3. GAUSS' THEORY OF THE MAIN FIELD

This chapter deals with the geomagnetic field in the static approximation: that is we limit our attention to an instant in time and consider the problem of how to look at the structure of the field. We begin by thinking of Earth as a spherical body of radius a surrounded by an insulating atmosphere extending out to radius b; we call the region of space lying between the radii $a < r < b$, the shell $S(a, b)$. In this approximation the magnetic field can be considered the gradient of a scalar potential that satisfies Laplace's equation everywhere outside the source region.

3:1 Gauss' Separation of Harmonic Fields into Parts of Internal and External Origin

Let us suppose that in the atmospheric cavity $S(a, b)$ the magnetic field is derived from a scalar potential, thus:

$$\mathbf{B} = -\nabla\Psi \tag{31}$$

and that Ψ is harmonic

$$\nabla^2\Psi = 0. \tag{32}$$

From equation (I-8.7, p.31 of gravity notes) we know that Ψ, the solution to Laplace's equation, has a representation in terms of spherical harmonics

$$\Psi(r, \theta, \phi) = a \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \left[A_l^m \left(\frac{r}{a}\right)^l + B_l^m \left(\frac{a}{r}\right)^{l+1} \right] Y_l^m(\theta, \phi). \tag{33}$$

The coefficients A_l^m characterize fields of external origin and B_l^m describe those of internal origin. Gauss (in 1832) was the first to attempt to estimate the sizes of the internal and external contributions to the geomagnetic field, thereby demonstrating the predominance of the internal part. We will show here one means by which the separation into internal and external parts can in principle be performed.

But first, why does the expansion begin at $l = 1$ and not $l = 0$? The answer is that the Maxwell equation $\nabla \cdot B = 0$ rules out sources that look like monopoles, that is with $1/r$ behavior, and so $B_0^0 = 0$; A_0^0, which leads to a constant term in the potential, can have any value, so zero is fine. Hence we can exclude the $l = 0$ term for magnetic fields.

If we knew Ψ on two spherical surfaces we could get two independent equations for each of the Ψ_l^m involving the coefficients A_l^m and B_l^m and do the separation that way. However, in making observations of the geomagnetic field we do not measure Ψ, but $\mathbf{B}$. Let's assume that $\mathbf{B}$ is known everywhere on the surface of the sphere $r = a$, and we write

$$\mathbf{B} = \hat{\mathbf{r}} B_r + \mathbf{B}_s \tag{34}$$

where $\hat{\mathbf{r}} \cdot \mathbf{B}_s = 0$, in other words $\mathbf{B}_s$ is a tangent vector field on the surface of the sphere. Earlier (I-7.2) we defined the surface gradient ∇_1 as follows

$$\nabla = \hat{\mathbf{r}}\partial_r + \frac{1}{r}\nabla_1. \tag{35}$$

We first find B_r on $r = a$ from Ψ:

$$B_r = -\partial_r \Psi|_{r=a} = -\sum_{l,m} [l A_l^m \ - \ (l \ + \ 1) B_l^m] Y_l^m(\theta, \phi). \tag{36}$$

Now we recall from the table of spherical harmonic lore (Grav. p. 29) that if

$$f(\theta, \phi) = \sum_{l,m} c_l^m Y_l^m(\theta, \phi) \tag{37}$$

then using the orthonormality of the Y_l^m, that is, by

$$\int_{S(1)} Y_l^m(\theta, \phi) Y_{l'}^{m'}(\theta, \phi)^* d^2\hat{\mathbf{r}} = \delta_l^{l'} \delta_m^{m'} \tag{38}$$

the coefficients c_l^m are just

$$c_l^m = \int f(\theta, \phi) (Y_l^m)^* d^2\hat{\mathbf{r}}. \tag{39}$$

From this it follows that

$$l A_l^m \ - \ (l \ + \ 1) B_l^m = -\int B_r (Y_l^m)^* d^2\hat{\mathbf{r}}. \tag{40}$$

We can thus find the above linear combination of A_l^m and B_l^m from knowledge of B_r but to find each of them separately we need another equation. The obvious solution is to use the tangential field $\mathbf{B}_s$. Again using the spherical harmonic expansion for Ψ on $r = a$ we find

$$\mathbf{B}_s = -\sum_{l,m} [A_l^m \ + \ B_l^m] \nabla_1 Y_l^m. \tag{41}$$

Now we invoke an orthogonality relation for the $\nabla_1 Y_l^m$, namely (I-7.37)

$$\int_{S(1)} \nabla_1 Y_l^m \cdot (\nabla_1 Y_{l'}^{m'})^* d^2\hat{\mathbf{r}} = l(l \ + \ 1) \delta_l^{l'} \delta_m^{m'}. \tag{42}$$

Dot $(\nabla_1 Y_{l'}^{m'})$ into our equation for $\mathbf{B}_s$ on $r = a$, and integrate over the sphere:

$$\int \mathbf{B}_s \cdot (\nabla_1 Y_{l'}^{m'})^* d^2\hat{\mathbf{r}} = -\sum_{l,m} [A_l^m \ + \ B_l^m] \int \nabla_1 Y_l^m \cdot (\nabla_1 Y_{l'}^{m'})^* d^2\hat{\mathbf{r}}. \tag{43}$$

Thus we have

$$A_{l'}^{m'} \ + \ B_{l'}^{m'} = -\frac{\int \mathbf{B}_s \cdot (\nabla_1 Y_{l'}^{m'})^* d^2\hat{\mathbf{r}}}{l'(l' \ + \ 1)}. \tag{44}$$

Combining this with the equation derived from B_r we can always recover the A_l^m and B_l^m separately from our knowledge of $\mathbf{B}$ on $r = a$, except for $l = 0$ (Why?). Thus Ψ is determined to within an additive constant and $\mathbf{B}$ is determined uniquely within $S(a, b)$, by our knowledge of $\mathbf{B}$ on $S(a)$. It is important to keep in mind that equations (40) and (44) are theoretical results, not very useful in practice because they require knowledge of $\mathbf{B}$ all over the surface of the Earth. As we shall see later the estimation of the A_l^m and B_l^m from actual magnetic field measurements does not rely on knowing $\mathbf{B}$ everywhere, but makes use of traditional statistical estimation techniques and geophysical inverse theory.

None-the-less Gauss used his theory to answer the question of the origin of the magnetic field. He discovered to the accuracy available at the time that the external coefficients A_l^m were all negligible, so that Gilbert's idea that the Earth is a great magnet was shown to be essentially correct. We know today there are fields of external origin but they are usually three or more orders of magnitude smaller than the internal fields.

3:2 Upward Continuation

We will now specialize our interest in the geomagnetic field further and consider only those parts of internal origin. Suppose that we have a collection of observations on one surface, but would like to infer something about the source at some other altitude or radius; for example, we might want to study crustal sources from satellite observations or take measurements at Earth's surface and use them to study what's going on at the core-mantle boundary. The idea of upward continuation of a harmonic field has been touched on before in Part I, sections 16, and 19, but the reverse process, downward continuation, does not play much of role in gravity, in contrast to its very great importance in geomagnetism.

We will use as an example the case where we know $\mathbf{B} \cdot \hat{\mathbf{r}}$ everywhere on a sphere $S(a)$ containing the sources (**J**, **M**, current and magnetization). If we are prepared to assume that Earth's mantle is an insulator and there are no magnetic sources within it (approximations commonly adopted when studying the magnetic field at the core) then we can write the magnetic field **B** as the gradient of a scalar potential within that region too.

$$\mathbf{B} = -\nabla\Psi \tag{45}$$

and

$$\nabla^2\Psi = 0. \tag{46}$$

> *Exercise:*
> The radial component of the magnetic field is $B_r = \hat{\mathbf{r}} \cdot \mathbf{B}$. Show that $\nabla^2(rB_r) = 0$, that is rB_r is also harmonic outside of $S(a)$.

Using the result of the above exercise, we can define a potential function $\Omega = rB_r$ which is harmonic. Now we have an example of the exterior Dirichlet boundary value problem for a sphere (Part I, section 13) and for $r > a$, we can write Ω in terms of a spherical harmonic expansion

$$\Omega(r, \theta, \phi) = rB_r(r, \theta, \phi) = a \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \beta_l^m \left(\frac{a}{r}\right)^{l+1} Y_l^m(\theta, \phi). \tag{47}$$

If we know rB_r on $S(a)$ we can use the orthogonality of the Y_l^m in the usual way to get

$$\beta_l^m = \int_{S(a)} B_r(a, \theta', \phi') Y_l^m(\theta', \phi')^* \frac{d^2r'}{a^2}. \tag{48}$$

Knowing the β_l^m we can find B_r anywhere (47) converges.

