1. INTRODUCTION

Geomagnetism, the study of Earth’s magnetic field, has a long history and has revealed much about the way the Earth works. As we shall see, the existence and characteristics of the field essentially demand that the fluid outer core be made of electrically conducting material that is convecting in such a way as to maintain a self-sustaining dynamo. The study of the field as it is recorded in rocks is known as paleomagnetism. It allows us to track the past motions of continents and leads directly to the idea of sea-floor spreading. Mapping the signature of the magnetic field in Earth’s lithosphere provides information used in large scale tectonic studies. Paleomagnetism also allows the study of longevity of the field (~ 3.5 billion years), and how the geomagnetic field has evolved over geological time, through tracking of geomagnetic polarity reversals, and variations in the field’s strength and direction. Shorter term variations in the external part of the geomagnetic field induce secondary variations in Earth’s crust and mantle which are used to study the electrical properties of the Earth, giving insight into porosity, temperature, and composition in these regions. Changes in the external field are controlled by the interactions between the solar wind and Earth’s internal field and are of enormous interest in understanding solar terrestrial interactions.

The magnetic field was the first property attributed to the Earth as a whole, aside from its roundness. This was the finding of William Gilbert, physician to Queen Elizabeth I, who published his inference in 1600, predating Newton’s gravitational *Principia* by about 87 years. The magnetic compass had been in use, beginning with the Chinese, since about the second century B.C., but it did not find its way to Europe until much later, where it became an indispensable tool for maritime navigation. Petrus Peregrinus can be credited with producing the first scientific work devoted to magnetism, discovering magnetic meridians, the dipolar nature of the magnet, and describing two versions of the magnetic compass. The *Epistola de Magnete* was written in 1269, and subsequently widely circulated in Europe, but not actually published until the 16th century. Gilbert placed the source of magnetism within the Earth in 1600, but the temporal variations in the magnetic field (known as secular variation) were not well documented until the middle of the seventeenth century when Henry Gellibrand appreciated that the differences among repeated measurements were not just inaccuracy in the observations. In 1680 Edmund Halley (of Halley’s comet fame) published the first contour map of the geomagnetic variation as the declination was then known: he envisioned the secular variation of...
the field as being caused by a collection of magnetic dipoles deep within the earth drifting westward with time with about a 700 year period, a model not dissimilar to many put forward during the twentieth century, although he did not know of the existence of the fluid outer core. A formal separation of the geomagnetic field into parts of internal and external origin was first achieved by the German mathematician Karl Friedrich Gauss in the nineteenth century. Gauss invented spherical harmonics and deduced that by far the largest contributions to the magnetic field measured at Earth’s surface are generated by internal rather than external magnetic sources, thus confirming Gilbert’s earlier speculation. He was also responsible for beginning the measurement of the geomagnetic field at globally distributed observatories, some of which are still running today.

Figure 1.1

The magnetic field is a vector quantity, possessing both magnitude and direction; at any point on Earth a free compass needle will point along the local direction of the field. Although we conventionally think of compass needles as pointing north, it is the horizontal component of the magnetic field that is directed approximately in the direction of the North Geographic Pole. The difference in azimuth between magnetic north and true or geographic north is known as declination (positive eastward). The field also has a vertical contribution; the angle between the horizontal and the magnetic field direction is known as the inclination and is by convention positive downward (see Figure 1.1). Three parameters are required to describe the magnetic field at any point on the surface of the Earth, and the conventional choices vary according to subfields of geomagnetism and paleomagnetism. Traditionally, the vector $\mathbf{B}$ at Earth’s surface is referred to a right-handed coordinate system: north-east-down for $x$-$y$-$z$. But often instead of using the components in this system, three numbers used are: intensity, $B = |\mathbf{B}|$, declination, $D$, and inclination, $I$ as shown in the
sketch or \( D, H \) and \( Z \); \( H \), or equivalently \( B_h \), is the projection of the field vector onto the horizontal plane and \( Z \), or equivalently \( B_z \), is the projection onto the vertical axis. \( D \) is measured clockwise from North and ranges from \( 0 \to 360^\circ \) (sometimes \( -180^\circ \to 180^\circ \)). \( I \) is measured positive down from the horizontal and ranges from \( -90 \to +90^\circ \) (because field lines can also point out of the Earth, indeed it is only in the northern hemisphere that they are predominantly downward). From the diagram we have

\[
H = B \cos I; \quad B_z = B \sin I. \tag{1}
\]

When components of \( B \) are used they are called \( X, Y, Z \), and:

\[
X = B_x = B \cos I \cos D; \quad Y = B_y = B \cos I \sin D; \quad Z = B_z = B \sin I. \tag{2}
\]

The CGS unit of \( B \) is the gauss; smaller fields were once measured in gammas where \( 1 \, \gamma = 10^{-5} \, G \). Today SI units should be universally used: \( B \) is measured then in tesla (T); 1 T is a very large field. More commonly in geophysics the unit of choice is the submultiple nanotesla (nT); \( 1 \, nT = 10^{-9} \, T = 1 \, \gamma \), by pure coincidence; occasionally the \( \mu T \) is also used, with \( 1 \, \mu T = 10^{-6} \, T \).

When the standard geocentric spherical coordinate system is used the magnetic field elements are usually designated \( B_r, B_\theta, \) and \( B_\phi \), corresponding to locally radial, southward, and eastward unit vectors referred to a position vector \( \mathbf{r} \) on a spherical surface \( S(\alpha) \). It is generally important to account for the distinction between geocentric and geographic latitude, especially when combining surface and satellite observations. Detailed maps of the present day field show that it is a complicated function of position on the surface of the Earth although it is dominantly dipolar, and can be approximated to first order by a dipole located at the center of the Earth, with its axis tilted about \( 11^\circ \) relative to the geographic axis. The magnitude of the field, the magnetic flux density passing through Earth’s surface, is about twice as great at the poles (about \( 60 \, \mu T \)) as at the equator (about \( 30 \, \mu T \)).

The present and historical magnetic field is measured at observatories, by surveys on land and at sea, and from aircraft. Since the late 1950s a number of satellites, each carrying a magnetometer in orbit around Earth for months at a time, have provided more uniform coverage than previously possible. Early satellites only measured the magnitude of the field: however, it was shown in the late 1960s that measurements of the field’s direction are also required to specify the field accurately. Prehistoric magnetic field records can also be obtained through paleomagnetic studies of remanent magnetism recorded in rocks and archeological materials. These are useful for geomagnetic studies if there is an independent means of determining the timing that the magnetization was acquired.
Contours of the radial component of the geomagnetic field at Earth’s surface for the year 2000 are shown on Figure 1.2(a). The magnetic equator corresponds to the zero contour $B_r = 0$ and differs significantly from the geographic equator. At the magnetic poles the field is vertical (inclusion is $\pm 90^\circ$, and declination is undefined). Note that the magnetic poles are distinguished from the geomagnetic poles which correspond to
Figure 1.3: (a) Scalar magnetic field at Earth’s surface in μT and (c) its rate of change in nT/yr for the year 2000. (b) and (d) are the same with the dipole part of the field subtracted out.

the axis of the best fitting geomagnetic dipole. A different representation of the field is given in Figure 1.3 where the scalar field intensity is plotted along with its rate of change in the upper panel. In the lower part the dipole contribution to the field is omitted. From these figures we see that the field strength is lowest in the South Atlantic region, and at high latitudes it is dominated by pairs of approximately symmetric lobate structures. The non-dipole part of the field is weakest in the Pacific hemisphere. Looking at the secular variation or rate of change of the field, the largest changes are occurring over the Americas and the Atlantic and in the southern hemisphere over Africa and the Indian Ocean. Again the Pacific region shows relatively weak variations. The longevity of these features is an active area of research.

Figures 1.2 and 1.3 show the largest scale features of the internal magnetic field which originate in Earth’s core, but there are a number of different physical sources that contribute to the measured field. A whirlwind tour of spatial and temporal variations of both internal and external parts of the field is given in Chapter 1 of Foundations of Geomagnetism, by Backus, Parker and Constable (called Foundations henceforth). Figure 1.4 gives a simplified view of the parts of the magnetic field that are most important for our purposes: these can be roughly divided according to spatial scale and the frequency range in which they operate. The corresponding amplitude spectrum of variations as a function of frequency is given in Figure 1.5.

The bulk of Earth’s magnetic field is generated in the liquid outer core, where fluid flow is influenced
Earth's magnetic field is generated in the liquid outer core, where fluid flow is influenced by Earth rotation and the inner core geometry (which defines the tangent cylinder). Core fluid flow produces a secular variation in the magnetic field, which propagates upward through the relatively electrically insulating mantle and crust. Above the insulating atmosphere is the electrically conductive ionosphere, which supports Sq currents as a result of dayside solar heating. Outside the solid Earth the magnetosphere, the manifestation of the core dynamo, is deformed and modulated by the solar wind, compressed on the sunside and elongated on the nightside. At a distance of about 3 Earth radii, the magnetospheric ring current acts to oppose the main field and modulated by solar activity. Magnetic fields generated in the magnetosphere and ionosphere propagate by induction into the conductive Earth, providing information on electrical conductivity variations in the crust and mantle. Magnetic satellites fly above the ionosphere, but below the magnetospheric induction sources. The ring current and solar wind are not drawn to scale.
by Earth rotation and the geometry of the inner core. Core fluid flow produces a secular variation in the magnetic field (see Figure 1.2(b); 1.3(c), (d)), which propagates upward through the relatively electrically insulating mantle and crust. Short term changes in core field are attenuated by their passage through the mantle so that at periods less than a few months most of the changes are of external origin. At Earth’s surface the crustal part is orders of magnitude weaker than that from the core, but remanent magnetization carried by crustal rocks has proved very important in establishing seafloor spreading and plate tectonics, as well as a global magnetostratigraphic timescale. The crust makes a small static contribution to the overall field, which only changes detectably on geological time-scales making an insignificant contribution to the long period spectrum. On very long timescales (about $10^6$ years) the field in the core reverses direction, so that a compass needle points south instead of north, and inclination reverses sign relative to today’s field. The present orientation of the field is known as normal, the opposite polarity is reversed. The occurrence of reversals is unpredictable and the average rate varies with time. Figure 1.6 supplements the long period part of Figure 1.5, showing the power spectrum of dipole moment variations inferred from various kinds of paleomagnetic data.
Figure 1.6: Power spectrum of paleomagnetic dipole moment variations as a function of frequency. At longest periods the spectrum is derived from the magnetostratigraphic time scale (black, gray lines), intermediate (red, blue, orange, green, brown) are from marine sediment paleomagnetic records, and shortest periods (purple) are from archeomagnetic and lake sediment data.

Returning to Figure 1.4 we note that above the insulating atmosphere the electrically conductive ionosphere supports \( S_q \) currents with a diurnal variation as a result of dayside solar heating. Lightning generates high frequency Schumann resonances in the Earth/ionsphere cavity. Outside the solid Earth the magnetosphere, the manifestation of the core dynamo, is deformed and modulated by the solar wind, compressed on the sunside and elongated on the nightside. At a distance of about 3 earth radii, the magnetospheric ring current acts to oppose the main field and is also modulated by solar activity. Although changes in solar activity probably occur on all time scales the associated magnetic variations are much smaller than the changes in the core field at long periods, and only make a very minor contribution to the power spectrum.

The Earth’s magnetosphere plays an important role in protecting us from cosmic ray particle radiation, because the incoming ionized particles can get trapped along magnetic field lines, preventing them from reaching Earth. One consequence of this is that rates of production of radiogenic nuclides such as \(^{14}\text{C}\) and \(^{10}\text{Be}\) are inversely correlated with fluctuations in geomagnetic field intensity. This means that knowledge of Earth’s dipole moment in the past plays an important role in paleoclimate studies that use cosmogenic nuclide production to infer solar insolation during prehistoric times.
Supplemental Reading


2. CLASSICAL ELECTRODYNAMICS IN GEOMAGNETISM

As we turn to the geomagnetic part of this course we will apply many of the same mathematical tools as are used in studying Earth’s gravitational potential. However, instead of Newton’s law, the fundamental physics are described by the equations of classical electrodynamics. This chapter starts with Helmholtz’s theorem, Maxwell’s equations, and Ohm’s law in a moving medium, and motivates the equations that are used in static geomagnetic field modeling. Much of the material covered here is to be found in Chapter 2 of Foundations; a less advanced treatment is given in Chapters 2 and 4 of Blakely’s book on Potential Theory. Those needing a refresher on magnetic and electric fields could also consult 2:1 Helmholtz’s Theorem and Maxwell’s Equations

The universe of classical electrodynamics begins with a vacuum containing matter solely in the form of electric charges, possibly in motion, and electric and magnetic fields. We can detect the presence of these fields by the forces they exert on a moving point charge \( q \). If the charge \( q \) is located at position \( r \) at time \( t \) and moves with velocity \( v \) relative to an inertial frame, then

\[
\mathbf{f} = q[\mathbf{E}(r, t) + \mathbf{v} \times \mathbf{B}(r, t)].
\]  

(3)

This expression allows us in principle to measure the electric and magnetic fields using a moving charge as a detector in an inertial reference frame.

