer plane. If we can write this in terms of Fourier transforms, this would be a faster way of computing the terrain correction than doing the integrals by summing contributions, if you recall our earlier discussion. We find the potential, and assume a constant density ρ:

$$-\frac{V(\mathbf{r}+z_0\hat{\mathbf{z}})}{G\rho} = \int_{\mathbb{R}^2} d^2\mathbf{s} \int_0^{h(\mathbf{s})} \frac{dz}{|\mathbf{r}+(z-z_0)\hat{\mathbf{z}}-\mathbf{s}|} \ . \tag{1}$$

We take the Fourier transform with respect to $\mathbf{r} \in \mathbb{R}^2$:

$$-\frac{\hat{V}(\mathbf{k})}{G\rho} = \int_{\mathbb{R}^2} d^2\mathbf{r}\, e^{-2\pi i\mathbf{k}\cdot\mathbf{r}} \int_{\mathbb{R}^2} d^2\mathbf{s} \int_0^{h(\mathbf{s})} \frac{dz}{|\mathbf{r}+(z_0-z)\hat{\mathbf{z}}-\mathbf{s}|} \tag{2}$$

$$= \int_{\mathbb{R}^2} d^2\mathbf{s} \int_0^{h(\mathbf{s})} dz \int_{\mathbb{R}^2} d^2\mathbf{r}\, e^{-2\pi i\mathbf{k}\cdot\mathbf{r}} \frac{1}{|\mathbf{r}+(z_0-z)\hat{\mathbf{z}}-\mathbf{s}|} \ . \tag{3}$$

The last integral we recognize as being the Fourier transform of the potential of a point mass, (z_0-z) below the plane and at $\mathbf{s}$, the solution to the Exercise 17.1, almost. We make a small adjustment and plug this in:

$$-\frac{\hat{V}(\mathbf{k})}{G\rho} = \int_{\mathbb{R}^2} d^2\mathbf{s}\, e^{-2\pi i\mathbf{k}\cdot\mathbf{s}} \int_0^{h(\mathbf{s})} dz \int_{\mathbb{R}^2} d^2\mathbf{r}\, e^{-2\pi i\mathbf{k}\cdot(\mathbf{r}-\mathbf{s})} \frac{1}{|+(z_0-z)\hat{\mathbf{z}}|} \tag{4}$$

$$= \int_{\mathbb{R}^2} d^2\mathbf{s}\, e^{-2\pi i\mathbf{k}\cdot\mathbf{s}} \int_0^{h(\mathbf{s})} dz \, \frac{e^{-2\pi(z_0-z)|\mathbf{k}|}}{|\mathbf{k}|} \tag{5}$$

$$= \int_{\mathbb{R}^2} d^2\mathbf{s}\, e^{-2\pi i\mathbf{k}\cdot\mathbf{s}}\, e^{-2\pi|\mathbf{k}|z_0} \frac{e^{2\pi|\mathbf{k}|h(\mathbf{s})}-1}{2\pi|\mathbf{k}|^2} \ . \tag{6}$$

This is not a Fourier transform yet. But if we expand the right-most exponential in its Taylor series we get:

$$-\frac{\hat{V}(\mathbf{k})}{G\rho} = 2\pi\, e^{-2\pi|\mathbf{k}|z_0} \int_{\mathbb{R}^2} d^2\mathbf{s}\, e^{-2\pi i\mathbf{k}\cdot\mathbf{s}} \sum_{n=1}^{\infty} \frac{(2\pi|\mathbf{k}|)^{n-2} h(\mathbf{s})^n}{n!} \tag{7}$$

$$= 2\pi\, e^{-2\pi|\mathbf{k}|z_0} \sum_{n=1}^{\infty} \frac{(2\pi|\mathbf{k}|)^{n-2}}{n!} \mathcal{F}[h^n] \tag{8}$$

which is a sum over a series of Fourier transforms of successive powers of the topography h. To find the gravity anomaly we just use (17.21):

$$\Delta\hat{g}(\mathbf{k}) = 2\pi G\rho\, e^{-2\pi|\mathbf{k}|z_0} \sum_{n=1}^{\infty} \frac{(2\pi|\mathbf{k}|)^{n-1}}{n!} \mathcal{F}[h^n] \tag{9}$$

and invert. The method has been universally adopted for marine gravity and magnetics analysis.