Finding B_r further away from the sources is known as *upward continuation*. One way to do this is to apply (47) directly. We imagine performing the integrals (48) on the known field over $S(a)$. Let us look at the function Ω on sets of spheres of constant radius. In the rest of this section we consider upward continuation to be a mapping of a function defined on $S(a)$ onto another function on $S(r)$. On $S(r)$, a sphere of radius r, we will say the function Ω has a surface harmonic expansion $\Omega_l^m(r)$. It is clear from (47) that

$$\Omega_l^m(r) = \beta_l^m \left(\frac{a}{r}\right)^{l+1} . \tag{49}$$

When $r > a$ we see that the magnitude of the a given harmonic is more strongly attenuated the shorter the wavelength of the harmonic (recall Jean's rule: $\lambda \approx 2\pi r/(l + \frac{1}{2})$; so short-wavelength energy disappears from the field preferentially as we go to spheres of larger radius.

In Part I, section 13, we derived another way of finding a harmonic function with internal sources from values given on a sphere: here is (I-13.7) again, rewritten in terms of B_r

$$B_r(r, \hat{\mathbf{r}}) = \int_{S(a)} (a/r)^2 K(a/r,\ \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}')\, B_r(a, \hat{\mathbf{r}}')\, d^2\hat{r}' \tag{50}$$

where

$$K(x, \cos\theta) = \frac{1}{4\pi} \frac{1 - x^2}{(1 + x^2 - 2x\cos\theta)^{3/2}} - 1 \tag{51}$$

and this function is shown in the figure. Note that the constant term comes from the need to subtract the influence of the monopole in the generating function for Legendre polynomials.

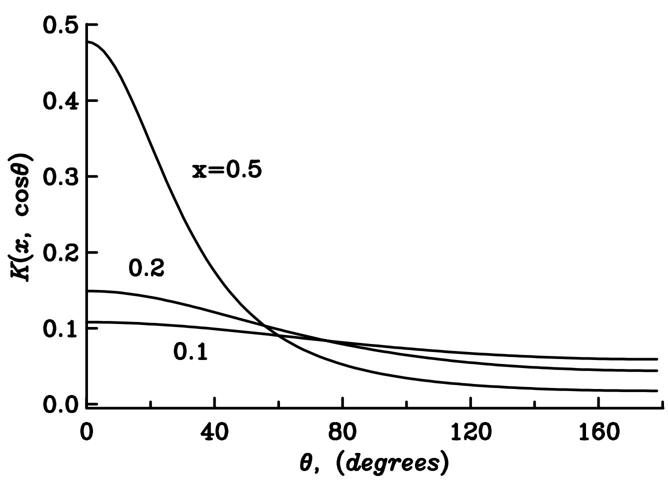

Figure 3.2.1

We can use the integral to illustrate another diminishing property of upward continued fields, not obvious from (49). A fundamental property of integrals is that

$$|B_r(r, \hat{\mathbf{r}})| \leq \int_{S(a)} |(a/r)^2 K(a/r, \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}')|\ |B_r(a, \hat{\mathbf{r}}')|_{max}\ d^2\hat{r}'. \tag{52}$$

Since the term in B_r under the integral is constant, and the function K is positive we can evaluate the right side, with $x = a/r$:

$$\int_{S(a)} |(a/r)^2 K(a/r, \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}')| \ d^2\hat{r}' = \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta \ \frac{1}{4\pi} \left[\frac{x^2(1-x^2)}{(1+x^2-2x\cos\theta)^{3/2}} - x^2 \right]$$
$$= x^2 = (a/r)^2. \tag{53}$$

It is an exercise for the student to evaluate the integral! From (53) it follows that

$$|B_r(r, \hat{\mathbf{r}})| < |B_r(a, \hat{\mathbf{r}}')|_{max}, \quad r > a. \tag{54}$$

In words, the magnitude of B_r on a sphere of radius $r > a$, is always less than the maximum magnitude on the sphere of radius a; in fact, (53) implies a stronger result, that the maximum value falls off like r^{-2}. Technically upward continuation is bounded linear mapping.

A consequence of (54) and of (49) is that considered as a mapping from one sphere to another larger sphere, upward continuation is stable. This means loosely that a small error in the field on the inner sphere remains small on the outer one.

This is important when we consider practical upward continuation. Suppose we have the true field on $S(a)$ corrupted by measurement error: $B_r + e$; we upward continue this inaccurate field to $S(r)$, but because we can imagine upward continuing B_r and e separately (by applying (50) to each part) we see that if the maximum error on $S(a)$ is less than some number $\eta > 0$, then (54) assures us the maximum error on $S(r)$ will also be less than η, indeed less than $\eta(a/r)^2$.

3:3 Downward Continuation

The process of upward continuation has some practical uses, for example: predicting magnetic anomalies measured at the sea surface from values obtained near the seafloor, or using ground-based surveys to predict the magnetic field at satellite altitudes. But it is more common to want to project field values towards their sources; as we have already mentioned, from the earth's surface towards the core, or from satellite orbits down to the crust.

The ideas of the previous section work up to a point. If there are no sources in the region we want to downward continue through then (45) and (46) are certainly valid. We can use (47) and (48) to represent the field, and provided the series still converges on $S(r)$, we can use (47) even if $r < a$. But (49) shows that now the shorter wavelength components of the field are magnified relative to the longer wavelength ones. There is no integral formula like (50) because, in its derivation, you will find the required series diverges, and so it is impossible to interchange the sum and integral.

In fact the mapping from $S(a)$ to $S(r)$ when $r < a$ is an example of an unstable process. Roughly this means small errors in the field may grow when the field is downward continued. We saw in the upward continuation process that simply knowing that the magnitude of the measurement error everywhere on $S(a)$ was less than η guaranteed the error in the upward continued field would be smaller than this. That result is no longer true in downward continuation, and is at the heart of the instability.

We illustrate the growth of error by the following simple example. As before we have a field plus error $B_r + e$ on $S(a)$ and $|e| \leq \eta$ everywhere on the sphere. Now we downward continue to the surface $S(r)$, with $r < a$. I will assert that here e can be arbitrarily large!. How can that be? Suppose the error term has a spherical harmonic expansion comprising the single term:

$$\begin{aligned} re_r(a, \theta, \phi) &= \alpha Y_l^{-l}(\theta, \phi) \\ &= \alpha N_{l,l} \sin^l \theta \mathrm{e}^{-\mathrm{i}l\phi} \end{aligned} \tag{55}$$

where $N_{l,l}$ is a normalizing factor. Then $\eta = |\alpha N_{l,l}|$. From (49) we know that on $S(r)$ e has the single coefficient in its SH (spherical harmonic) expansion $\alpha(a/r)^{l+1}$ and so the maximum error on $S(r)$ is $|\alpha N_{l,l}|(a/r)^{l+1}$ or $\eta(a/r)^{l+1}$. For fixed $r < a$ we can always choose l large enough, so that the error magnitude on $S(r)$ is as big as we please, even though it is always exactly η on $S(a)$.

A somewhat less dry example is given pictorially on the next page, in Figure 3.3. The top two maps are the radial field B_r at the surface of the earth over the northern hemisphere, the right map from the IGRF model for 1980, the left one deliberately disturbed by noise. Each provides a spherical harmonic model out to degree and order 10. The noise is not very large and it is only just possible to see by eye the difference between the two surface maps. Below, in the second row, is the radial field, upward continued to a radius of 1.5 earth-radii. Now the noisy and original fields are totally indistinguishable. Notice two factors: first, the overall field size is much smaller (the units for all the maps are μT), conforming to the property that upward continued fields are always smaller than the originals; second, observe how the upward continued field is almost featureless – small-scale components have been filtered out, as predicted.

At the bottom is B_r downward continued to the depth of the core. The right map (the noise-free picture) and the left one are obviously very different – the error component has grown considerably in downward continuation. You should notice there is still some resemblance between the two fields, but there is more energy in the noisy map because incoherent signal has been added.

> *Exercise:* We have shown that an error field must decrease during upward continuation, but that is true of the field itself. Is it true that the more important quantity, the signal-to-noise ratio always improves with upward continuation? And must this ratio always deteriorate with downward continuation?