Maxwell’s equations provide the curl and divergence of the electric fields and magnetic fields in terms of other things. The reason this is useful is that Helmholtz’s Theorem tells us that if we know the curl and the divergence of a vector field, we can explicitly calculate the field itself, and furthermore, the curl and the divergence represent sources for the field, essentially creating the field. Here is Helmholtz’s theorem. A vector field \( \mathbf{F} \) in \( \mathbb{R}^3 \) which is continuously differentiable (except for jump discontinuities across certain surfaces) is uniquely determined by its divergence, its curl and jump discontinuities if it approaches 0 at infinity. The field can be written as the sum of two parts

\[
\mathbf{F} = -\nabla V + \nabla \times \mathbf{A}
\]  

(4)

where \( V \) is called a scalar potential and \( \mathbf{A} \) a vector potential. These two potentials can be explicitly computed from the following two integrals:

\[
V(r) = \frac{1}{4\pi} \int d^3s \frac{\nabla \cdot \mathbf{F}(s)}{|r - s|}
\]  

(5)

\[
\mathbf{A}(r) = \frac{1}{4\pi} \int d^3s \frac{\nabla \times \mathbf{F}(s)}{|r - s|}.
\]  

(6)

What these two equations state is that the field \( \mathbf{F} \) is generated by two kinds of sources: one is the divergence of \( \mathbf{F} \), the other its curl. Recall that in classical gravity, the gravitational potential \( V \) is generated by matter density \( \rho \) and we see this through:

\[
V(r) = -G \int d^3s \frac{\rho(s)}{|r - s|}.
\]  

(6a)
But this is just equation (5), because of Poisson’s equation, $\nabla^2 V = 4\pi G\rho$ and the fact that here $\mathbf{F} = \mathbf{g} = -\nabla V$. Helmholtz tells us that if there is no divergence or curl anywhere in space, then $\mathbf{F}$ must vanish, again confirming that $\nabla \cdot \mathbf{F}$ and $\nabla \times \mathbf{F}$ are the sources of the field. Now back to Maxwell’s equations in a vacuum.

Recall the universe we are operating in comprises an infinite vacuum containing electrical charges, represented by a local density $\rho$, which may be moving, and hence generating electric current, represented by a local current density $\mathbf{J}$. Here are Maxwell’s equations:

$$\begin{align*}
\nabla \times \mathbf{E} &= -\partial_t \mathbf{B} & \text{Faraday’s Law} \quad (7) \\
\nabla \cdot \mathbf{E} &= \rho/\varepsilon_0 & \text{Coulomb’s Law} \quad (8) \\
\nabla \times \mathbf{B} &= \mu_0 (\mathbf{J} + \varepsilon_0 \partial_t \mathbf{E}) & \text{Ampère’s Law} \quad (9) \\
\nabla \cdot \mathbf{B} &= 0 & \quad (10)
\end{align*}$$

where $\rho$ is charge density (in SI units coulombs/m$^3$), $\mathbf{J}$ is current density (amperes/m$^2$), $\mu_0$ is permeability of vacuum ($4\pi \times 10^{-7}$ henries/m), $\varepsilon_0$ is capacititivity of vacuum ($10^7/4\pi c^2$ farads/m); $\mathbf{B}$ is in teslas, and $\mathbf{E}$ is in volts/m. The quantities $\mu_0$ and $\varepsilon_0$ are exactly defined constants of the SI measurement conventions; the quantity $c$ is the velocity of light in a vacuum, which is also an exact number in SI. Notice I have introduced the somewhat unconventional, streamlined notation for time derivative: $\partial/\partial t = \partial_t$.

Viewed from the perspective of Helmholtz’s Theorem we see that the Maxwell equations (7) and (8) tells how the electric field is generated (by changing magnetic fields — Faraday’s law) or by the presence of electric charges (Coulomb’s law); and equations (9) says we can generate a magnetic field by a combination of moving charges (Ampere’s law) and by changing the electric field in time (Maxwell’s discovery, which does not have the word law associated with it). Equation (10) says there are no isolated magnetic charges, that is, no magnetic monopoles.

### 2:2 The Static Case for Geomagnetic Field Modeling

We can make use of Helmholtz’s theorem, Maxwell’s equations and the appropriate constitutive relations in describing any electromagnetic problem in geophysics. For some purposes we can neglect time variation in geomagnetic processes and imagine a system of stationary charges and steady current flows. Many geomagnetic phenomena take place over long time scales and certainly for the purposes of modeling the present internal geomagnetic field this seems like a reasonable approximation. In (7), the first of the Maxwell equations, we set $\partial_t \mathbf{B} = 0$; then the curl of the electric field vanishes. Making use of this in (4) and (6) we find that the electric field may be written as the gradient of a scalar $\phi$ (the electric potential). Thus

$$\mathbf{E} = -\nabla \phi. \quad (11)$$

Putting this together with (8) we get $\nabla^2 \phi = -\rho/\varepsilon_0$ (Poisson’s equation again, but notice the sign!) and from (5) we then get

$$\phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \int d^3 \mathbf{s} \frac{\rho(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|}. \quad (12)$$
For the magnetic field it follows from $\nabla \cdot \mathbf{B} = 0$ and (5),(4) that we can always write $\mathbf{B} = \nabla \times \mathbf{A}$. The vector field $\mathbf{A}$ is known as the magnetic vector potential. Now if we specialize to the static case with $\partial_t \mathbf{E} = 0$, we find from (9), and (6) that

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int d^3 s \frac{\mathbf{J}(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|}. \quad (13)$$

Again from (9) we have that $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$ and taking the divergence yields

$$\nabla \cdot \mathbf{J} = 0. \quad (14)$$

### 2:3 Constitutive Relations

Maxwell’s equations as written in (7)-(10) apply to a vacuum. When we need equations describing the behavior of electromagnetic fields inside a material we require some mechanism for spatial averaging of the charge and current distributions due to the atoms making up the material. This question is considered in most courses on electromagnetism, and in *Foundations*. These lead us to a form of Maxwell’s equations capable of describing field within various materials

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B} \quad (16)$$

$$\epsilon_0 \nabla \cdot \mathbf{E} = \rho - \nabla \cdot \mathbf{P} \quad (17)$$

$$\nabla \times \mathbf{B}/\mu_0 = (\mathbf{J} + \partial_t \mathbf{P} + \nabla \times \mathbf{M} + \epsilon_0 \partial_t \mathbf{E}) \quad (18)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (19)$$

where $\mathbf{P}$ and $\mathbf{M}$ are the electric polarization per unit volume and the magnetization, or magnetic polarization per unit volume of the material. Physically what happens is that the presence of an electric field (for simplicity) polarizes the material, causing charge separation. This introduces a large number of tiny electric dipoles into the medium, quantified by the term $\mathbf{P}$ – this is simply the density of electric dipole moment present in the material. If the dipole density were precisely constant, there would be no effect on $\mathbf{E}$, because the dipole fields would cancel on average (except at the ends of the specimen, where charges would accumulate). But variations in the dipole density do cause electric fields – this is seen in the fact that the term in the modified equations is $\nabla \cdot \mathbf{P}$. The magnetic effect is similar, but more complicated because electrons’ intrinsic magnetic moments and their motions within atoms cause magnetic fields.

The solution of Maxwell’s equations for $\mathbf{E}$ and $\mathbf{B}$ in a material thus requires knowledge of $\mathbf{J}$, $\mathbf{P}$ and $\mathbf{M}$ and these in turn depend on the way the material responds to the fields. These are called the constitutive relations for the material and are often determined by $\mathbf{E}$ and $\mathbf{B}$ themselves. They are not fundamental like Maxwell’s equations, but are the result of empirical observations and experiments done on different materials. The simplest possible behavior is linear. For example for many materials over a wide range of field values, we find

$$\mathbf{J} = \sigma \mathbf{E} \quad (20)$$
where $\sigma$, $\chi_E$ and $\beta$ are constants. Of course, we recognize $\sigma$ as the electrical conductivity (so that (20) is a statement of Ohm’s law, $\chi_E$ is the electrical susceptibility, and $\beta$ a kind of magnetic susceptibility.

We can simplify Maxwell’s equations by defining new fields $\mathbf{H}$ and $\mathbf{D}$,

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$$  \hspace{1cm} (23)

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M}.$$  \hspace{1cm} (24)

$\mathbf{D}$ is called the electric displacement vector. $\mathbf{H}$ has traditionally been called the magnetic field vector, while $\mathbf{B}$ was called the flux intensity or magnetic induction. The two are often confused. In view of its primary place in the theory we shall call $\mathbf{B}$ the magnetic field vector and by analogy with $\mathbf{D}$, $\mathbf{H}$ will be the magnetic displacement vector. You should be aware these names are not yet standard, but they ought to be. With these definitions in place we achieve a form of Maxwell’s equations for the second and third relations:

\[ \nabla \cdot \mathbf{D} = \rho \] \hspace{1cm} (25)

\[ \nabla \times \mathbf{H} = \mathbf{J} + \partial_t \mathbf{D}. \] \hspace{1cm} (26)

The last term is called the displacement current and is a way of generating magnetic fields without any charges having to move. This is how electromagnetic waves propagate in a vacuum. Notice that the first and last of Maxwell’s equations remain unchanged from their vacuum forms, (7) and (10). In fact we rarely use $\mathbf{D}$ in geomagnetism; one reason is that most Earth materials are not highly polarizable, and another is that we almost always drop the term involving $\mathbf{D}$ in (26) as we shall see next.

### 2:4 Application to the Geomagnetic Field

A reasonable approximation in geophysical problems is to neglect the displacement current $\partial_t \mathbf{D}$ in (26). This can be shown by a crude dimensional analysis as follows. (For more details see *Foundations*, Section 2.4) Take the time derivative of (26), and insert Ohm’s law (20); for simplicity assume $\chi_E$ and $\beta$ in (21)-(22) are negligible; then (26) becomes:

\[ \nabla \times \nabla \times \mathbf{E} + \frac{\mu_0 \varepsilon_0}{\mu_0} \partial_t \mathbf{E} - \frac{\mu_0 \varepsilon_0 \partial_t^2 \mathbf{E}}{\mu_0} = 0. \] \hspace{1cm} (28)

Now we use $\nabla \times \mathbf{E} = -\partial_t \mathbf{B}$ (16) and rearrange slightly

\[ \nabla \times \nabla \times \mathbf{E} + \mu_0 \sigma \partial_t \mathbf{E} + \mu_0 \varepsilon_0 \partial_t^2 \mathbf{E} = 0. \] \hspace{1cm} (28)

We would like to estimate the approximate size of each of the three terms in (28). If we assume length scales of variation are $L$ or larger and time scales $T$ or larger, very roughly we can replace space derivatives by $1/L$ and time derivatives by $1/T$; then

\[ 0 = [1 + \mu_0 \sigma (L^2/T) + \mu_0 \varepsilon_0 (L/T)^2] |\mathbf{E}|. \] \hspace{1cm} (29)
Because $\mu_0\epsilon_0 = 1/c^2$, where $c$ is the speed of light, the last term represents the ratio of typical speeds in the system over $c$ squared. In geomagnetism scales are typically many thousands of kilometers, and time scales can be as low as minutes, but may be years: even for $L = 10^4$ km and $T = 10$ s, the last term in (29) is $10^{-5}$. The term with conductivity is much larger than this in the interior, say $\sigma \approx 10^{-3}$ S/m, then the second term is roughly $4\pi \times 10^{-7} \times 10^{-3} \times 10^{12}/10$ or about 120. So displacement current is unimportant and the balance is between the first two terms. The four equations (7)-(10), or the set valid within material (7), (10), (25), (26), in which the displacement current is neglected ($\partial_t E$ or $\partial_t D$) are sometimes referred to as the pre-Maxwell equations.

But in the atmosphere $\sigma$ is so small, the second term is small too. When this happens we see the simple-minded analysis breaks down and the discover that the size of $\nabla \times \nabla \times E$ cannot be $|E|/L^2$ – terms in the spatial derivative cancel among themselves and the corresponding term in (27) vanishes by itself.