20. Isostasy and the Strength of the Lithosphere

We have already mentioned that the idea behind isostasy is that the Earth has little strength over long periods. If there is a raised region, like a mountain chain, this is not a pile of rock resting on a rigid foundation, but the visible result of a low density root that allows the upper portion float higher above the geoid. Seismic studies support the theory of Airy, that the upper surface of the mountain is mirrored by a discontinuity (the Moho) where the low density crust penetrates more deeply into the high-density mantle. The Airy mechanism is the example of *local* compensation, in which the low density supporting the topography at any point can be found immediately under the point, and not to the side. An alternative local mechanism was given by Pratt, in which the density of each column under a topographic high was reduced throughout its length, down to the (constant) compensation depth. Before the seismic studies, geologists believed the Pratt mechanism was a good approximation. If, on the other hand, the Earth had strength and could elastically support a mountain, you can see that the compensation of a compact load would be spread to the sides – *regional compensation*. We shall return to this idea in a short while.

Suppose we believe in local compensation, then we can compute the density structure responsible by doing a gravity and topographic survey of a region. This is the idea of Dorman and Lewis (*JGR,* 75, p 3357, 1970); it is a nice application of Fourier methods and is used today to learn about the strength of the oceanic lithosphere as we shall see. We make the (at first sight) rather strange hypothesis that for topographic elevations $h(\mathbf{s})$ above the geoid the density change directly under the point is given by

$$\Delta\rho(\mathbf{s}, z) = \sigma(z)\, h(\mathbf{s}) \tag{1}$$

where z is a depth coordinate immediately below the elevated point. A little thought will show that the Airy compensation mechanism gives a spike for σ at the mean Moho depth. The Pratt mechanism gives a flat function, down to the depth of compensation. Obviously it is a good idea to determine σ from observations if we can.

Let us compute the gravitational perturbation from the compensating density, but instead of doing this directly we find the Fourier transform. Of course we assume the region is small enough that a flat-Earth approximation holds. At a depth z we will find a sheet of matter, thickness dz, with density change given by (1). To a first approximation the gravity field just above the sheet will be given by the slab formula:

$$dg(\mathbf{s}, z) = 2\pi G \Delta\rho\; dz = 2\pi G\, \sigma(z)\, h(\mathbf{s})\; dz\ . \tag{2}$$

To get the surface value we must upward continue this signal. In the flat Earth world $g = \partial U / \partial z$, which is a harmonic function. We Fourier transform (2) on the horizontal variable $\mathbf{s}$:

$$d\hat{g}(\mathbf{k}, z) = 2\pi G\, \sigma(z)\, \hat{h}(\mathbf{k})\; dz \tag{3}$$

which represents a signal with wavenumber k at depth z. At the surface this signal is attenuated by the factor $e^{-2\pi |\mathbf{k}| z}$. Now all we do is integrate through the layers vertically:

$$\Delta\hat{g}_B(\mathbf{k}) = \int_0^\infty dz\ 2\pi G\ \sigma(z)\ \hat{h}(\mathbf{k}) e^{-2\pi|\mathbf{k}|z} \tag{4}$$