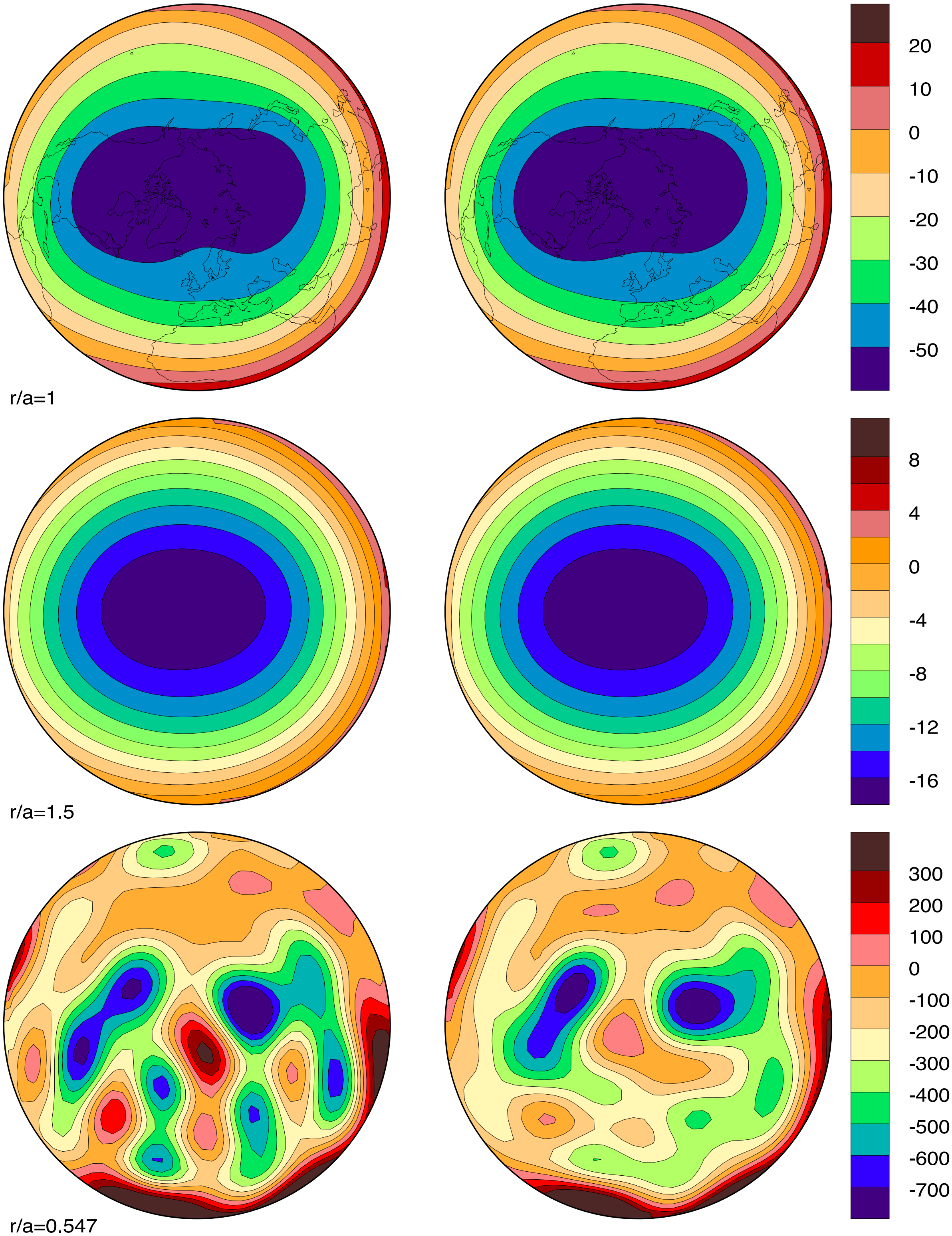

Figure 3.3

Figure 3.3: B_r, radial component of magnetic field in μT, evaluated from IGRF1980 at various radii. Left column has been perturbed by noise, right side is IGRF

3:4 Geomagnetic Field Models

In our discussions of the field so far we have not dealt with the issue of how to determine the magnetic field from a practical set of observations. In an earlier lecture it was stated that the spherical harmonic representation provides a unique description of any magnetic field represented by a harmonic potential. The field coefficients for rB_r can, in principle be derived from equation (48):

$$\beta_l^m = \int_{S(a)} B_r(a, \theta, \phi) Y_l^m(\theta, \phi)^* \frac{d^2\mathbf{r}}{a^2} \tag{56}$$

but this evidently requires knowledge of B_r all over $S(a)$, which is not possible in practice. Before we tackle the question of finding approximate models, and here we follow custom and say that a geomagnetic field model is a finite collection of SH coefficients, we discuss the mathematical question of what are sufficient measurements to define the field uniquely.

3:4.1 Geomagnetic Elements and Uniqueness

As we already noted, it is impossible to acquire complete knowledge of the radial magnetic field (or indeed any other component) on any spherical surface. Our data are always incomplete and noisy, and consequently there will always be ambiguity in the models derived from them. Because of this it might be argued that the issue of uniqueness in the case of complete and perfect data is of purely academic interest. However, experience in making magnetic field models based exclusively on one kind of observation (the magnitude of the field) shows that this is not the case.

Equation (56) relies on knowledge of the radial magnetic field, but these are not the only kind of observations that are made; typically when both surface survey and satellite data are involved we must consider all of the following kinds of observations:

B_r, B_θ, B_ϕ – orthogonal components of the magnetic field in geocentric reference frame.
X, Y, Z orthogonal components of the geomagnetic field in local coordinate system,
directed north, east, and downwards respectively. This is the geodetic coordinate system.
$H = (X^2 + Y^2)^{\frac{1}{2}}$ – horizontal magnetic field intensity
$B = (B_\theta^2 + B_\phi^2 + B_r^2)^{\frac{1}{2}} = (X^2 + Y^2 + Z^2)^{\frac{1}{2}}$ – total field intensity
$D = \tan^{-1}(Y/X)$ – declination
$I = \tan^{-1}(Z/H)$ - inclination.

If we are prepared to accept the approximation that the Earth is a sphere, then

$$X = -B_\theta; \quad Y = B_\phi; \quad Z = -B_r. \tag{57}$$

As we saw in Part I, the shape of the Earth is much better approximated by a spheroid or ellipsoid of revolution, with equatorial radius a, polar radius b, eccentricity, e

$$e^2 = \frac{(a^2 - b^2)}{a^2} \tag{58}$$

and flattening

$$f = \frac{(a\ -\ b)}{a} = 1/298.257. \tag{59}$$

Exercise:

What is the size of the error you make if you neglect to correct for the ellipsoidal shape of the earth and use geocentric latitude and longitude in calculating the field from a spherical harmonic model?

The fact that B_r on a surface $S(a)$ uniquely determines the field (because it is the solution of a Neumann boundary value problem for Laplace's equation) might lead one to hope that other measurements such as $|\mathbf{B}|$ would uniquely specify the magnetic field to within a sign. Most marine observations are of $|\mathbf{B}|$ and so are many satellite measurements, and scalar observations are generally more accurate than their vector counterparts so there might be advantages to just using those data in modeling. Or one might anticipate that exact knowledge of the direction of the field all over $S(a)$ would determine the field to within a multiplicative scaling constant. Paleomagnetic observations can determine ancient field directions much more reliably than field intensities, so one might hope to discover the shape of the geomagnetic field in those epochs when paleomagnetic data are plentiful. For the case of intensity data alone, George Backus showed in 1968 (*Q. J. Mech and Appl. Math* 21, pp 195-221, 1968; and *J. Geophys. Res.* 75, pp 6339-41, 1970; see also Jacobs, *Geomagnetism*, Vol 1, p 347) by constructing a counter example that knowledge of $|\mathbf{B}|$ was not sufficient. This provided an explanation of the poor quality of directional information predicted by models based on B alone, especially in equatorial regions. Early satellites had only measured total field intensity, and this result led to the launching of the first vector field satellite in 1979, known as MAGSAT. Backus' counter example is constructed in the following way: first find a magnetic field with potential Ψ_M that is everywhere orthogonal to a dipole field; then $\Psi_D + \Psi_M$ and $\Psi_D - \Psi_M$ have the same B everywhere.

Subsequently, Khokhlov, Hulot and Le Mouël (*Geophys. J. Internat.* 130, pp 701-3, 1997) showed that if knowledge of the location of the dip equator is added to knowledge of $|\mathbf{B}|$ everywhere then uniqueness of the solution is guaranteed.