The magnetic field can always be written as the curl of a vector potential (because of Helmholtz’s Theorem, (4)-(6), and $\nabla \cdot B = 0$). In certain circumstances there is an alternative representation in terms of a scalar potential for $B$. In our application to the geomagnetic field we will make the approximation that Earth’s atmosphere is an insulator with no electrical currents (actually $\sigma \approx 10^{-13}$ S/m close to the ground so $J = 0$ seems like a reasonable approximation). The atmosphere is also only very slightly polarizable magnetically so we can set $M = 0$, thus within the atmospheric cavity we find the essential content of (26) is

$$\nabla \times B = 0.$$  \hspace{1cm} (30)

(30) tells us that $B$ can be written as the gradient of a scalar because when $B = -\nabla \phi$, (30) is automatically satisfied (recall $\nabla \times \nabla = 0$). Since $B$ is also solenoidal (divergence free) from (19), the scalar potential $\phi$ is harmonic: $\nabla^2 \phi = 0$. This why we can use so much of our gravity machinery in geomagnetism, for example, spherical harmonics.
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 \(\Psi\) is harmonic

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

From equation (I-8.7, p.31 of gravity notes) we know that \(\Psi\), 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 \(\nabla \cdot \mathbf{B} = 0\) 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 \(\Psi\) on two spherical surfaces we could get two independent equations for each of the \(\Psi_{l}^{m}\) 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 \(\Psi\), but \(\mathbf{B}\). Let’s assume that \(\mathbf{B}\) is known everywhere on the surface of the sphere \(r = a\), and we write

\[
\mathbf{B} = \hat{r} B_{r} + \mathbf{B}_{s} \tag{34}
\]

where \(\hat{r} \cdot \mathbf{B}_{s} = 0\), 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 \(\nabla_{1}\) as follows

\[
\nabla = \hat{r} \partial_{r} + \frac{1}{r} \nabla_{1}. \tag{35}
\]
We first find $B_r$ on $r = a$ from $\Psi$:

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

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

$$f(\theta, \phi) = \sum_{l,m} c^m_l Y^m_l(\theta, \phi) \tag{37}$$

then using the orthonormality of the $Y^m_l$, that is, by

$$\int_{S^1} Y^m_l(\theta, \phi) Y^m_l(\theta, \phi)^* d^2 \hat{r} = \delta^m_l \delta^m_m \tag{38}$$

the coefficients $c^m_l$ are just

$$c^m_l = \int f(\theta, \phi) Y^m_l(\theta, \phi)^* d^2 \hat{r}. \tag{39}$$

From this it follows that

$$lA^m_l - (l + 1)B^m_l = - \int B_r Y^m_l(\theta, \phi)^* d^2 \hat{r}. \tag{40}$$

We can thus find the above linear combination of $A^m_l$ and $B^m_l$ 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 $B_s$. Again using the spherical harmonic expansion for $\Psi$ on $r = a$ we find

$$B_s = - \sum_{l,m} [A^m_l + B^m_l] \nabla_1 Y^m_l. \tag{41}$$

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

$$\int_{S^1} \nabla_1 Y^m_l \cdot (\nabla_1 Y^m_l)^* d^2 \hat{r} = l(l + 1) \delta^m_l \delta^m_m. \tag{42}$$

Dot $(\nabla_1 Y^m_l)$ into our equation for $B_s$ on $r = a$, and integrate over the sphere:

$$\int B_s \cdot (\nabla_1 Y^m_l)^* d^2 \hat{r} = - \sum_{l,m} [A^m_l + B^m_l] \int \nabla_1 Y^m_l \cdot (\nabla_1 Y^m_l)^* d^2 \hat{r}. \tag{43}$$

Thus we have

$$A^m_l + B^m_l = - \frac{\int B_s \cdot (\nabla_1 Y^m_l)^* d^2 \hat{r}}{l(l + 1)}. \tag{44}$$

Combining this with the equation derived from $B_r$ we can always recover the $A^m_l$ and $B^m_l$ separately from our knowledge of $B$ on $r = a$, except for $l = 0$ (Why?). Thus $\Psi$ is determined to within an additive constant and $B$ is determined uniquely within $S(a, b)$ by our knowledge of $B$ 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 $B$ all over the surface of the Earth. As we shall see later the estimation of the $A^m_l$ and $B^m_l$ from actual magnetic field measurements does not rely on knowing $B$ 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 $A^m_l$ 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 $B \cdot \hat{r}$ everywhere on a sphere $S(a)$ containing the sources $(J, M, \text{current and magnetization})$. 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.

$$B = -\nabla \Psi$$

and

$$\nabla^2 \Psi = 0.$$  \hspace{1cm} (46)

**Exercise:**

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

Using the result of the above exercise, we can define a potential function $\Omega = r B_r$, 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 $\Omega$ in terms of a spherical harmonic expansion

$$\Omega(r, \theta, \phi) = r B_r(r, \theta, \phi) = a \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \beta_m^l \left( \frac{a}{r} \right)^{l+1} Y_l^m(\theta, \phi).$$  \hspace{1cm} (47)

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

$$\beta_m^l = \int_{S(a)} B_r(a, \theta', \phi') Y_l^m(\theta', \phi') \frac{d^2r'}{a^2}.$$  \hspace{1cm} (48)

Knowing the $\beta_m^l$ we can find $B_r$ anywhere (47) converges.

17
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 $\Omega$ 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 $\Omega$ has a surface harmonic expansion $\Omega_{lm}(r)$. It is clear from (47) that

$$\Omega_{lm}(r) = \beta_{lm} \left( \frac{a}{r} \right)^{(l+1)\frac{1}{2}}.$$  \hspace{1cm} (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 $B_r$

$$B_r(r, \vec{r}) = \int_{S(a)} (a/r)^2 K(\cos \theta) B_r(a, \vec{r}) d^2 \vec{r}.$$  \hspace{1cm} (50)

where

$$K(x, \cos \theta) = \frac{1}{4\pi} \frac{1 - x^2}{(1 + x^2 - 2x \cos \theta)^{3/2}} - 1.$$  \hspace{1cm} (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](image-url)

**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

$$|B_r(r, \vec{r})| \leq \int_{S(a)} |(a/r)^2 K(\cos \theta) B_r(a, \vec{r})| |B_r(a, \vec{r})|_{\text{max}} d^2 \vec{r}.$$  \hspace{1cm} (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 $x = a/r$:

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

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

\[ |B_r(r, \hat{r})| < |B_r(a, \hat{r}')|_{\text{max}}, \quad r > a. \] (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 $\eta$, 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 $\eta$ 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:

$$re_r(a, \theta, \phi) = \alpha Y^{-1}_{l,l}(\theta, \phi)$$

$$= \alpha N_{l,l} \sin^l \theta \mathrm{e}^{-i l \phi} \quad (55)$$

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 $\eta$ 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 $\mu T$), 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: $B_r$, radial component of magnetic field in $\mu$T, 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, and in Section 3.1 we saw how in principle this allows unique separation of the internal and external field contributions to spherical harmonic expansions. Continuing with the specialization to the internal field coefficients begun in Section 3.2, we note that the field coefficients for $\Omega(r, \theta, \phi) = r B_r(r, \theta, \phi)$ can, in principle be derived from equation (48):

$$\beta_{lm}^n = \int_{S(a)} B_r(a, \theta, \phi) Y_{lm}^n(\theta, \phi) \frac{d^2 r}{a^2}$$

but this evidently requires knowledge of $B_r$ all over $S(a)$, which is not possible in practice. We will need a strategy for dealing with incomplete and noisy observations. But before we tackle the question of finding approximate models, (and here we will 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.

**Geomagnetic Elements**

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

- $B_r, B_\theta, B_\phi$ – 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.
- $H = (X^2 + Y^2)^{1/2}$ – horizontal magnetic field intensity
- $B = (B_\theta^2 + B_\phi^2 + B_r^2)^{1/2} = (X^2 + Y^2 + Z^2)^{1/2}$ – 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 we can write

$$X \approx -B_\theta; \quad Y \approx B_\phi; \quad Z \approx -B_r.$$  \hspace{1cm} (57)

But 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, $e$

$$e^2 = \frac{(a^2 - b^2)}{a^2}$$ \hspace{1cm} (58)

and flattening

$$f = \frac{(a - b)}{a} = 1/298.257.$$ \hspace{1cm} (59)
In paleomagnetic field modeling it’s probably not necessary to take account of the difference between geodetic and geocentric latitude but in modern field modeling the data may be accurate enough for it to really matter.

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?

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 has shown that this is not the case.

3.4.1 Limitations on Uniqueness

Non-Uniqueness and $|\mathbf{B}|$

The fact that $B_r$ on a surface $S(\alpha)$ 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 complete measurements of $|\mathbf{B}|$ (the magnitude of the field) would uniquely specify the internal 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. 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 $\Psi_M$ that is everywhere orthogonal to a dipole field; then $\Psi_D + \Psi_M$ and $\Psi_D - \Psi_M$ have the same $B$ everywhere.

One might reasonably ask whether this counter-example matters or is merely an artificial mathematical construct. If it does not one would surely be better off using scalar measurements (with typical accuracy of < 1nT) than vector components which are usually several nT, because of the need to orient a low earth orbiting measurement platform within an Earth centered reference frame. However the effect is real. Independent field models using exclusively vector or scalar Magsat data predict similar intensity at Earth’s surface, but show differences of up to 2000 nT in the vector components. This reflects the fact that optimization in the $|\mathbf{B}|$ model is insensitive to errors perpendicular to the field, while the full vector is sensitive misfit in all directions. This perpendicular error effect is also referred to as the Backus effect.

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 |B| everywhere then uniqueness of the solution is guaranteed. The practical usefulness of this result is debatable, and full field B measurements are now routine.

Non-Uniqueness and B

Another potentially interesting case is exact knowledge of the direction of the field all over Earth’s surface, S(a), which one might imagine could 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. Similarly, (despite an earlier inaccurate argument to the contrary) it was shown in 1990 by Proctor and Gubbins, (Geophys. J. Internat. 100, pp 69-79, 1990) that \( \hat{B} \) on a spherical surface does not determine the field to within a scalar multiple.

The counter examples are again questionable representatives of Earth-like fields, being axisymmetric, antisymmetric wrt the Equator, and octupolar in type. This result too has been investigated further and it has now been shown that if the magnetic field only has two poles on the surface \( S(a) \) 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).

For much more discussion of uniqueness questions, including some extensions to the case where both external and internal field are considered see Sabaka et al. (2010, Handbook of Geomathematics, doi: 10.1007/978-3-642-01546-5_17). The essential result of importance for geomagnetic field modeling using observations from satellites and geomagnetic observatories is summarized in their Figure 4 and reproduced in Figure 3.4.1.1 below. Despite intrinsic non-uniqueness one can recover sufficient information for field modeling to be a useful activity.
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 $B_i = B \cdot \hat{r}_i$. 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 $B_i(a, \theta, \phi)$ 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

$$\Psi(r, \theta, \phi) = a \sum_{l=1}^{L} \sum_{m=-l}^{l} b_{lm} \left( \frac{a}{r} \right)^{l+1} Y_{lm}(\theta, \phi)$$

and we find

$$B = -\nabla \Psi.$$  

If we measure orthogonal components of the field $B_r, B_\theta, B_\phi$ we have a set of observations at $N$ sites designated $r_j$

$$d_j = \hat{s}_j \cdot B(r_j), \quad j = 1, ..., N$$

$$= \sum_{l=1}^{L} \sum_{m=-l}^{l} b_{lm} \left( \frac{a}{r_j} \right)^{l+1} \hat{s}_j \cdot \nabla \left[ \frac{Y_{lm}(\hat{r}_j)}{r_j^{l+1}} \right] + \epsilon_j. \quad (62)$$

The data $d_j$ are then linear functionals of the $b_{lm}$ specifying the field. With an appropriate indexing scheme for the $b_{lm}$ we can write a prediction for our observations $d_j$ as a matrix equation.

$$d = Gb + n \quad (63)$$

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

$$b = \begin{bmatrix} b_1^{-1} \\ b_1^0 \\ b_1^1 \\ \vdots \\ b_L^L \end{bmatrix}.$$  \hspace{1cm} (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 $b$ to get an exact match to the data, and so $n$ is not a vector of zeros. Least squares estimation involves finding the values for $b$ that minimize $||n||^2 = ||d - G \cdot b||^2$, 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:

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

Note that (65) is for several reasons not a good way to find $\hat{b}$ 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

$$E[||n||^2] = (N - K)\sigma^2$$  \hspace{1cm} (66)

where $\sigma^2$ 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:

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

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:

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

subject to

\[
\sum_{j=1}^{N} \left( \frac{f(x_j) - y_j}{\sigma_j^2} \right)^2 = T.
\]

The size of the Lagrange multiplier \(\lambda\) is dictated by the constraint that the data be fit to a reasonable tolerance level. If the uncertainty in the observations, \(\sigma_j\), is known one reasonable choice for \(T\) is the expected value of \(\chi^2_N\), 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

\[
E = \int_{r>a} |B|^2 d^3r.
\]

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

\[
E = \int \nabla \Psi \cdot \nabla \Psi d^3r.
\]

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

\[
E = \int [\nabla \cdot (\Psi \nabla \Psi) - \nabla^2 \Psi] d^3r
\]

\[
= \int \nabla \cdot (\Psi \nabla \Psi) d^3r.
\]

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

\[
E = \int_{S(a)} -\Psi \frac{\partial \Psi}{\partial r} d^2r.
\]
Now we simply substitute the spherical harmonic expansion (60) for the potential $\Psi$.