$$= \hat{R}(\mathbf{k})\ \hat{h}(\mathbf{k}) \tag{5}$$

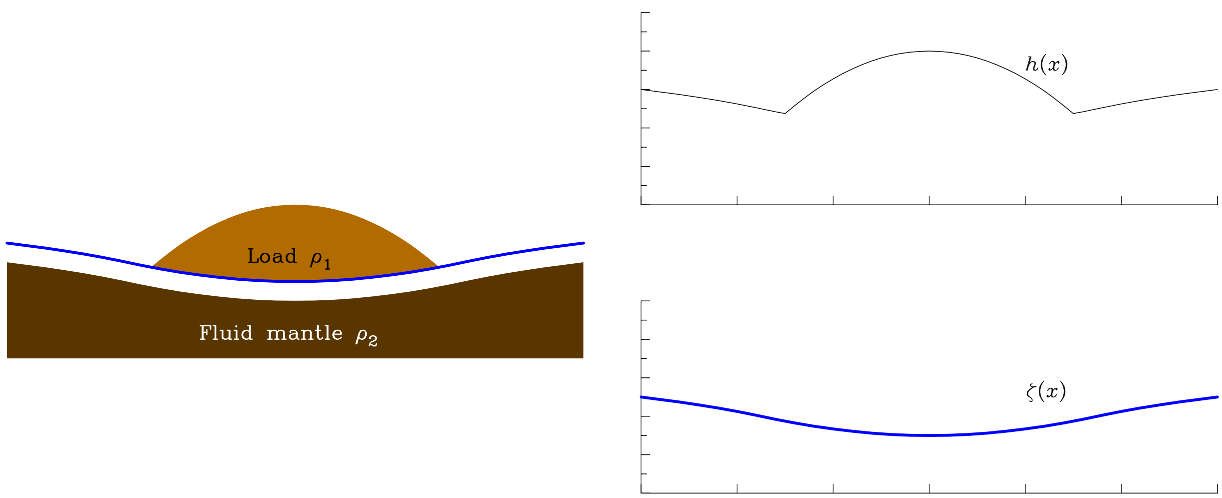
where $\hat{R}$ is called the *isostatic response function* of the lithosphere.

The nice thing about (5) is that in principle we can measure h and Δg_B, the Bouguer-corrected gravity anomaly, so that we can compute their transforms and hence get $\hat{R}$. Thus we can get an experimental estimate for the response, and see if it matches expectations. In general we expect $\hat{R}$ to be axisymmetric unless there is a strong fabric in the geology. We expect it to be a decreasing function of k because density signals at depth are more weakly felt at shorter wavelengths (higher wavenumbers). These things are indeed seen. But when we come to try to model $\Delta\rho$ things are less satisfactory. For example, when Dorman and Lewis did this for the USA they found positive and negative densities below each point. The general consensus about this observation was that it was most unlikely, and that instead the assumption of local compensation was simply wrong. It seems there ought to be useful information in the response function, but how to get at it?

We need a new model. This one involves some regional compensation, in the form of elastic support from strength in the crust. What we get now is information about crustal strengths. We begin with a simple elastic model with some free parameters; we ask for the deformation of the simple model from a load, and then we ask for the Bouguer anomaly generated by the deformation. Everything is done in the Fourier domain. It turns out that, even though the model is obviously of regional compensation, it predicts a response function in the form of (5) which can be checked in the same way as the local compensation model could.

We imagine an elastic layer, bent down by a load. Elasticity theory says that the equation for the deformation of the layer, if it is thin and the deformation is not too great, is:

Figure 10



$$D\,\nabla^4\zeta = p \tag{6}$$

where ζ is the upward displacement and p the applied upward pressure; D is called the flexural rigidity of the plate and is given in terms of more common elastic parameters by:

$$D = E\,T^3/12(1-s^2) \tag{7}$$

where T is the thickness, s is Poisson's ratio and E Young's modulus for the material. In our model the pressure can be related to the sum of two effects, the buoyancy force from the mantle below, and the weight of the load:

$$p = -\,\rho_1\,gh_0 - \rho_2\,g\zeta\ . \tag{8}$$

(The signs seem strange because positive ζ is upwards.) The measured topography h is the sum of ζ, the upward displacement of the layer, and h_0, the height of the load itself:

$$h = h_0 + \zeta\ . \tag{9}$$

We eliminate the h_0, which cannot be observed, between (8) and (9) to get

$$D\nabla^4\zeta + g(\rho_2 - \rho_1)\zeta = -\,\rho_1\,gh\ . \tag{10}$$

We see on the left there are two forces supporting the load: the first is an elastic term and second a buoyancy term. We solve this equation by taking the Fourier transform. Just as the FT of ∇^2 was $-4\pi k^2$, so that of ∇^4 is its square, namely $16\pi^4k^4$; so we get

$$[16\pi^4k^4D + g(\rho_2 - \rho_1)]\hat{\zeta}(\mathbf{k}) = -\,\rho_1 g\,\hat{h}(\mathbf{k})\ . \tag{11}$$