Similarly, (despite an earlier argument to the contrary) it was shown in 1990 by Proctor and Gubbins, (*Geophys. J. Internat.* 100, pp 69-79, 1990) that $\hat{\mathbf{B}}$ on a spherical surface does not determine the field to within a scalar multiple. This is of some importance in the consideration of paleomagnetic data, many of which only provide approximate geomagnetic field directions with no associated information about the intensity. This result too has been investigated further and it has now been shown that if the magnetic field only has two poles then knowledge of its direction everywhere allows the geomagnetic field to be recovered, except for a constant multiplier (see Hulot, Khokhlov and Le Mouël, *Geophys. J. Internat.* 129, pp 347-54, 1997).

3:4.2 Construction of Field Models

We will suppose here that we have a finite collection of inaccurate observations of orthogonal components of the geomagnetic field $B_i = \mathbf{B} \cdot \hat{\mathbf{r}}_i$. Our goal is to derive from these observations the spherical harmonic coefficients that best represent the real geomagnetic field. Clearly, we cannot use (56) because our knowledge of $B_r(a, \theta, \phi)$ is incomplete.

Also it would not make use of our measurements of the tangential part of the field. The problem is analogous to (and in some respects identical with) that of interpolation to find a curve passing through a finite number of data. Many curves will do the job, and we cannot choose which one is more desirable without supplying additional information of some kind. One extra thing we know is that the magnetic field obeys a differential equation, but that turns out not to be enough information by itself. When the observations are not exact, but uncertain as in all real situations the issue is even more complicated. We obviously shouldn't expect the interpolant to pass through all (perhaps even any) of the observations. In the case where we would like to downward continue a geomagnetic field model to the core-mantle boundary, we need to be especially careful about how we deal with noise in the data; if we fit models with small scale structure derived from this noise, then it may dominate the real signal after downward continuation.

3:4.3 Least Squares Estimation

The time-honored technique for the construction of geomagnetic field models (invented by Gauss for this very purpose!) used to be that of least squares estimation of the spherical harmonic coefficients in a truncated spherical harmonic expansion for the measured field components. But, instead of the exact expansion with infinitely many terms, we decide ahead of time to model the data with an expansion truncated to degree L: so the scalar potential is

$$\Psi(r,\theta,\phi) = a \sum_{l=1}^{L} \sum_{m=-l}^{l} b_l^m \left(\frac{a}{r}\right)^{l+1} Y_l^m(\theta,\phi) \tag{60}$$

and we find

$$\mathbf{B} = -\nabla\Psi. \tag{61}$$

If we measure orthogonal components of the field B_r, B_θ, B_ϕ we have a set of observations at N sites designated $\mathbf{r}_j$

$$\begin{aligned} d_j &= \hat{\mathbf{s}}_j \cdot \mathbf{B}(\mathbf{r}_j), \quad j = 1, ..., N \\ &= \sum_{l=1}^{L} \sum_{m=-l}^{l} b_l^m a^{l+2} \hat{\mathbf{s}}_j \cdot \nabla \left[\frac{Y_l^m(\hat{\mathbf{r}}_j)}{r_j^{l+1}}\right] + \epsilon_j. \end{aligned} \tag{62}$$

The data d_j are then linear functionals of the b_l^m specifying the field. With an appropriate indexing scheme for the b_l^m we can write a prediction for our observations d_j as a matrix equation.

$$\mathbf{d} = \mathbf{G}\mathbf{b} + \mathbf{n} \tag{63}$$

with $\mathbf{d}, \mathbf{n} \in R^N$, $\mathbf{b} \in R^K$. $\mathbf{G}$ is an $N \times K$ matrix, and for each d_j we can compute g_{kj} the contribution of the relevant spherical harmonic at that point; the vector $\mathbf{n}$ contains the misfits between the model predictions and the actual measurements. The total number of parameters, the length of the vector $\mathbf{b}$, is determined by the truncation level: $K = L(L + 2)$. Thus

$$\mathbf{b} = \begin{bmatrix} b_1^{-1} \\ b_1^0 \\ b_1^1 \\ \vdots \\ b_L^L \end{bmatrix}. \tag{64}$$

The value of L is chosen so that K is (much) less than N, the number of data, so there are fewer free parameters than data to be fit. This means that it is impossible to choose $\mathbf{b}$ to get an exact match to the data, and so $\mathbf{n}$ is not a vector of zeros. Least squares estimation involves finding the values for $\mathbf{b}$ that minimize $||\mathbf{n}||^2 = ||\mathbf{d} - \mathbf{G} \cdot \mathbf{b}||^2$, where the notation $|| \cdot ||$ is called a norm – in this case it is the ordinary length of the vector. The idea here is to do the best job possible with the available free variables and make the model predictions as close to the data as they can be measured by the length of the misfit vector.

Straightforward calculus can show that the LS vector can be written in terms of the solution to the normal equations:

$$\hat{\mathbf{b}} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}. \tag{65}$$

Note that (65) is for several reasons not a good way to find $\hat{\mathbf{b}}$ in a computer – first linear systems of equations ought never to be solved by calculating a matrix inverse (it wastes time and is inaccurate); second, there is a clever way of writing the LS equations that avoids a serious numerical precision problem arising in (65). A result from statistics, known as the Gauss-Markov Theorem, shows that provided the misfits are due to random, uncorrelated perturbations with zero mean, and have a common variance, then the least squares solution is the best linear unbiased estimate (BLUE) available, in the sense that they have the smallest variance amongst such estimates. Also

$$E[||\mathbf{n}||^2] = (N - K)\sigma^2 \tag{66}$$

where σ^2 is the variance of the noise process. If the noise has a known covariance structure, then the theory can readily be adjusted to take this into account.

A fundamental problem with this approach is that although we might have an idea of the size of the uncertainty in the various observations, and we could choose the truncation level of the spherical harmonic expansion so (66) is approximately satisfied, we do not know that the K spherical harmonics we have chosen adequately describe the geomagnetic field. In other words, the misfit has two sources, not one: measurement error and an insufficiently detailed model. To guarantee a complete description of the real field $L = \infty$; in that case the Gauss-Markov theorem does not apply, nor does (66), and we have no uncertainty estimates for our model. Truncation at finite L corresponds to an assumption about simplicity in the model that has no physical basis – the resulting field model may be biased by the truncation procedure.

3:4.4 Regularization – an Alternative to Least Squares

An alternative to LS fitting that has been very widely used since the mid 1980s and is the basis of almost all modern geomagnetic field modeling is to choose "simple" models for the field of an explicit kind. To illustrate the concept we return to our one-dimensional interpolation problem.

One widely used solution to the problem of interpolation is to use what is known as cubic spline interpolation. Data are connected by piecewise cubic polynomials, with continuous derivatives up to second order. The cubic spline interpolant has the property that it is the smoothest curve connecting the points, in the precise sense of minimizing RMS second derivative:

$$\int_{x_1}^{x_N} \left[\frac{d^2 f}{dx^2}\right]^2 dx. \tag{67}$$

The spline solution is the solid line in the Figure 3.4.4.1; any other curve, like the dashed ones, have a larger value for the integral (67).

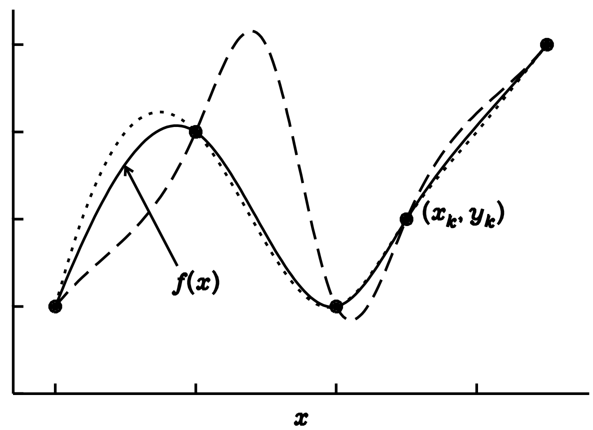

Figure 3.4.4.1

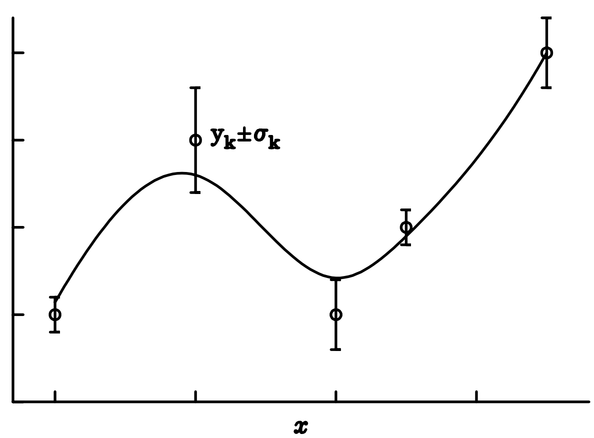

Figure 3.4.4.2

If the data are noisy we can still model them by the same kind of curve, but we no longer require the curve to pass exactly through the model points. Instead, we ask for a satisfactory fit, usually defined in terms of norm of the misfit. So we seek the curve with the smallest RMS second derivative, that has a target misfit. This is another example of a constrained minimization, and like the one we saw in Part I, we can solve it with a Lagrange multiplier. The solution to the constrained problem is determined by finding the stationary points of an objective functional of the following kind:

$$U = \sum_{j=1}^{N} \frac{(f(x_j) \ - \ y_j)^2}{\sigma_j^2} \ + \ \lambda \int_{x_1}^{x_N} \left[\frac{d^2 f}{dx^2}\right]^2 dx \tag{68}$$

subject to

$$\sum_{j=1}^{N} \frac{(f(x_j) - y_j)^2}{\sigma_j^2} = T. \tag{69}$$

The size of the Lagrange multiplier λ is dictated by the constraint that the data be fit to a reasonable tolerance level. If the uncertainty in the observations, σ_j, is known one reasonable choice for T is the expected value of χ_N^2, namely N.

The same ideas are used in geomagnetic field modeling. However, the vector nature of the field makes the choice of penalty functional more complicated – we want to find some property of the magnetic field that can be used like (67) to minimize wiggliness or complexity in our models of the geomagnetic field, either at Earth's surface or at the CMB (core-mantle boundary). The general idea of constructing models that minimize a penalty other than data misfit is called *regularization*.

One candidate penalty function is

$$E = \int_{r>a} |\mathbf{B}|^2 d^3\mathbf{r}. \tag{70}$$

Since $|\mathbf{B}(\mathbf{r})|^2/2\mu_0$ is the energy density of the magnetic field at $\mathbf{r}$, $E/2\mu_0$ is the total energy stored in $\mathbf{B}$ outside the sphere of radius a. We can reduce this integral to a manageable form in terms of spherical harmonics. We write $\mathbf{B} = -\nabla\Psi$, then

$$E = \int_{r>a} \nabla\Psi \cdot \nabla\Psi \; d^3\mathbf{r}. \tag{71}$$

Next we make use of a familiar vector identity [number 4, in our list: $\nabla \cdot (\phi\mathbf{A}) = \nabla\phi \cdot \mathbf{A} + \phi\nabla \cdot \mathbf{A}$, and letting $\mathbf{A} = \nabla\Psi$] followed by Laplace's equation to write

$$\begin{aligned} E &= \int [\nabla \cdot (\Psi\nabla\Psi) \ - \ \nabla^2\Psi] \; d^3\mathbf{r} \\ &= \int \nabla \cdot (\Psi\nabla\Psi) \; d^3\mathbf{r}. \end{aligned} \tag{72}$$

Using Gauss' Divergence Theorem we can rewrite the volume integral in terms of a surface integral over $S(a)$

$$E = \int_{S(a)} -\Psi \frac{\partial\Psi}{\partial r} \; d^2\mathbf{r}. \tag{73}$$

Now we simply substitute the spherical harmonic expansion (60) for the potential Ψ.

$$E = \int_{S(a)} [a \sum_{l,m} b_l^m Y_l^m(\hat{\mathbf{r}})][\sum_{l',m'} (l' + 1) b_{l'}^{m'} Y_{l'}^{m'}(\hat{\mathbf{r}})]^* a^2 d^2\hat{\mathbf{r}}$$

$$= a^3 \sum_{l,m} \sum_{l',m'} (l' + 1)\, b_l^m (b_{l'}^{m'})^* \int_{S(a)} Y_l^m(\hat{\mathbf{r}}) Y_{l'}^{m'}(\hat{\mathbf{r}})^* d^2\hat{\mathbf{r}}$$

$$= a^3 \sum_{l=1}^{\infty} \sum_{m=-l}^{l} (l + 1)|b_l^m|^2. \tag{74}$$

Hence E can be written as a positive weighted sum of the squared absolute values of the spherical harmonic coefficients. This sum is now in a form we can use in minimizing an objective function like (71) for the magnetic field. The weighting by increasing l means that higher degree (shorter wavelength) contributions to the model will be strongly penalized in minimizing the objective functional. (74) is one example of a set of norms (measures of size) of the kind

$$||\mathbf{B}||_w^2 = \sum_{l=1}^{\infty} w_l \sum_{m=-l}^{l} |b_l^m|^2, \quad w_l > 0. \tag{75}$$

Many interesting properties corresponding to smoothness or small size of the field can be written in this form

$$\int_{S(a)} \mathbf{B} \cdot \mathbf{B}\, d^2\hat{\mathbf{r}} \qquad w_l = (2l + 1)(l + 1) \tag{76}$$

$$\int_{S(a)} [\nabla_1 \hat{\mathbf{r}} \cdot \mathbf{B}]^2 d^2\hat{\mathbf{r}} \quad w_l = l(l + 1)^2(l + \frac{1}{2}) \tag{77}$$

$$\int_{S(a)} [\nabla_1^2 \hat{\mathbf{r}} \cdot \mathbf{B}]^2 d^2\hat{\mathbf{r}} \quad w_l = l^2(l + 1)^4 \tag{78}$$

$$\int_{r<a} \mathbf{J}_T \cdot \mathbf{J}_T d^3\mathbf{r} \qquad w_l = a^3 \mu_0^2 (l + 1)(2l + 1)^2(2l + 3). \tag{79}$$

In (79) $\mathbf{J}_T$ is the toroidal part of the current flow in Earth's core, whose significance is discussed in section 3.5. These ideas were first set out in a paper by Shure, Parker, and Backus, *Phys. Earth Planet. Inter.* 28, pp 215-29, 1982.

3:4.5 Results - Gauss Coefficients

In Part I you may recall that the expansion of the Earth's gravitational potential was called Stokes' expansion. In geomagnetism the honor goes to Karl Friedrich Gauss: the expansion coefficients are invariably called *Gauss coefficients*. But there are some differences. First the fundamental one, that makes geomagnetism more interesting than gravity – the field is changing on times scales of a human lifetime. So the coefficients must always be dated. Secondly, and almost trivially, geomagnetists never use fully normalized spherical harmonics – instead they employ an awkward real (as opposed to complex) representation in which the basis functions (spherical harmonics) are normalized so that

$$\int_{S(1)} (Y_l^m)^2 d^2\hat{\mathbf{r}} = \frac{4\pi}{2l+1}. \tag{80}$$

Then the scalar potential is written

$$\Psi(r,\theta,\phi) = a\sum_{l=1}^{\infty}\left(\frac{a}{r}\right)^{l+1}\sum_{m=0}^{l} N_{lm}(g_l^m \cos m\phi \ + \ h_l^m \sin m\phi)P_l^m(\cos\theta) \tag{81}$$

where $a = 6,371.2$ km, the mean Earth radius, the P_l^m are the Associated Legendre functions from Part I, and

$$\begin{aligned} N_{lm} &= 1, \qquad\qquad m = 0 \\ &= \sqrt{\frac{2(l-m)!}{(l+m)!}}, \ \ m > 0. \end{aligned} \tag{82}$$

You will find the numbers g_l^m, h_l^m tabulated in many places. There are official models designated IGRF for the International Geomagnetic Reference Field, agreed upon every five years by the International Association for Geomagnetism and Aeronomy as a good approximation to the field. The linearized rate of change of the field is also given for interpolation or extrapolation across the five year period. We are currently at Version 11 and the model coefficients can be found at http://www.ngdc.noaa.gov/IAGA/vmod/igrf.html. See Langel's chapter in Jacob's book *Geomagnetism*, an article rich in nuance, with a detailed description of how these are computed, and many tables. Below is the IGRF-9, 2000. The values are in nanotesla. Figure 3.4.5 shows various views of the radial component of the magnetic field from IGRF 2000: top is B_r at Earth's surface; middle the non-dipole contribution to B_r at Earth's surface and the bottom panel gives B_r downward continued to the core-mantle boundary.