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

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

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

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

$$||B||_2^2 = \sum_{l=1}^{\infty} w_l \sum_{m=-l}^{l} |b_l^m|^2, \quad w_l > 0. \quad (75)$$

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

$$\int_{S(o)} B \cdot B d^2 \hat{r} \quad w_l = (2l + 1)(l + 1) \quad (76)$$

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

$$\int_{S(o)} [\nabla_2^2 \hat{r} \cdot B]^2 d^2 \hat{r} \quad w_l = l^2(l + 1)^4 \quad (78)$$

$$\int_{r<a} J_T \cdot J_T d^3 r \quad w_l = a^3 \mu_0^2 (l + 1)(2l + 1)^2(2l + 3). \quad (79)$$

In (79) $J_T$ 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

$$\int_{S(1)} (Y_l^m)^2 d^2 \hat{r} = \frac{4\pi}{2l + 1}. \quad (80)$$
Then the scalar potential is written

\[ \Psi(r, \theta, \phi) = a \sum_{l=1}^{\infty} \left( \frac{a}{r} \right)^{l+1} \sum_{m=0}^{l} N_{lm} (g_{lm}^m \cos m \phi + h_{lm}^m \sin m \phi) P_l^m(\cos \theta) \]  

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

\[ N_{lm} = \begin{cases} 1, & m = 0 \\ \sqrt{\frac{2(l-m)!}{(l+m)!}}, & m > 0. \end{cases} \]

You will find the numbers \( g_{lm}^m, h_{lm}^m \) 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 13 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, onto degree and order 10. The values are in nanotesla. The most recent version of IGRF (2015-2020) can be found online at http://www.ngdc.noaa.gov/IAGA/vmod/igrf.html. 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. IGRF models come with a Health Warning about their limitations. Recent versions extend to degree and order 13, but they do not include crustal anomalies.

| \( l \) | \( m \) | \( g_{lm}^m \) | \( h_{lm}^m \) | \( l \) | \( m \) | \( g_{lm}^m \) | \( h_{lm}^m \) | \( l \) | \( m \) | \( g_{lm}^m \) | \( h_{lm}^m \) | \( l \) | \( m \) | \( g_{lm}^m \) | \( h_{lm}^m \) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 0 | -29615 | 0 | 6 2 | 74 | 64 | 9 0 | 5 | 0 | 9 1 | 9 | -20 | 2 0 | -2267 | 0 | 6 4 | -5 | -61 |
| 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 | 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 | 30 |
Field values in microteslas

Figure 3.4.5
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 of 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

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

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

(83)

(84)

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

$$\frac{2\mu_0 E(a)}{a^2} = \sum_{l,m} \sum_{l',m'} [(l + 1)(l' + 1)b^m_l(b^{m'}_{l'})^* \int_{S(1)} Y^m_l(Y^{m'}_{l'})^* \, d^2\mathbf{s} + b^m_l(b^{m'}_{l'})^* \int_{S(1)} \nabla Y^m_l \cdot (\nabla Y^{m'}_{l'})^* \, d^2\mathbf{s}]$$

(85)

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

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

$$= \sum_{l=1}^{\infty} (2l + 1)(l + 1) \sum_{m=-l}^{l} |b^m_l|^2$$

(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

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

(87)

where we define

$$R_l = \frac{(2l + 1)(l + 1)}{4\pi} \sum_{m=-l}^{l} |b^m_l|^2 = (l + 1) \sum_{m=0}^{l} [(g^m_l)^2 + (h^m_l)^2].$$

(88)

Here $R_l$ is called the geomagnetic spectrum or the Lowes spectrum (or even the Mauersberger-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 Olsen et al., 2009 (Geophys. J. International, vol. 179, p. 1477-87, doi: 32.
Fig. 3.4.6.1
Spatial power spectrum of the geomagnetic field at the Earth’s surface. Black dots represent the spectrum of a recent field model (Olsen et al. 2009; Maus et al. 2008). Also shown are theoretical spectra (Voorhies et al., 2002) for the core (blue) and crustal (magenta) part of the field, as well as their superposition (red curve).

10.1111/j.1365-246X.2009.04386.x, 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 $l \leq 14$ come almost entirely from the core, those with $l > 14$ 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

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

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 up to about $l = 14$. A flat spectrum is one in which there is equal energy at every scale, the sort of 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, I will draw attention to the use of the spatial power spectrum in building statistical models of paleosecular variation - known as Giant Gaussian process models. 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 $b^m_l$ 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 which exhibits significant changes over time. Subsequent work has shown that the true field has some definite departures from the uniform Gaussian model, but the Giant Gaussian Process remains a useful framework for describing expected averages and statistical variability produced by secular variation at any location. Such statistical behavior should be predicted by Earth-like numerical simulation of the geodynamo.
3:5 Toroidal and Poloidal Fields

Now we turn to a brief discussion of a fundamental limitation of observations made outside the magnetic source region in Earth’s core. Before we start, let us assemble two useful vector calculus identities, numbers 5 and 9 in our list:

\[ \nabla \times (\mathbf{U} \mathbf{s}) = (\nabla \times \mathbf{U}) \mathbf{s} - \mathbf{U} \times \nabla \mathbf{s} \]  
(90)

\[ \nabla \times \nabla \times \mathbf{U} = -\nabla^2 \mathbf{U} + \nabla(\nabla \cdot \mathbf{U}). \]  
(91)

Our magnetic field modeling so far has only dealt with solutions of Laplace’s equation in regions free of sources. In an earlier lecture we showed that the magnetic field can always be written as the curl of the magnetic vector potential \( \mathbf{A} \), i.e., \( \mathbf{B} = \nabla \times \mathbf{A} \).

Now we consider a sphere of radius \( c \) (Earth’s core) surrounded by an insulator and divide the vector potential into parts parallel to and perpendicular to \( \mathbf{r} \) by writing

\[ \mathbf{A} = \mathbf{T} \mathbf{r} + \nabla \times \mathbf{P} \mathbf{r} = \mathbf{T} \mathbf{r} + \nabla \times (\mathbf{P} \mathbf{r}) \]  
(92)

where \( \mathbf{T} \) and \( \mathbf{P} \) are scalar functions of \( \mathbf{r} \), 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 \( \mathbf{B} \) we take the curl:

\[ \mathbf{B} = \nabla \times (\mathbf{T} \mathbf{r}) + \nabla \times \nabla \times (\mathbf{P} \mathbf{r}) \]

\[ = \mathbf{B}_T + \mathbf{B}_P \]  
(93)

and \( \mathbf{B}_T \) is called the toroidal part of \( \mathbf{B} \), while \( \mathbf{B}_P \) is the poloidal part. This decomposition for \( \mathbf{B} \) 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 \( \mathbf{B} \) in an insulator; when \( \mathbf{J} \) does not vanish, we need two scalars, not one to describe \( \mathbf{B} \) completely. Conventionally, the scalars are always restricted to a class of functions whose average value over every sphere is zero, that is

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

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

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

\[ \mathbf{B}_T = (\nabla \times \mathbf{r}) \mathbf{T} - \mathbf{r} \times \nabla \mathbf{T} \]

\[ = -\mathbf{r} \times \nabla \mathbf{T} \]  
(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) \( J = 0 \) and

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

Now \( B \) is continuous at \( S(c) \) and since \( \hat{r} \cdot B_T = 0 \) just inside we conclude that \( \hat{r} \cdot B_T \) 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 \( B_T \) vanishes outside the conducting sphere. Hence the toroidal part of \( B \) in Earth’s core is invisible outside the core and only the poloidal part, \( B_P \), 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:

\[
\nabla \times B = \mu_0 J \tag{97}
\]

As we saw earlier, taking the divergence yields that \( J \) is solenoidal

\[
\nabla \cdot J = 0 \tag{98}
\]

and thus we can also find a unique decomposition into toroidal and poloidal parts for the current density

\[
J = J_T + J_P. \tag{99}
\]

By substituting for \( B \) in terms of \( B_T \) and \( B_P \) in (93) we will show that \( B_P \) comes from \( J_T \), that is, poloidal fields are generated by toroidal currents and similarly that \( B_T \) comes from \( J_P \), so toroidal fields are generated by poloidal current systems.

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

\[
\mu_0 J = \nabla \times B = \nabla \times \nabla \times (Tr) + \nabla \times \nabla \times \nabla \times (Pr). \tag{100}
\]

It is obvious from the first term on the right that the toroidal part of \( B \) is associated with a poloidal current, whose scalar is just \( T/\mu_0 \). To show the last term, with three curls, is in fact toroidal, we analyze \( S = \nabla \times \nabla \times (Pr) \). From the identity (91)

\[
S = -\nabla^2(rP) + \nabla(\nabla \cdot (rP)). \tag{101}
\]

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

\[
\nabla^2(rP) = \partial_j \partial_j (x_k P) = \partial_j ((\partial_j x_k) P + x_k \partial_j P) = \partial_j (\delta_{jk} P + x_k \partial_j P) \tag{102}
\]

\[
= \partial_k P + \delta_{jk} \partial_j P = \partial_k P + (\delta_{jk} \partial_j P) + x_k \partial_j \partial_j P = \partial_k P + x_k \partial_j \partial_j P = 2 \partial_k P + x_k \partial_j \partial_j P \tag{103}
\]

\[
= 2 \nabla P + r \nabla^2 P.
\]

36
Then (101) becomes

\[ S = -r \nabla^2 P + \nabla(\nabla \cdot (rP) - 2P) \]
\[ = -r \nabla^2 P + \nabla(r \partial_r P + P). \] (104)

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

\[ \nabla \times S = -\nabla \times (r \nabla^2 P) \]

and we recover (100)

\[ \mu_0 J = -\nabla \times (r \nabla^2 P) + \nabla \times \nabla \times (Tr) = \mu_0 (J_T + J_P). \] (105)

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 4\pi in the bounds given there):

\[ \mu_0^2 \int_{r<c} J_T \cdot J_T d^3 r \geq c \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \frac{(l+1)(2l+1)^2(2l+3)}{l} |b_{lm}|^2 \] (106)
where the coefficients $b_{lm}$ 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:

$$Q = \int_{r < c} \frac{\mathbf{J}_T \cdot \mathbf{J}_T}{\sigma} d^3r \geq \frac{c}{\mu_0 \sigma} \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \frac{(l + 1)(2l + 1)^2(2l + 3)}{l} |b_{lm}|^2. \quad (107)$$

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 $\mathbf{B}$ in the core is large compared with the poloidal part. Some models predict $|B_T| \sim 50 \times |B_P|$. 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 $\mathbf{B}_T$?
4. MAGNETOHYDRODYNAMICS IN THE CORE

Observations of the geomagnetic field show that the internal part changes on times scales ranging from about a year upwards. It is generally believed that secular variation at shorter time scales will be greatly attenuated by passage through the Earth’s mantle which is only approximated by an insulator (perhaps more on this later). Observationally, the secular variation is usually distinguished from the more rapidly varying external field by its time spectrum. However, it should be noted that the external field generated by magnetic storms does not average to zero, and the number of these storms is greater at times of sunspot maximum than at sunspot minimum. Since the sunspot cycle has an approximate periodicity of 11 years there is some overlap in the time spectra of internal and external field variations. Geomagnetic jerks, which are defined by very rapid changes in the second derivative of the field with time, are also believed to be of internal origin.