Next suppose the gravity anomaly is generated by the density differences that arise from pushing the material density ρ_1 into the mantle density ρ_2 and that this takes place at a depth z_1. Then it is easily seen by the arguments already given that the FT of the Bouguer anomaly is simply:

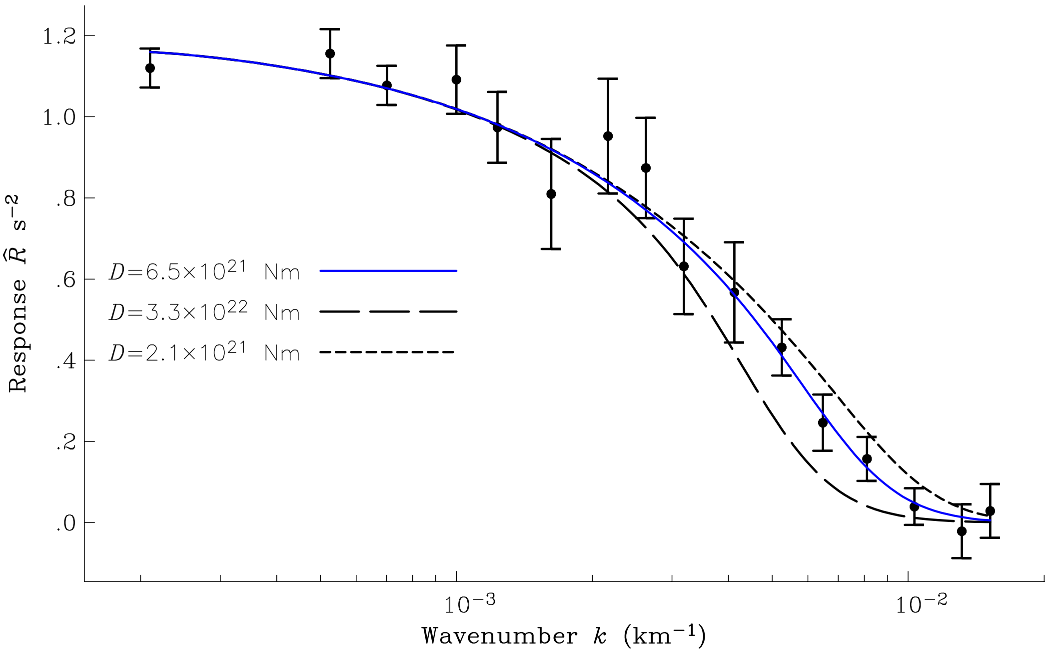
$$\Delta\hat{g}_B(\mathbf{k}) = 2\pi G(\rho_2 - \rho_1)\,\mathrm{e}^{-2\pi|\mathbf{k}|z_1}\,\hat{\zeta}(\mathbf{k})\ . \tag{12}$$

Now we just eliminate $\hat{\zeta}$ between these two equations and

$$\hat{R}(\mathbf{k}) = \frac{\Delta\hat{g}_B(\mathbf{k})}{\hat{h}(\mathbf{k})} = \frac{2\pi\rho_1\,gG(\rho_2 - \rho_1)\,\mathrm{e}^{-2\pi|\mathbf{k}|z_1}}{16\pi^4|\mathbf{k}|^4D + g(\rho_2 - \rho_1)}\ . \tag{13}$$

Suitable choices of the various free parameters give great fits to the data. See Banks et al. (*Geophys. J. Roy. Astr. Soc.,* 54, p 431, 1977). Most of the parameters can be estimated rather well, with the exception of D, the flexural rigidity, which corresponds to the strength of the lithosphere. If we plot $\hat{R}$ for various values of D we can see how this parameter controls the transition of the response function from local to region compensation in the wavenumber, or length-scale, domain. Figure 11 shows the response estimates for the continental United States, and how varying the rigidity parameter alters the model response. In view of the uncertainties, the fit is excellent. How would you expect the response curve to look for Oceanic crust? For a largely Precambrian continent, like Australia?

Figure 11



Isostasy is still a hot topic in geophysics. See, for example, the paper by K. M. Fischer, *Nature*, 417, pp 933-6, 2002. The author shows evidence that low-density roots of mountain chains that have been eroded away do not entirely disappear; so there are fossil roots under some long-vanished mountains.