Gauss Coefficients for IGRF-2000

l	m	g_l^m	h_l^m	l	m	g_l^m	h_l^m	l	m	g_l^m	h_l^m
1	0	-29615	0	6	2	74	64	9	0	5	0
1	1	-1728	5186	6	3	-161	65	9	1	9	-20
2	0	-2267	0	6	4	-5	-61	9	2	3	13
2	1	3072	-2478	6	5	17	1	9	3	-8	12
2	2	1672	-458	6	6	-91	44	9	4	6	-6
3	0	1341	0	7	0	79	0	9	5	-9	-8
3	1	-2290	-227	7	1	-74	-65	9	6	-2	9
3	2	1253	296	7	2	0	-24	9	7	9	4
3	3	715	-492	7	3	33	6	9	8	-4	-8
4	0	935	0	7	4	9	24	9	9	-8	5
4	1	787	272	7	5	7	15	10	0	-2	0
4	2	251	-232	7	6	8	-25	10	1	-6	1
4	3	-405	119	7	7	-2	-6	10	2	2	0
4	4	110	-304	8	0	25	0	10	3	-3	4
5	0	-217	0	8	1	6	12	10	4	0	5
5	1	351	44	8	2	-9	-22	10	5	4	-6
5	2	222	172	8	3	-8	8	10	6	1	-1
5	3	-131	-134	8	4	-17	-21	10	7	2	-3
5	4	-169	-40	8	5	9	15	10	8	4	0
5	5	-12	107	8	6	7	9	10	9	0	-2
6	0	72	0	8	7	-8	-16	10	10	-1	-8
6	1	68	-17	8	8	-7	-3				

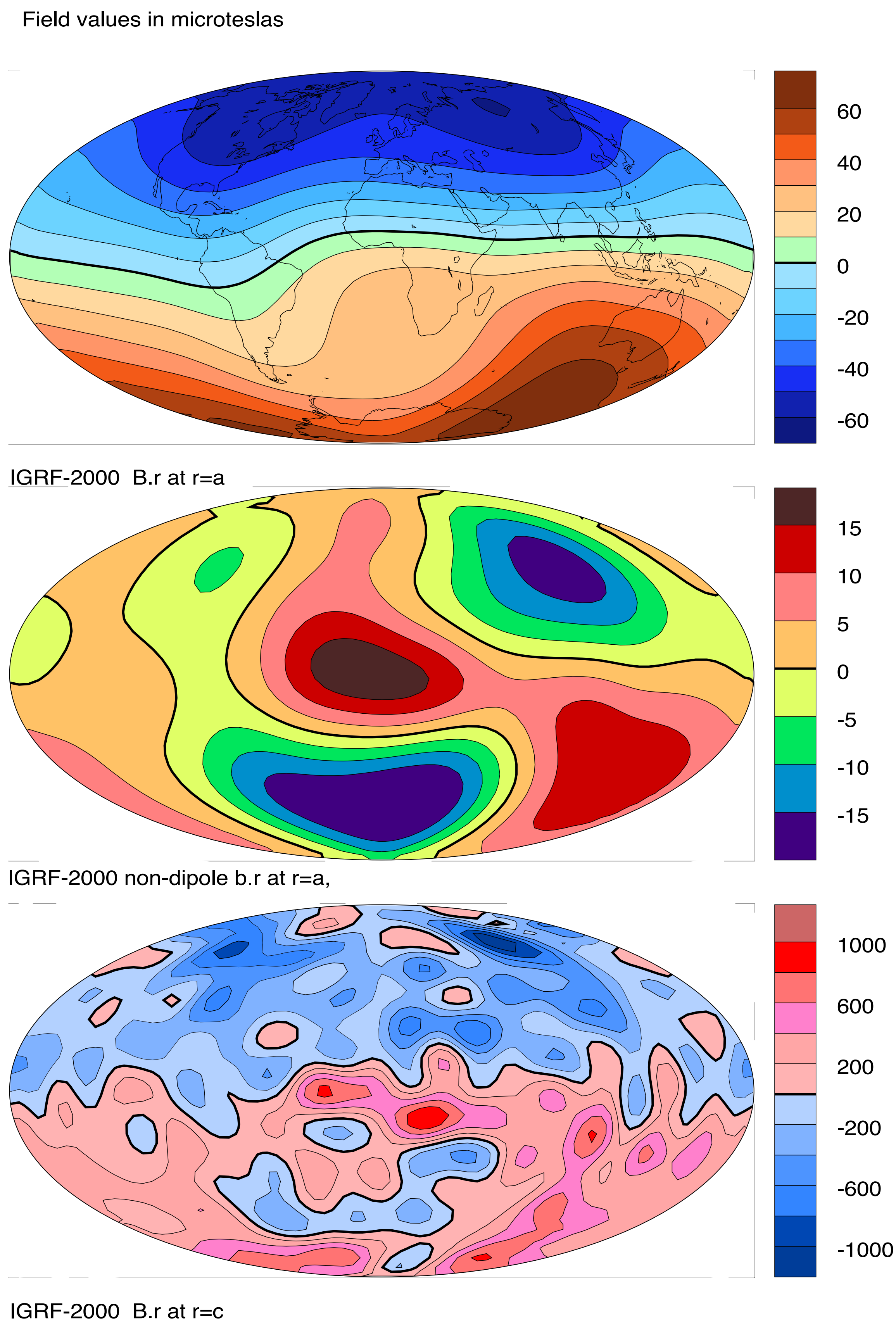
Field values in microteslas
60
40
20
0
-20
-40
-60
IGRF-2000 B.r at r=a
15
10
5
0
-5
-10
-15
IGRF-2000 non-dipole b.r at r=a,
1000
600
200
-200
-600
-1000
IGRF-2000 B.r at r=c

Figure 3.4.5

3:4.6 The Geomagnetic Spectrum

In 1974 Frank Lowes (*Geophys. J. Royal Astron. Soc.* 36, pp 717-730, 1974) had the idea of plotting the magnetic energy averaged over the Earth's surface as a function of wavelength. Of course on a sphere this really means as a function SH degree, because by Jeans' rule $\lambda = 2\pi a/(l + \frac{1}{2})$. Recall the energy density in a magnetic field is given by $\mathbf{B} \cdot \mathbf{B}/2\mu_0$ so the energy integrated over a spherical surface is

$$E(a) = \int_{S(a)} \frac{\mathbf{B} \cdot \mathbf{B}}{2\mu_0} \, d^2\mathbf{s}$$

$$= \frac{1}{2\mu_0} \int_{S(a)} \nabla\Psi \cdot \nabla\Psi \, d^2\mathbf{s} \tag{83}$$

$$= \frac{a^2}{2\mu_0} \int_{S(1)} [(\partial_r \Psi)^2 \; + \; |\nabla_s \Psi|^2] \, d^2\hat{\mathbf{s}} \tag{84}$$

where Ψ is the usual scalar potential. Inserting the familiar SH expansion for Ψ [equation (60) with $L = \infty$] we have:

$$\frac{2\mu_0 E(a)}{a^2} = \sum_{l,m} \sum_{l',m'} [(l+1)(l'+1) b_l^m (b_{l'}^{m'})^* \int_{S(1)} Y_l^m (Y_{l'}^{m'})^* d^2\hat{\mathbf{s}}$$

$$+ \, b_l^m (b_{l'}^{m'})^* \int_{S(1)} \nabla_1 Y_l^m \cdot (\nabla_s Y_{l'}^{m'})^* \; d^2\hat{\mathbf{s}}]. \tag{85}$$

By the orthogonality relations of Y_l^m and its surface gradient (3 and 12 in the table of lore) the double sums all collapse

$$\frac{2\mu_0 E(a)}{a^2} = \sum_{l,m} \, [(l+1)^2 \; + \; l(l+1)] |b_l^m|^2$$

$$= \sum_{l=1}^{\infty} (2l+1)(l+1) \sum_{m=-l}^{l} |b_l^m|^2. \tag{86}$$

This is essentially the result stated by (76). The idea is to see how the various harmonic degrees (equivalently, length scales) contribute to the total energy, and so traditionally one writes

$$R = \sum_{l=1}^{\infty} \; R_l \tag{87}$$

where we define

$$R_l = \frac{(2l+1)(l+1)}{4\pi} \sum_{m=-l}^{l} |b_l^m|^2 = (l+1) \sum_{m=0}^{l} \, [(g_l^m)^2 + (h_l^m)^2]. \tag{88}$$

Here R_l is called the *geomagnetic spectrum* or the *Lowes spectrum* (or even the Mausersberger-Lowes spectrum, because Mauersberger mentioned R_l some years before Lowes, though he didn't do anything very original with it).