In the magnetic field models that we looked at in an earlier lecture the temporal variation is incorporated by allowing the Gauss coefficients to vary as a function of time; the temporal variation of the $g^m_l$, $h^m_l$ can be constrained to be smooth through a parameterization in terms of cubic or higher order splines and the minimization of a penalty function like (68) involving the second temporal derivative of $B$, as well as some penalty on spatial structure, like (76)-(79). Spherical harmonic models now extend back to 100 ka, although the level of detail in paleofield models is much less than in those constructed from modern observations with good spatial coverage (see Treatise on Geophysics, volume 5, Chapter 9 for a review). Further back in time it is possible to reconstruct time variations in the dipole moment back to about 2 Ma, and intermittently at more remote times in the geological past. Because of lack of detailed age constraints in paleomagnetic work statistical representations of geomagnetic field variability are useful for earlier times. The occurrence times for geomagnetic reversals are well documented back to about 200 Ma via marine magnetic anomalies, and there have been numerous attempts to link changes in reversal frequency to long term changes in field strength.

Ultimately such observations and the models derived from them are used to try and make inferences about the state of Earth’s core and the way the geodynamo operates. This requires that we go back to the physical cause of the magnetic field, and look at magnetohydrodynamics (MHD); these are the dynamics of fluids in which electromagnetic forces are important.

For our purposes this means looking at the time evolution of the magnetic field in the fluid part of Earth’s core, where the fluid has non-zero conductivity and is permeated by a magnetic field. The temporal variation is described by the pre-Maxwell equations:

$$\nabla \cdot \mathbf{B} = 0 \quad (108a)$$

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B} \quad (108b)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \quad (108c)$$

We have assumed that there is no permanent magnetization in the core so that in the last equation $\mathbf{H} = \mathbf{B}/\mu_0$. In static
media we have Ohm’s law $\mathbf{J} = \sigma \mathbf{E}$ but in a fluid moving with respect to the (inertial) frame of reference with velocity $\mathbf{u}$

$$\mathbf{J} = \sigma (\mathbf{E} + \mathbf{E}')$$  \hspace{1cm} (109)

where $\mathbf{E}' = \mathbf{u} \times \mathbf{B}$. You can see this might be correct from the force on a moving charge (3). For a derivation, consult *Foundations*. So the modified Ohm’s law is

$$\mathbf{J} = \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}).$$  \hspace{1cm} (110)

Setting this result into (108c), we find

$$\nabla \times \mathbf{B} = \mu_0 \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}).$$  \hspace{1cm} (111)

Taking the curl of this equation and assuming that $\sigma$ is uniform throughout the core, we get

$$\nabla \times \nabla \times \mathbf{B} = \mu_0 \sigma [\nabla \times \mathbf{E} + \nabla \times (\mathbf{u} \times \mathbf{B})].$$  \hspace{1cm} (112)

Recall the vector identity (91) which we apply to $\mathbf{B}$:

$$\nabla \times \nabla \times \mathbf{B} = -\nabla^2 \mathbf{B} + \nabla (\nabla \cdot \mathbf{B}).$$ \hspace{1cm} (113)

But $\mathbf{B}$ is solenoidal – equation (108a) – so the second term vanishes. We put (113) into (112). Next we appeal to Maxwell’s equation (108b) to eliminate $\mathbf{E}$, and (112) becomes after some rearrangement:

$$\partial_t \mathbf{B} = \eta \nabla^2 \mathbf{B} + \nabla \times (\mathbf{u} \times \mathbf{B}).$$  \hspace{1cm} (114)

This is called the *magnetic induction equation* for the geodynamo. Here $\eta = 1/\mu_0 \sigma$ is called the *magnetic diffusivity* – it has the usual units for a diffusion constant namely, m$^2$/s. $\eta$ is the analog of the kinetic viscosity, $\nu$ in ordinary fluid flow.

For the Earth’s core recent revisions to the electrical conductivity (see Pozzo et al., 2012, doi:10.1038/nature11031) yield a $\sigma \approx 1.1 \times 10^6$ S/m, and so $\eta \approx 0.7$ m$^2$/s under core conditions.

Jacobs’ book *Geomagnetism, Vol 2*, Chapter 1, pp 1-23 gives a good description of the equations required to describe the basic state of the core. These include the magnetic induction equation (114) above and the nature of the magnetohydrodynamic approximation leading to it. In addition to this we need an equation describing the dynamics of the flow. This is the Boussinesq Navier-Stokes equation which assumes that the fluid is incompressible except for thermal expansion. The frame of reference is fixed in and rotating with the earth’s mantle. Then the equation of motion for a volume element within the fluid is

$$\rho_0 (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + 2 \Omega \times \mathbf{u}) = -\nabla p + \rho' \mathbf{g} + \mathbf{J} \times \mathbf{B} + \rho_0 \nu \nabla^2 \mathbf{u}$$  \hspace{1cm} (115)

where $\rho_0$ is the hydrostatic density

$\rho'$ is departure from hydrostatic density

$\Omega$ is the angular velocity vector of Earth’s rotation

$p$ is the non-hydrostatic pressure
\( g \) is gravitational acceleration

\( J \) is current density

\( J \times B \) is called the Lorentz force.

You will recognize in (115) the terms on the left correspond to mass times acceleration, and those on the right to a sum of forces, in a statement of Newton’s second law of motion for an element of the fluid. Equations (114) and (115) are the fundamental equations of magnetohydrodynamics. In solving the geodynamo problem one can identify three important aspects.

(1) Energy source: we need a source of energy for convection to occur in Earth’s core. This source must be sufficient to overcome diffusion, not just at the present time, but to maintain the dynamo throughout Earth history. In most theories the driving term appears in additional equations that define how \( \rho' \), the density differential is governed; but in some the driving force enters through boundary conditions.

(2) Kinematic dynamo problem: this is basically the question of whether a particular fluid motion, that is \( u \), is capable of generating a magnetic field. The problem can be studied mathematically using the magnetic induction equation (114), \( \nabla \cdot B = 0 \) and appropriate boundary conditions for the geomagnetic field. A candidate fluid velocity \( u \) is tested for dynamo action by solving (114) for \( B \) and testing to see if the solution grows with time. The fluid velocity needs to be sufficiently large to overcome the effects of diffusion as well as having the right form to cause its regeneration. Two important types of kinematic dynamos are \( \alpha - \omega \) models and \( \alpha^2 \) models which we discuss later. In the first the \( \omega \) effect is strong, core fluid motion is influenced by the Coriolis force, and is a combination of differential rotation and convective helical motion. In \( \alpha - \omega \) type models the toroidal field is expected to be much stronger than the poloidal contribution, but when the fields are comparable the dynamo is an \( \alpha^2 \) type.

(3) Dynamical dynamo problem: the feasibility of kinematic dynamo problems has now been amply demonstrated, and in dynamo theory the emphasis is now on MHD, that is studying the dynamics of the fluid flow. The full dynamo problem requires the solution of (114), (115) in three-dimensions and an appropriate equation of state for the core with suitable boundary conditions. This is a formidable problem, currently only approachable by computer modeling, and even then not with parameters accurately reflecting the conditions inside the core. For example the Ekman number, which measures the relative importance of viscosity and the Coriolis force, is usually larger than \( 10^{-5} \) in simulations: but the molecular value is \( \approx 10^{-15} \) and even the turbulent value is \( \approx 10^{-9} \). None-the-less the solutions are suggestive and resemble the real system in many ways. Well known models by Glatzmaier, G. A, and Roberts, P. H., *Phys. Earth Planet. Inter.*, 91, pp 63-75, 1995 were the first to spontaneously exhibit geomagnetic reversals among other Earth-like properties. For animations and other good stuff see these websites:

http://www.psc.edu/science/glatzmaier.html

http://es.ucsc.edu/~glatz/geodynamo.html
Of the above list (2) and (3) each involve mathematical analysis of simplified models of Earth’s core, while (1) is fundamentally a thermodynamic problem. Much of (1) depends on ideas about Earth’s history and constitution which lie outside the scope of this course; despite these uncertainties (2) and (3) may be pursued essentially independently of the exact nature of the energy source.

4:1 Energy Sources for the Geodynamo

The amount of energy actually required to drive the geodynamo is at best poorly understood. The magnitude of ohmic dissipation depends on the electrical conductivity and the strength of the magnetic field in the core, and this must be made up by conversion from kinetic energy of the flow. The poloidal field at the base of the mantle is of the order of 500 $\mu$T, but this may not reflect the strength in the interior of the core. Surface observations place a lower bound on heating of $10^8$ W, but strong toroidal fields may make $10^{12}$ W a more realistic estimate.

Two possible mechanisms can drive the flow, (1) internal buoyancy forces and (2) external forcing by boundary motion. Suggested energy sources have been:

1. Thermal buoyancy from heat sources in the core: these are cooling of the core, release of latent heat during freezing of the outer core to form the inner core, and contributions from radioactive decay. The significance of the latter is disputed.

2. Compositional buoyancy is expected because light constituents in the core are released at the inner core boundary, as the heavy fraction is preferentially incorporated into the solid part.

3. In a Gravitationally powered dynamo the energy source is gravitational potential energy stored in the outer core, which is released as Earth cools and the inner core grows.

4. Precessional driving of core flows due to gravitational torques from the Sun and Moon. Precession of the earth can produce fluid instability (and thus flow?), but this requires momentum transfer from mantle to core. Nevertheless this has not been completely discounted as a partial energy source.

Most numerical simulations rely on a cooling Earth model, with both compositional and thermal buoyancy effects incorporated into a so-called co-density. For a recent review of progress in understanding the origin of Earth’s magnetism see Roberts & King (2013, Rep. Prog. Phys., 76, doi:10.1088/0034-4885/76/9/096801) We will return briefly in 4.6 to the geodynamo after we have studied the behavior of a magnetic field in a conducting fluid.

4:2 Secular Variation

Secular variation is the name given to the changes in the geomagnetic field with timescales of the order of a few years to many centuries. As we have seen there are a number of regular features in the secular variation, such as the westward
drift, that ought to be capable of theoretical explanation. Westward drift is illustrated in Figure 4.2 below, which shows the nondipole field at two times, exactly 10 years apart. The fields are contoured at 5 $\mu$T intervals. Notice there is good evidence for a general westward drift in the Atlantic ocean, but little sign over Siberia or the central Pacific oceans.

![Nondipole $B_r$: IGRF 1990 (dashed) and 2000](image)

**Figure 4.2**

We gain some insight into the motions of the fluid outer core by considering various approximations to the magnetic induction equation in which each term in turn is taken to be negligibly small. The idea here is to forget about the forces and just look at how a magnetic field is affected by motions in the fluid core. Here is the magnetic induction equation again:

$$\frac{\partial}{\partial t} B = \eta \nabla^2 B + \nabla \times (u \times B).$$  \hspace{1cm} (116)

The left side is the local rate of change of the geomagnetic field. Setting it to zero corresponds to a steady state condition in which the field is unchanging; this is of no interest in describing secular variation. If we neglect the second term on the right (which contains effects of advection) then (116) becomes a vector diffusion equation, corresponding to no fluid motion in the outer core. Alternatively, if the core were a perfect conductor $\sigma \to \infty$, we would have $\eta = 0$, and the $\nabla^2$ term could be neglected. What determines the relative importance of advection and diffusion? When $L$ and $U$ are typical length and velocity scales then the ratio of the two terms right is roughly

$$\frac{|\nabla \times (u \times B)|}{|\eta \nabla^2 B|} = \frac{UB/L}{\eta B/L^2} \approx \frac{UL}{\eta} = \mu_0 \sigma UL = R_m.$$  \hspace{1cm} (117)
The dimensionless number $R_m$ is called the magnetic Reynold’s number by analogy with the Reynold’s number, $Re$, used in fluid mechanics. Large magnetic Reynold number corresponds to the dominance of advection over diffusion, while small values indicate a diffusive situation with negligible advection. We will study each of these next. Note that if we use the core diameter as $L$, and $U \approx 0.4 \text{ mm/s}$, we find a rather large value: $R_m \approx 4 \times 10^{-4} \times 7 \times 10^6 / 0.7 = 5,000$.

4.3 Diffusion of the Magnetic Field

Suppose that $R_m$ is small. Then we neglect the advection term in (116) and we study

$$\frac{\partial \mathbf{B}}{\partial t} = \eta \nabla^2 \mathbf{B}. \quad (118)$$

You will recall the diffusion equation, as the one that governs the behavior of heat for example, or salt dissolving in water. Equation (118) is a vector equation and the various components will not in general diffuse separately because they get mixed together in the boundary conditions. It is not hard to solve (118) exactly for a uniform conducting sphere: one uses the decomposition into spheroidal and poloidal parts, and there quickly emerges a diffusion equation for each of the two scalars $T$ and $P$; the only trick is in fitting proper boundary conditions: see Foundations if you are interested.