What Lowes discovered when he plotted the log of R_l against l, and what we see with much more complete data, is that the spectrum breaks very obviously into two parts, the lower l behavior well-fitted by a straight line on a linear-log plot. See Figure 3.4.6.1; the field model is due to Cain, Wang, Kluth, and Schmitz, *Geophys. J. Royal Astron. Soc.* 97, pp 431-42,

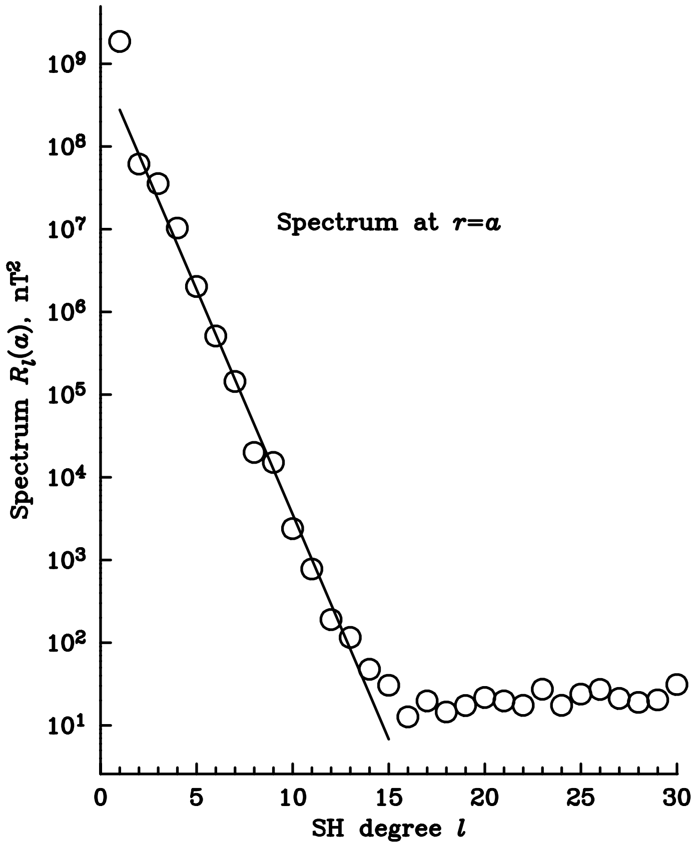

Figure 3.4.6.1

1989. We have exponential behavior, completely different from the gravity spectrum's, which is essentially a power law (See Fig 7, Part I, though we plot a slightly different kind of spectrum, without the multiplying quadratic in l).

The natural interpretation of this result is that the two parts of the spectrum reflect different source regions – the long wavelength fields with $l \leq 13$ come almost entirely from the core, those with $l > 13$ from the crust. This idea is given considerable support from the observation that the equation of the best straight line through the core spectrum is $\alpha(r_1/a)^{2l}$ where $r_1 = 3,407$km, which is not very different from the core radius $c = 3,486$ km, according to seismologists. Why would one expect this equation for fields with their sources in the core? One argument runs as follows: consider the spectrum on spheres of different radii r. If it is permissible to downward continue the field through mantle, treating the mantle as a nonmagnetic insulator, then we find that

$$R_l(r) = (a/r)^{2l+4} R_l(a). \tag{89}$$

When we substitute $r = c$ and plot the spectrum that would be observed at the surface of the core we obtain the spectrum in Figure 3.4.6.2, which is almost flat. A flat spectrum is one in which there is equal energy at every scale, the sort of

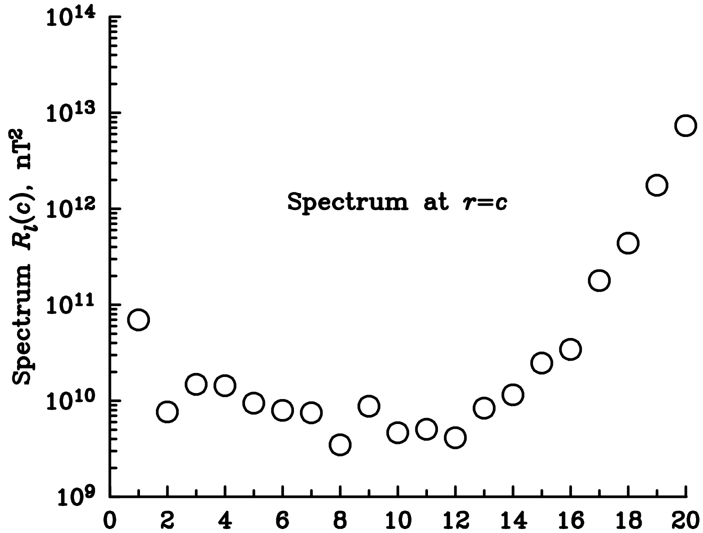

Figure 3.4.6.2

thing predicted in homogeneous turbulence. A plausible argument might be made that the fluid motions in the core cause a distribution of magnetic energy evenly into the different scales. This seems to work, except for the dipole term, which clearly is unusually large, and doubtless this fact reflects the dominance of the Earth's rotation in the geodynamo.

The large-l spectrum makes no sense if the sources are the core; why should the energy apparently increase exponentially with l? Rather we can show (Jackson, *Geophys. J. Int.* 103, pp 657-74, 1990) that randomly distributed sources of magnetization in the crust would yield a mildly rising spectrum like the one seen in Figure 3.4.6.1. We suspect the core contribution continues on its exponential decline, but is completely obscured by the crustal field at shorter wavelengths.

Finally, Constable and Parker (*J. Geophys. Res.* 93, pp 11569-81, 1988) suggested that the flatness of $R_l(c)$ was most unlikely to be an accident of the 20th century, but instead it may be a persistent feature of the geomagnetic field common to the geodynamo throughout geologic history, with the possible exception of times near reversals. They also discovered that the coefficients b_l^m when normalized to the radius of the core resembled a set of numbers drawn from a zero mean, Gaussian random process, a remarkable fact in itself. If this property holds for other geologic times, it becomes possible to predict what constitutes "normal" statistical behavior of the ancient geomagnetic field. Subsequent work has shown that the true field had some small but definite departures from the uniform Gaussian model, but it remains a useful approximation.

3:5 Toroidal and Poloidal Fields

Before we start, let us assemble two useful vector calculus identities, numbers 5 and 9 in our list:

$$\nabla \times (\mathbf{U}s) = (\nabla \times \mathbf{U})s \;-\; \mathbf{U} \times \nabla s \tag{90}$$

$$\nabla \times \nabla \times \mathbf{U} = -\nabla^2\mathbf{U} \;+\; \nabla(\nabla \cdot \mathbf{U}). \tag{91}$$

In an earlier lecture we showed that the magnetic field can always be written as the curl of the magnetic vector potential **A**. Now we consider a sphere of radius c surrounded by an insulator and divide the vector potential into parts parallel to and perpendicular to **r** by writing

$$\mathbf{A} = T\mathbf{r} \;+\; \nabla P \times \mathbf{r} = T\mathbf{r} \;+\; \nabla \times (P\mathbf{r}) \tag{92}$$

where T and P are scalar functions of **r**, known as the defining scalars of the toroidal and poloidal fields. That this is always possible is shown in detail in *Foundations*, Chapter 5. To find **B** we take the curl:

$$\begin{aligned}\mathbf{B} &= \nabla \times (T\mathbf{r}) \;+\; \nabla \times \nabla \times (P\mathbf{r}) \\ &= \mathbf{B}_T \;+\; \mathbf{B}_P\end{aligned} \tag{93}$$

and $\mathbf{B}_T$ is called the *toroidal part* of **B**, while $\mathbf{B}_P$ is the *poloidal part.* This decomposition for **B** is unique and can always be done for all solenoidal vector fields (those with $\nabla \cdot \mathbf{F} = 0$). We are very familiar with the idea of a potential for **B** in an insulator; when **J** does not vanish, we need two scalars, not one to describe **B** completely. Conventionally, the scalars are always restricted to a class of functions whose average value over every sphere is zero, that is

$$0 = \int_{S(r)} T(r\hat{\mathbf{r}})\, d^2\hat{\mathbf{r}} = \int_{S(r)} P(r\hat{\mathbf{r}})\, d^2\hat{\mathbf{r}} \tag{94}$$

With this property, the scalars become unique in (92), which means that if $\mathbf{B}_T$ vanishes, then $T = 0$, and similarly for $\mathbf{B}_P$.