We can get an idea of what is going on by assuming that the conductor is infinite (thus getting rid of the nuisance of boundary conditions), and looking at the evolution of a magnetic field that initially is in the form of an infinite sine wave with wavelength $\lambda$: we simply assume a behavior of the form $\mathbf{B}(x, t) = B_0 e^{2\pi i x / \lambda} f(t)$. Substitute into (118):

$$\frac{\partial f}{\partial t} = -\frac{4\pi^2 \eta}{\lambda^2} f. \quad (119)$$

The solution to this differential equation is easily seen to be

$$f(t) = f(0) \exp\left(\frac{-t}{t_0}\right)$$

where $t_0 = \lambda^2 / 4\pi^2 \eta$, the characteristic time of exponential decay. Thus

$$\mathbf{B}(x, t) = B_0 e^{2\pi i x / \lambda} e^{-t/ t_0}. \quad (120)$$

Thus we expect all magnetic fields to decay away in time, at a rate inversely proportional to their length scales. The energy stored in the field is dying away as the electric currents transform the field energy into heat through Joule heating $\mathbf{J} \cdot \mathbf{J} / \sigma$.

A more precise argument is the following. Suppose the core is a sphere radius $c$, in an insulator. Dot equation (118) with $\mathbf{B}$ and integrate over all space:

$$\int_{|R^3|} d^3 \mathbf{s} \cdot \mathbf{B} \cdot \partial_t \mathbf{B} = \int_{|R^3|} d^3 \mathbf{s} \eta \mathbf{B} \cdot \nabla^2 \mathbf{B} \quad (121)$$

On the left you will easily verify that $\mathbf{B} \cdot \partial_t \mathbf{B} = \frac{1}{2} \partial_t (\mathbf{B} \cdot \mathbf{B})$ and so the left side of (121) is proportional to the rate of increase of the total field energy. On the right introduce the summation convention and then apply vector identity 4:

$$\int_{|R^3|} d^3 \mathbf{s} \left[\frac{1}{2} \partial_t |\mathbf{B}|^2 = \int_{|R^3|} d^3 \eta \mathbf{B}_j \nabla^2 B_j = \int_{|R^3|} d^3 \eta (\nabla \cdot (B_j \nabla B_j) - \nabla B_j \cdot \nabla B_j) \right]$$ \quad (122)
When Gauss’s theorem is applied to the first term on the right and we integrate over a very large sphere, it is clear this term vanishes, leaving us with
\[ \int |R| d3s \frac{1}{2} \partial_t |B|^2 = - \int |R| d3s \eta \sum_j |\nabla B_j|^2 \] (123)

Since the right side is always negative, this equation shows the total magnetic energy must decrease for all times. Because the field vanishes at infinity, \( B_j \) cannot be constant, and thus the field energy declines inexorably to zero. In fact it can be shown (Foundations, p 271) that
\[ \int |R| d3s |\nabla B_j|^2 \geq \frac{(\pi/c)^2}{c^2} \int |R|^3 d3s B_j^2 \] (124)

from which we can conclude the magnetic field energy dies away at least as fast as \( \exp(-\pi^2 \eta/c^2) t \).

Recall in the earth’s core we have \( \sigma \approx 1.1 \times 10^6 \) S/m or \( \eta \approx 0.7m^2/s \); the core radius is \( c = 3.5 \times 10^6 \) m; then the longest characteristic decay time, \( c^2/\pi^2 \eta = 7.7 \times 10^{11} \) s or 55,000 years. In fact, the same answer is obtained from (120) if we fit a half wavelength across the diameter of the core, and if we determine the longest-lived field in a spherical conductor (see Foundations, Chap 5). This calculation shows that any magnetic field for which \( u = 0 \) cannot persist very long without motions in the fluid, yet we know there has been a field for over 3 billion years from paleomagnetism. Thus we need velocities \( u \) to prevent the death of the main field through Joule heating.

4:4 Frozen Flux and Alfvén’s Theorem

In this section we investigate the much more interesting, and geophysically more appropriate, case in which the magnetic Reynolds number is so large that we can neglect the diffusion term in the magnetic induction equation. Then the fluid effectively becomes infinitely conducting and is called a perfectly conducting fluid. Superconductors conduct perfectly too, but they don’t allow magnetic fields inside them, so a superconductor is different from a perfect conductor.

We look at the way the magnetic field changes according to:
\[ \partial_t B = \nabla \times (u \times B) \] (125)

which is (116) with \( \eta = 0 \). If we use vector identity number 8 to write
\[ \nabla \times (u \times B) = B \cdot \nabla u - u \cdot \nabla B + u \nabla \cdot B - B \nabla \cdot u \] (126)

then since \( \nabla \cdot B = 0 \), and dropping the \( \eta \nabla^2 B \), we transform (125) into
\[ \partial_t B + (u \cdot \nabla) B + B(\nabla \cdot u) - (B \cdot \nabla) u = 0 \] (127)

or, equivalently
\[ \frac{DB}{Dt} = (B \cdot \nabla) u - B(\nabla \cdot u) \] (128)
where $\frac{DB}{Dt}$ is called the material derivative of $B$ and is the rate of change of the field as experienced by a particle moving with the fluid: in general the material derivative is given by

$$\frac{D}{Dt} = \partial_t + \mathbf{u} \cdot \nabla$$

(129)

where the first term on the right accounts for the time derivative at a fixed point, and second gives the apparent change as the particle moves along the spatial gradient.

Although it is unnecessary, let us simplify our discussion to incompressible fluids, which means, as you can easily verify from conservation of matter, that $\nabla \cdot \mathbf{u} = 0$. Now (128) reduces to

$$\frac{DB}{Dt} = (\mathbf{B} \cdot \nabla)\mathbf{u}.$$  
(130)

We wish to show that the same equation is satisfied by a line element joining two neighboring points in the fluid. At time $t$ we mark a particle at the point $x$ and a nearby particle is at point $y$. Time flows on by a small amount $dt$; now the first particle has moved to $x + dt\mathbf{u}(x)$ plus terms in $dt^2$, and the second has traveled to $y + dt\mathbf{u}(y)$. Since $y$ is near to $x$ we can write $\mathbf{u}(y) = \mathbf{u}(x) + (y - x) \cdot \nabla \mathbf{u}$ plus small terms in the square of $|y - x|$. Thus the new vector between the two particles is

$$y(t + dt) - x(t + dt) = (y + dt[\mathbf{u}(x) + (y - x) \cdot \nabla \mathbf{u}]) - (x + dt\mathbf{u}(x)) + O(dt^2)$$

$$= (y - x) + dt[(y - x) \cdot \nabla \mathbf{u}] + O(dt^2).$$

(131)

Hence the material derivative describing a short line element embedded in the fluid is

$$\frac{D(y - x)}{Dt} = (y - x) \cdot \nabla \mathbf{u}.$$  
(132)

As promised this is the same differential equation as (130). Suppose the two particles lie on the same magnetic line of force at time $t$. The two equations (130) and (132) show that at a later time they will remain on the same line of force because the vector $\mathbf{B}$ and the vector $y - x$ evolve identically and must remain parallel. This famous result is known as Alfvén’s Theorem after Hannes Alfvén a Swedish plasma physicist who in his later years was a faculty member at UCSD. The theorem tells us that in perfectly conducting material the magnetic lines of force are permanently attached to the material and move with it.

Equations (130) and (132) also show that the amplitude of the field is proportional to the density multiplied by the length of the fluid element, so that elongation of the material line will increase the field strength.

Next consider a simple surface $P$ with boundary points attached to the material of the fluid and enclosing a number of lines of force; the number of lines of force is proportional to the magnetic flux $\Phi$ threading $P$:

$$\Phi = \int_P \mathbf{B} \cdot \hat{n} d^2 s.$$  
(133)
At a later time $P$ has a different shape since the particles have moved, but none of the field lines that started inside $P$ has crossed the perimeter, because none of the interior particles has either – we assume the fluid is moving continuously without tearing. Thus the number field lines crossing $P$ is unchanged and therefore $\Phi$ is unchanged also. We call this the \textit{frozen flux} condition for perfect conductors. You may not have found this argument very convincing (I know that I don’t).

Here is a more conventional proof which I have found in several places though usually with mistakes. You may skip it if the line-of-force argument satisfies you. \textit{Foundations} has a very different and more difficult proof. We consider a simple patch $P$ in the fluid bounded by the line $\partial P$ which is defined by a set of points attached to material particles. We wish to show that the material derivative of magnetic flux through the patch vanishes:

$$\frac{D\Phi}{Dt} = \frac{D}{Dt} \int_P \mathbf{B} \cdot \hat{n} \, d^2s = 0. \quad (134)$$

In time $dt$ the change of the flux through a surface that moves with the fluid is the contribution of two terms: one from the time derivative with the patch fixed in space, the second from the constant magnetic field when the patch is moved from its old position to the new one, the same idea as in (129):

$$\frac{dt}{dt} \int_P \frac{D\Phi}{Dt} = dt \int_P \partial_t \mathbf{B} \cdot \hat{n} \, d^2s + \left[ \int_{P(t+dt)} \mathbf{B} \cdot \hat{n} \, d^2s - \int_{P(t)} \mathbf{B} \cdot \hat{n} \, d^2s \right]$$

$$= dt \int_P \partial_t \mathbf{B} \cdot \hat{n} \, d^2s + dF. \quad (135)$$

Consider the volume of space $V$ swept out by the particles as they move from $P(t)$ to $P(t + dt)$; the flux of $\mathbf{B}$ out of this region is zero by the divergence theorem:

$$\int_{\partial V} \mathbf{B} \cdot \hat{n} \, d^2s = \int_V \nabla \cdot \mathbf{B} \, d^3s = 0 \quad (136)$$
where \( \hat{s} \) denotes the outward normal to the surface \( \partial V \). Noting that the normal \( \hat{n} \) on \( \partial P(t) \) points inward we see that the flux integral in (136) is also given by

\[
\int_{\partial V} B \cdot \hat{s} d^2s = \int_{P(t+dt)} B \cdot \hat{n} d^2s - \int_{P(t)} B \cdot \hat{n} d^2s + \int_{dS} B \cdot \hat{s} d^2s
\]

\[= dF + \int_{dS} B \cdot \hat{s} d^2s \tag{137}\]

where \( dS \) is the surface strip joining \( P(t) \) and \( P(t + dt) \). Since (136) shows that the left side of (137) vanishes, we have:

\[
\int_{dS} B \cdot \hat{s} d^2s = -dF. \tag{138}\]

Now we calculate \( dF \) in a different way. From Figure 4.4.2 \((\hat{\tau} ds) \times (udt) = \hat{\tau} \times u dt ds\) is the elementary area swept out by \( \partial P \) as it moves from \( \partial P(t) \) to \( \partial P(t + dt) \). The cross product points in the direction of \( \hat{s} \) normal to this surface and so the flux integral can also be written

\[
\int_{dS} B \cdot \hat{s} d^2s = \int_{\partial P} B \cdot (\hat{\tau} \times u dt ds)
\]

\[= dt \int_{\partial P} (u \times B) \cdot \hat{\tau} ds = dt \int_{\partial P} \nabla \times (u \times B) \cdot \hat{n} d^2s \tag{139}\]

where we have applied Stokes’ integral theorem to turn the line integral to a surface integral. We can substitute this expression into \( dF \) in (138) and then into (135); then divide by \( dt \):

\[
\frac{D\Phi}{Dt} = \int_{\partial P} \partial_t B \cdot \hat{n} d^2s - \int_{\partial P} \nabla \times (u \times B) \cdot \hat{n} d^2s
\]

\[= \int_{\partial P} [\partial_t B - \nabla \times (u \times B)] \cdot \hat{n} d^2s. \tag{140}\]

Finally we simply appeal to (125) which shows that the right side must vanish, and therefore \( D\Phi/Dt = 0 \), which is the frozen flux condition.

Roberts and Scott (J. Geomag. Geoelec. 17, p 137, 1965) noted that the time scale for diffusion, \( t_0 \), can be expected to be much longer than that for advection, given the high electrical conductivity in the core: in effect they are saying that
for fields with the scales we can detect from the Earth’s surface, \( R_m \) is large – the same calculation we did at the end of section 4.2, but now with \( L \approx 1,000 \) km (then \( R_m = 180 \)). They called this idea the \textit{frozen-flux hypothesis}, and declared that under this approximation we are allowed to use (118) in studying the secular variation. If fluid particles and field lines are so closely linked, it seems likely that we can learn something rather directly about \( \mathbf{u} \) in the core by observing \( \mathbf{B} \) and its evolution in time. This is indeed the case, as we shall see in the next section.