Using the vector identity (90) on the toroidal part of the field, we see

$$\begin{aligned}\mathbf{B}_T &= (\nabla \times \mathbf{r})T \;-\; \mathbf{r} \times \nabla T \\ &= -\,\mathbf{r} \times \nabla T\end{aligned} \tag{95}$$

because $\nabla \times \mathbf{r} = 0$; hence $\hat{\mathbf{r}} \cdot \mathbf{B}_T = 0$ always; in other words, the toroidal magnetic field has no radial component – it is a tangent vector field on every concentric spherical surface. The lines of force lie on spherical surfaces and are thus confined to the interior of the conducting sphere. If we think of the sphere as Earth's core, outside the core we have (for the sake of argument) $\mathbf{J} = 0$ and

$$\mathbf{B} = -\nabla\Psi, \quad \text{with } \nabla^2\Psi = 0. \tag{96}$$

Now **B** is continuous at $S(c)$ and since $\hat{\mathbf{r}} \cdot \mathbf{B}_T = 0$ just inside we conclude that $\hat{\mathbf{r}} \cdot \mathbf{B}_T$ is also zero just outside the core. But we showed in Part I section 8, that any harmonic function with internal sources and vanishing radial component on $S(c)$

is identically zero outside. Therefore $\mathbf{B}_T$ vanishes outside the conducting sphere. Hence the toroidal part of $\mathbf{B}$ in Earth's core is invisible outside the core and only the poloidal part, $\mathbf{B}_P$, has any detectable influence at the Earth's surface.

When we neglect the displacement current term in Maxwell's equations we can make a similar decomposition for the current flow in core, as we now show:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \tag{97}$$

$$\nabla \cdot \mathbf{J} = 0 \tag{98}$$

$$\mathbf{J} = \mathbf{J}_T + \mathbf{J}_P. \tag{99}$$

By substituting for $\mathbf{B}$ in terms of $\mathbf{B}_T$ and $\mathbf{B}_P$ in (93) we will show that $\mathbf{B}_P$ comes from $\mathbf{J}_T$ and $\mathbf{B}_T$ from $\mathbf{J}_P$, that is, toroidal fields are generated by poloidal current systems, and poloidal fields by toroidal currents.

Here is the proof: take the curl of (93)

$$\mu_0 \mathbf{J} = \nabla \times \mathbf{B} = \nabla \times \nabla \times (T\mathbf{r}) + \nabla \times \nabla \times \nabla \times (P\mathbf{r}). \tag{100}$$

It is obvious from the first term on the right that the toroidal part of $\mathbf{B}$ is associated with a poloidal current, whose scalar is just T/μ_0. To show the last term, with three curls, is in fact toroidal, we analyze $\mathbf{S} = \nabla \times \nabla \times (P\mathbf{r})$. From the identity (91)

$$\mathbf{S} = -\nabla^2(\mathbf{r}P) + \nabla(\nabla \cdot (\mathbf{r}P)). \tag{101}$$

We expand the first term in (101) with the Einstein notation:

$$\nabla^2(\mathbf{r}P) = \partial_j \partial_j (x_k P) = \partial_j((\partial_j x_k)P + x_k \partial_j P) = \partial_j(\delta_{jk} P + x_k \partial_j P) \tag{102}$$

$$= \partial_k P + \partial_j(x_k \partial_j P) = \partial_k P + (\partial_j x_k)\partial_j P + x_k \partial_j \partial_j P$$

$$= \partial_k P + \delta_{jk} \partial_j P + x_k \partial_j \partial_j P = 2\partial_j P + x_k \partial_j \partial_j P$$

$$= 2\nabla P + \mathbf{r}\nabla^2 P. \tag{103}$$

Then (101) becomes

$$\mathbf{S} = -\mathbf{r}\nabla^2 P + \nabla(\nabla \cdot (\mathbf{r}P) - 2P)$$

$$= -\mathbf{r}\nabla^2 P + \nabla(r\partial_r P + P). \tag{104}$$

Finally, we take the curl of (104), noting the vanishing of the term involving a grad because $\nabla \times \nabla = 0$, so that

$$\nabla \times S = -\nabla \times (\mathbf{r}\nabla^2 P)$$

and we recover (100)

$$\mu_0 \mathbf{J} = -\nabla \times (\mathbf{r}\nabla^2 P) + \nabla \times \nabla \times (T\mathbf{r}) = \mu_0(\mathbf{J}_T + \mathbf{J}_P). \tag{105}$$

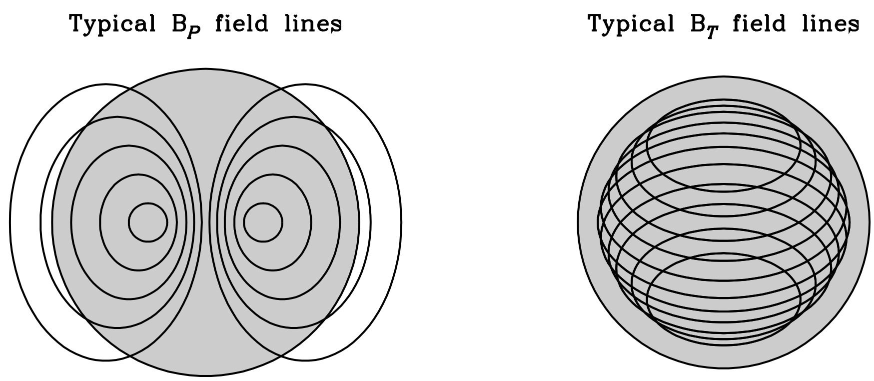

Figure 3.5.1

Thus the toroidal part of the current is derived from the toroidal scalar $-\nabla^2 P/\mu_0$.

The lengthy algebra may obscure what a toroidal or poloidal magnetic field might actually look like. The sketches in Figure 3.5.1 give an example of the simplest types of fields. You can easily see the toroidal system on the right as the currents generating the poloidal field on the left. But if you reverse the process, the field lines that stray outside the conductor (gray area) on the left are not allowed. Obviously not every toroidal or poloidal field has axial symmetry like these.

So we know that the geomagnetic field that we see is connected to a poloidal field in the core, which is generated by a toroidal current system J_T. The following nice result is also true (shown first by Gubbins 1975, and in more detail in *Foundations*, section 5.5, although there is an error of a factor 4π in the bounds given there):

$$\mu_0^2 \int_{r<c} \mathbf{J}_T \cdot \mathbf{J}_T d^3\mathbf{r} \ \geq \ c \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \frac{(l+1)(2l+1)^2(2l+3)}{l} |b_{lm}|^2 \tag{106}$$

where the coefficients b_l^m are those obtained on the surface of the core, whose radius is c. The amount of heat generated by ordinary Ohmic losses (also called *Joule heating*) at a point is given by $\mathbf{J} \cdot \mathbf{E}$. Thus (106) allows us to compute the minimum Joule heating from toroidal currents:

$$Q = \int_{r<c} \frac{\mathbf{J}_T \cdot \mathbf{J}_T}{\sigma} d^3\mathbf{r}$$

$$\geq \ \frac{c}{\mu_0^2 \sigma} \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \frac{(l+1)(2l+1)^2(2l+3)}{l} |b_{lm}|^2. \tag{107}$$

By using Q as a regularizing norm, we can find the minimum amount of heating in the core associated with the poloidal magnetic field, the only part we can see. Of course the actual ohmic heating could be much larger than this because of the poloidal current term which remains invisible to us and because the estimate from (96) is a lower bound on toroidal current power in any case. It is controversial in dynamo theory whether the toroidal component of **B** in the core is large compared with the poloidal part. Some models predict $|\mathbf{B}_T| \sim 50 \times |\mathbf{B}_P|$.

The Joule heating in the core is probably a small fraction of the heat budget (see the exercise), but it is believed that in certain neutron stars, called *magnetars*, the collapse of the magnetic field (with intensity 10^9 T!) is the major source of energy in the body. See S. Kalkarni, *Nature*, v 419, pp 121-2, 2002.

Exercise:

(a) Make an estimate of the minimum rate of Joule heating generated in core. Compare this with terrestrial heat flow at the surface.

(b) A toroidal magnetic field $\mathbf{B}_T$ fills a conducting sphere of radius a; its scalar is the function $T(\mathbf{r})$. Imagine creating a contour map of the values of T on the surface $S(b)$ with $b < a$. What connection does the contour map of the scalar T have with the magnetic field $\mathbf{B}_T$?