\textit{Exercise:}

The proof in the notes of the frozen-flux condition is too long. Here is a shorter one. Consider three closely neighboring, non-colinear particles in the surface \( P \) at \( x, y \) and \( z \) at time \( t \). An element of area is given by \( \mathbf{s} = (y - x) \times (z - x) \), and an elementary contribution to the flux integral over \( P \) is given by \( \mathbf{s} \cdot \mathbf{B} \) where \( \mathbf{B} \) is measured at \( x \). Consider a fourth point also close to \( x \) at \( \mathbf{p} = x + \kappa \mathbf{B} \). The product \( dV = (\mathbf{p} - x) \cdot ((y - x) \times (z - x)) \) is the volume of a small parallelepiped of fluid. As time goes on it deforms with the motion. If the fluid is incompressible, \( dV \) cannot change; hence \( \kappa dV \), which is an element of flux, is invariant in time. Then the sum of all such contributions, which gives the total flux, is also invariant in time.

Show from \( \nabla \cdot \mathbf{u} = 0 \) that the elementary volume doesn’t change as the fluid unit moves. Does the long proof in the notes require the fluid to be incompressible as this one does?

4:5 Applications of the Frozen Flux Hypothesis

The beauty of the frozen flux hypothesis is that it allows one to visualize a very direct connection between the core fluid motions and the magnetic field. Indeed it is the case that we can calculate the magnetic field at all future times if we know it at one instant, and we are given the velocity at all times. We won’t do exactly this calculation (again, see \textit{Foundations}), but we will look at a simplified version for the surface of the core. Is the frozen-flux hypothesis really valid for the core? It is so useful for other calculations and deductions we would like to check it from observations. We will discuss how that is done.

4:5.1 Observational Consequences of the FFH

Let us continue to assume the FFH (frozen-flux hypothesis) applies to the core. We are going to look at the magnetic fields right at the surface of the core, a place where we hope downward continuation from the surface makes them observable. We start with the evolution equation (125), and apply standard vector identity 8:

\[
\partial_t \mathbf{B} = \nabla \times (\mathbf{u} \times \mathbf{B})
= \mathbf{B} \cdot \nabla \mathbf{u} - \mathbf{u} \cdot \nabla \mathbf{B} + \mathbf{u} \nabla \cdot \mathbf{B} - \mathbf{B} \nabla \cdot \mathbf{u}. \tag{141}
\]

Now dot this equation with \( \mathbf{\hat{r}} \) which is a fixed vector and so it can travel through the time derivative:

\[
\partial_t B_r = \mathbf{B} \cdot \nabla u_r - \mathbf{u} \cdot \nabla B_r + u_r \nabla \cdot \mathbf{B} - B_r \nabla \cdot \mathbf{u}
\]
\[ = \nabla \cdot (Bu_r) - \nabla \cdot (uB_r). \]  \hspace{1cm} (142)

Remember \( B \cdot \nabla u \) is just the dot product of \( B \) with the gradients of each of the components of \( u \); so when we dot with \( \hat{r} \) this simply picks out the \( r \) component under the gradient. And the second line is just two applications of that most useful identity number 4. At the CMB the core fluid cannot flow radially because it is confined by the solid mantle; hence \( u_r = 0 \).

Similarly radial components of \( u \) are irrelevant in the first term on the right and thus (142) at the CMB reduces to:

\[ \partial_t B_r = -\nabla_s \cdot (u_s B_r). \]  \hspace{1cm} (143)

This equation shows us that if we know \( B_r \) at one time, and we are given the surface velocities on the top of the core, we can calculate \( B_r \) at all future times. Of course given \( B_r \) we can also find \( B \) outside the core uniquely too. This is a remarkably simple solution to what would appear to be a complicated system.

In reality we have the other side of the coin: we know \( B_r \) moderately accurately for 100 years – can we deduce \( u_r \) from (143)? The answer is no. We can see this crudely by observing that \( u_r \) has two unknown components at every point, while (143) supplies only a single constraint, not enough to pin down the velocity without further information. As we will see in a moment there are places where we can gain partial information about \( u_r \).

We now derive an observable consequence of (142). At the surface of the earth there is a line running very roughly around the equator where \( B_r = 0 \) called, appropriately enough, the magnetic equator. When we downward continue the field we find there are at least seven such lines (see Figure 4.5.2.1) that George Backus named null-flux curves. These mini-equators are the site of some interesting properties.

First we show the material particles on a null flux curve move with the curve. We name \( \partial P \) to be a particular (closed) curve on the core on which \( B_r = 0 \); it is the boundary of a patch \( P \) on the core. Consider a particle on \( \partial P \) and a neighboring particle at the position \( \epsilon B \); since \( B \cdot \hat{r} = 0 \) the neighboring particle is in \( \partial P \) as well. As time moves on the two particles can only move tangentially because they are at the CMB, confined to the boundary, thus the line connecting them remains tangential. Since by Alfvén’s theorem that line remains a magnetic field line, magnetic field remains tangential too, and therefore the radial component \( B_r \) continues to be zero for all time.

This already tells us something interesting. By following \( \partial P \) in time we can see the fluid velocity normal to \( \partial P \); the particles on \( \partial P \) cannot escape from it, but they can move along \( \partial P \) so we have no way of finding the component tangential to \( \partial P \); see Figure 4.5.3.1.

But there is more. Since \( \partial P \) is a line of material particles, the frozen-flux condition applies to the flux coming through the area \( P \):

\[ \int_P B_r d^2s = \text{constant}. \]  \hspace{1cm} (144)

This is an observable consequence of the FFH. We can model the magnetic field on the core over a period of time, and calculate the fluxes through the various null flux curves as a function of time. If the FFH is correct the values within each patch should not change, even though the shapes and field values will evolve through secular variation.
4:5.2 Testing the FFH

Does the set of magnetic observations pass this test? The answer turns out to be disappointing and various authors (e.g. Bloxham and Gubbins, *Nature* 317, pp 777-81, 1985) claimed in fact that the geomagnetic field at the core’s surface did not satisfy the FFH because of disagreement between fluxes calculated at various times. But their conclusions are suspect because it is quite difficult to come up with a truly believable estimate for the uncertainty in the geomagnetic models. Whether or not the test succeeds or fails depends vitally on how accurate you think the calculated fluxes really are.

The problem is that the magnetic field at the surface of the core requires infinitely many numbers to describe it completely, but we have only finitely many observations. Therefore while it is possible to say that we know the first \( L \) SH coefficients to a certain accuracy, it is unreasonable to expect us to know every \( b^{(m)}_l \) to some precision. Another equivalent way of looking at the question, which derives from inverse theory, is to say that we know the field at the core up to a certain resolution, meaning that features below a certain length scale are unresolved, and therefore inaccessible to us in the present state of knowledge.

Constable, Parker, and Stark (*Geophys. J. Royal Astron. Soc.* 113, pp 419-33, 1993) argued that the issue could be studied another way: start with magnetic data at two epochs and see if it is possible to actually construct a core field that has satisfies the integral constraints from the FFH. At first this would seem to settle the accuracy problem because now we are asking questions about the accuracy of the data, something we have an idea about. If we can find models obeying the constraint, then the constraint is a physically plausible one; if we cannot, then we must reject the FFH.

But, as these authors showed the question is more subtle than that: if one is allowed to manipulate features on an arbitrarily fine scale, *it is always possible to find pairs of models satisfying the FFH*. In other words, you cannot actually decide the issue with models of a given resolution, or from a finite data set! The best you can say is that the models that fit the observations are implausible in some way, for example, in having to much short-wavelength energy.

In that paper and a later one (O’Brien, Constable, and Parker, *Geophys. J. Int.* 128, pp 434-50, 1997) reasonable-looking geomagnetic models were constructed satisfying the FFH for the epochs 1915, 1945, 1980, and the idea that the FFH is generally agreed to have passed the test. For the past decade continuous satellite observations have improved recent field models (see for example Chulliat & Olsen, *J. Geophys. Res.*, 115, 2010, doi:10.1029/2009JB006994) and the fit to the observation is beginning to look less plausible. It is also true that when we look back on historical timescales to 1590AD we see violation of the FFH as features like the South Atlantic Anomaly appear, growing a reverse flux patch at the CMB.

4:5.3 Calculation of Fluid Flow Models at the Surface of the Core

If we agree to accept the FFH we can hope to say something about the velocity field \( \mathbf{u} \) on \( S(c) \). As discussed earlier it turns out not to be possible to get the velocity from \( B_r \) and its time derivatives alone, as we crudely indicated by a counting argument. A more sophisticated treatment is given in *Foundations*. But as we already indicated, we can find
unambiguously the component of $\mathbf{u}$, normal to the null-flux curves. This is easily done – just calculate the radial field at two close times $t_1$ and $t_2$, then draw the null-flux curves at each time. The normal distance moved between the two is just $\mathbf{u} \cdot \mathbf{\nu}(t_2 - t_1)$. Figure 4.5.3.1 shows one such calculation based on the IGRF-1980 model and its time derivative.

The map is drawn in Mercator projection, so it doesn’t preserve area or scales, but locally angles between features are correct. The sizes of the normal velocity vectors are drawn to scale, except for the imploding patch under the western Pacific, where the lines have been reduced by a factor of four. Recall the velocity component of $\mathbf{u}$, tangent to the null-flux curves is unknown. The map of the world is there for orientation purposes only.

![Map](image)

*Figure 4.5.3.1*

There are several other additional assumptions one can make about the field, which make it possible to find surface fluid flows. One is the *geostrophy*. In the atmosphere a geostrophic wind would arise from exact balance between Coriolis force and the pressure gradient force. In the core this means that the pressure balance is between Coriolis force and the Lorentz force $\mathbf{J} \times \mathbf{B}$. As Backus and Le Mouël showed (*Geophys. J. Internat.* 85, pp 617-29, 1986) however, this constraint does
not make the flow unique – there are large patches where we have no information still!

This is a currently fast-developing field. For a thorough, if now somewhat dated review see Bloxham and Jackson, Fluid flow near the surface of Earth’s outer core, *Rev. Geophys.* 29, pp 97-120, 1991. A more up to date discussion can be found in the Space Science Reviews article by Hulot et al (2010, DOI 10.1007/s11214-010-9644-0).

4:5.4 Determination of the Magnetic Core Radius of the Earth and Planets

The pole strength of the earth is given by the magnetic flux coming out of the core, without regard to the sign:

\[
N = \int_{S(r)} |\mathbf{B} \cdot \hat{s}| \, d^2s = \int |B_r| \, d^2s.
\] (145)

We can obviously calculate this number for any radius spherical surface:

\[
N(r) = \int_{S(r)} |B_r(r, \theta, \phi)| \, d^2s.
\] (146)

Recall that the flux coming through each null-flux curve is invariant in time under the FFH. This means that (145) is a constant of the magnetic field. But there is no reason to suppose the integral (146) at any other radius is constant in time, so this singles out the core radius as a special one. Suppose we compute this number from field models as a function of time and radius. The radius that gives the most nearly constant value for \(N(r)\) is an estimate of the magnetic core radius. The result claimed is \(c_m = 3,484 \pm 48\) km, which is in surprisingly good agreement with the value derived from seismic observations \(c = 3,485\) km. We have already seen another way to get the core’s magnetic radius, by the slope of the spectrum, assuming the spectrum is white at the core.

The method was invented by Raymond Hide (*Nature* 271, pp 640-1, 1978) who also proposed it to determine the structure of other magnetic planets, such as Jupiter, Saturn and Mercury, if we could obtain sufficiently good measurements of their magnetic fields and secular variations.

I have repeated the calculations with the IGRF models and the results are on the next page. I have plotted values of \(N\) through time, normalized by the 1960 values, and at the various radii shown; one needs to normalize because the variations in \(N(r)\) are large as \(r\) varies. You will see I don’t get such great agreement as other workers claim, I find approximate constancy for \(r/a = 0.58\), or \(r = 3,695\) km.

4:6 A Very Little Dynamo Theory

We have seen in 4.3 that if the conductor does not move the magnetic field must die away; in 4.4 we saw at the other extreme, when advection dominates, the field is trapped in the conductor, but there is no increase in magnetic energy. On this evidence it seems unlikely that with an intermediate conductivity fluid motions could maintain or even amplify the magnetic field. Indeed, in 1934 Thomas Cowling published a proof that many believed demonstrated dynamo action to be impossible; however, the proof showed only that purely axially symmetric fields could not be maintained. Intuition
Figure 4.5.4.1
is wrong: we will discuss results proving that the proper kinds of \( \mathbf{u} \) can amplify \( \mathbf{B} \) for intermediate magnetic Reynolds numbers.

Section 4.5 was devoted to the approximation that \( \eta = 0 \) in (114), the induction equation. While this is likely to be a good approximation over centuries, we saw it surely fails over timescales of \( 10^4 \) years or more. Diffusion and advection each play a part in the real dynamo. But what we have learned from Alfvén’s theorem is a handy way of thinking about how the conducting fluid interacts with the magnetic field. When the magnetic Reynolds number \( UL/\eta \) is very high, the field is "dragged" around by the fluid motion, or more precisely, by the component of motion normal to the field lines. When diffusion is important, the fluid is not 100 percent effective in dragging the field lines, and they tend to slip. It can be shown from energy considerations that the field lines act dynamically like stretched strings in these circumstances and exert a force resisting the dragging motion. These qualitative ideas allow us to do an arm-waving kind of analysis of some aspects of kinematic dynamo theory. In particular we will briefly discuss what is commonly believed to be a major mechanism for sustaining the magnetic field: the \( \alpha - \omega \) effect.

Recall that the effectiveness of diffusion is increased for small scale fields (the decay time is like \( L^2 \)). The difficulty in getting a dynamo to work (which means not losing all its energy through diffusion to Joule heating) is to find a way so that short wavelength fields, which are easily created by stirring the conducting fluid (and dragging the field lines into small eddies) can be combined together to create long-wavelength fields.

First the \( \alpha \) effect. Recall the form of Ohm’s law we needed in a moving medium (109) which we will write as \( \mathbf{J} = \sigma \mathbf{E}_1 \), where

\[
\mathbf{E}_1 = \mathbf{E} + \mathbf{u} \times \mathbf{B}.
\]

The effective electromotive force (emf) for driving currents has the unfortunate tendency from this equation to run perpendicularly to already existing magnetic fields, thus making it hard for advection to reinforce fields already present. Steenbeck, Krause and Rädler (see Roberts, P. H., and M. Stix, The turbulent dynamo: a translation of a series of papers by F. Krause, et al., Tech. Note 60, NCAR, Boulder, CO, 1971) in an astrophysical study of magnetic fields in the galaxy, postulated that there might be small-scale \( \mathbf{u} \) superimposed by turbulence on the large-scale flows; writing \( \mathbf{u} = \mathbf{u}_0 + \mathbf{u}' \). Let us write averaging over a large length scale as \( \langle \cdot \rangle = \frac{1}{L^3} \int \cdot \, dV \) so that \( \langle \mathbf{u}' \rangle = 0 \). Next we introduce a quantity called helicity:

\[
h = \mathbf{h} = \langle \mathbf{u} \cdot \nabla \times \mathbf{u} \rangle.
\]

Helicity gives the average amount of "screw-like" motion, because we are dotting the velocity \( \mathbf{u} \) at a point with its vorticity, which is just (twice) the local angular vector velocity. It is known that in a rapidly rotating convecting system, the turbulent motion has a net average helicity, as given by (148). Steenbeck and Krause showed by averaging over a velocity field with a nonzero average helicity (147) becomes, through interaction of the small-scale terms:

\[
\langle \mathbf{E}_1 \rangle = \langle \mathbf{E} \rangle + \alpha \langle \mathbf{B} \rangle
\]

where \( \alpha \) is a constant (or a tensor) depending on the details of the motion. But the important fact is that the helical motions on a fine scale produce emfs that are parallel to the large scale \( \mathbf{B} \); this is the (unimaginatively named) \( \alpha \) effect.
With the $\alpha$ effect it is trivial to maintain a dynamo. Even if the average fluid flow is zero, we can get growing fields. This is shown in *Foundations*, Chapter 6.

Armed with the $\alpha$ effect many dynamo theorists look only at large-scale $u$ and $B$ and assume that the small-scale motions are harmless and just produce the useful $B$-parallel emf, through (149). This is given the grand name of *mean-field electrodynamics*. Here is a qualitative description of a popular dynamo, the $\alpha$-$\omega$ model. We begin by assuming there is a large-scale poloidal field in a rotating spherical conductor, aligned with the spin axis. The fluid near the equator is assumed to rotate a bit faster than that in higher latitudes (observed in rotating fluids, like the atmosphere) and over time this drags the poloidal field into a wound up spiral through the attachment of the fluid to the field. But diffusion causes the spiral to condense into a pair of opposite-signed toroidal fields, one in each hemisphere. This process is called the $\omega$ effect, and though we have described it as a discrete process, it will go on continuously, transforming the large poloidal $B_P$ into a quadrupole $B_T$. Now the $\alpha$ effect operates, but you will perhaps believe it would possesses opposite signs in the two hemispheres, because the helicity derives from the overall rotation. So the toroidal $B_T$ generates an effective electric field, with the same direction in the upper and lower hemispheres, in turn causing a toroidal $J$ to be created, which is the source for more poloidal $B_P$. The loop is closed. The figure illustrates the fate of a single field line in this scenario. I don’t know how satisfactory this kind of "explanation" is without the algebra, but the equations do back it up. There are many other mechanisms.

For an advanced but very satisfactory review of dynamo theory read the two articles in *Geomagnetism, Vol 2* by Gubbins and Roberts.
5. THE CRUSTAL MAGNETIC FIELD

Magnetic anomalies are the signals of the internal geomagnetic field left behind after the part generated by Earth’s core has been removed from the observations. Most of the mantle and some of the lower portions of the crust are above the Curie point of magnetic minerals; thus the source of the magnetic anomalies must lie in the crust or uppermost mantle. We see this from the Lowes spectrum; if there were significant mantle sources, the spectrum would not die away exponentially, but more as a power law, just like the gravitational SH spectrum. Magnetic anomalies have of course played a very important role in earth sciences: the correlation of lineated magnetic anomalies roughly parallel to mid-oceanic ridge crests with the geomagnetic reversal time scale provided the final evidence for seafloor spreading.

Regional magnetic surveys, both on land and under the oceans used to be performed by taking spot readings of the magnetic field. Now, however, aircraft or shipboard surveys are commonly carried out with a towed proton precession magnetometer, which measures only the field intensity not its direction. In order to study magnetic anomalies a standard (usually IGRF) core field model, \( B_0 \), is subtracted from the observations; this is often called the regional magnetic field.

Suppose the sum of all contributions to the field is \( B \). The total field anomaly is the difference between the magnitudes of the observed and core magnetic fields:

\[
\Delta T = |B| - |B_0|.
\]  
(150)

Now we will let \( \Delta B \) be the contribution to \( B \) of some anomalous magnetic source. Then

\[
B = B_0 + \Delta B.
\]  
(151)

We would like to know \( \Delta B \) to study the crustal source, so

\[
\Delta T = |B_0 + \Delta B| - |B_0|
\]

\[= \sqrt{|B_0|^2 + |\Delta B|^2 + 2\Delta B \cdot B_0} - |B_0|.\]  
(152)

Neglecting quantities \( O|\Delta B|^2/|B_0|^2 \) we find

\[
\Delta T = |B_0|[1 + 2\Delta B \cdot B_0/|B_0|^2]^{\frac{1}{2}} - |B_0|
\]

\[= |B_0|[1 + \Delta B \cdot B_0/|B_0|^2] - |B_0|
\]

\[= \Delta B \cdot B_0/|B_0| = \Delta B \cdot \hat{B}_0.\]  
(153)

Thus \( \Delta T \approx \Delta B \cdot \hat{B}_0 \), that is, it is the component of the anomaly field in the direction of the regional field, provided the anomaly field is small in magnitude relative to the total field. The validity of this approximation depends on the size of \( \Delta B \) relative to \( B \). Typical crustal magnetic anomalies range in magnitude from a few nT to several thousand nT, but are usually less than 5,000 nT, so this provides an adequate representation for total field anomalies.
Is the total field harmonic? Approximately, to the extent that the regional field direction remains constant in the survey region. We can see this easily, writing $\Delta \mathbf{B} = -\nabla V$ with harmonic $V$,

$$\nabla^2 \Delta T = -\nabla^2 (\nabla V \cdot \mathbf{B}_0) = -\mathbf{B}_0 \cdot \nabla \nabla^2 V = 0. \quad (154)$$

If $\mathbf{B}_0$ is not effectively constant, the algebra is messy and the subject not worth much effort.

As with satellite observations, the reason one prefers to measure the field magnitude $|\mathbf{B}|$ at sea is that there is no need to keep an accurately oriented platform, and total field data can be made to $\pm 1$ nT in 60,000 nT with very robust instruments, and $\pm 0.1$ nT with only a little more effort. As we asked in 3.4.1 we can inquire whether knowledge of $\Delta T$ is actually sufficient to describe the harmonic field of crustal sources. If $\mathbf{B}_0$ is effectively constant, and not horizontal, then knowledge of $\Delta T$ is fully equivalent to knowledge of $\Delta \mathbf{B}$ itself, since one can then construct $\Delta \mathbf{B}$ from $\Delta T$. But as usual, things aren’t quite so simple. One collects survey data on long lines in the oceans, and it really is not possible to know $\mathbf{B}$ everywhere on a surface. If instead of $\Delta T$ on a long line one actually has the vector data $\Delta \mathbf{B}$, then valuable information can be obtained about the accuracy of the measurements and other useful things, like across track lineation, by looking at the correlations between the components. See Parker and O’Brien, *J. Geophys. Res.* 102, pp 24815-24, 1997. Over the past two decades there has been a move towards measuring the vector field in some marine surveys, particularly on Japanese ships and on near-seafloor instruments.

5:1 Magnetic Permeability and Susceptibility

In chapter 2 we outlined the relationship between magnetic displacement $\mathbf{H}$ and magnetic induction $\mathbf{B}$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \quad (155)$$

with $\mathbf{M}$ representing the magnetic polarization or magnetization of the material. In the region where we make measurements it is unnecessary to distinguish between $\mathbf{B}$ and $\mathbf{H}$ because there are no currents flowing and no magnetization. Inside the crust, however, $\mathbf{M}$ depends on the atomic and macroscopic properties of the material. Materials can acquire a component of magnetization in the presence of an external magnetic field (such as that generated in Earth’s core). The so-called *induced magnetization* is often considered to be proportional in magnitude to and along the direction of the external field and

$$\mathbf{M} = \chi \mathbf{H} \quad (156)$$

and $\chi$ is called the *magnetic susceptibility*. Then we can write

$$\mathbf{B} = \mu_0 (1 + \chi) \mathbf{H} = \mu \mathbf{H} \quad (157)$$

where $\mu$ is the *magnetic permeability*. In practice $\chi$ may be dependent on field intensity, negative or need to be represented by a tensor (magnetically anisotropic materials).
Although diamagnetism due to perturbations of electron orbits in an applied field and paramagnetism (perturbations of atomic magnetic moments) are important physical processes, these are insignificant contributors to the geomagnetic field. The important contributions come from materials with atomic moments that interact strongly with each other as a result of quantum mechanical exchange interactions. These are called ferrimagnetic materials and they can carry either an induced or remanent magnetization. The total magnetization of a rock will result from the sum of these two contributions

\[ \mathbf{M} = \mathbf{M}_i + \mathbf{M}_r = \chi \mathbf{H} + \mathbf{M}_r. \]  

(158)

The stability and acquisition of magnetization depends on temperature: above the Curie temperature thermal perturbations destroy the spontaneous magnetization so that the only remaining magnetization is from diamagnetic or paramagnetic effects.

Magnetite (Fe₃O₄, Curie temperature 580°C) and its solid solutions with ulvospinel Fe₂TiO₄ are the most important magnetic minerals in crustal rocks, although hematite, pyrhotite also play a role in paleomagnetic studies.

5:2 Crustal Magnetic Models

The ultimate goal of a regional magnetic survey is to make some inferences about the spatial distribution and nature of the magnetic sources generating the anomalies and hence to draw conclusions about the geological processes active in the earth. The first step in understanding this problem is to calculate the fields from a model: that is for a given distribution of magnetization predict the expected observations; this is called solving the forward problem.

How can we do this? We start by considering the magnetic scalar potential \( \Psi \) at position \( \mathbf{r} \) due to a point dipole \( \mathbf{m} \) located at \( \mathbf{s} \):

\[ \Psi(\mathbf{r}) = \frac{\mu_0}{4\pi} \mathbf{m} \cdot \nabla_s \frac{1}{|\mathbf{r} - \mathbf{s}|}. \]  

(159)

To find the potential for a distribution of magnetization, we consider it to be a sum of contributions from elemental dipoles (recall the magnetization vector \( \mathbf{M} \) is just a density of dipole moment per unit volume):

\[ \Psi(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_V \mathbf{M}(\mathbf{s}) \cdot \nabla_s \frac{1}{|\mathbf{r} - \mathbf{s}|} d^3s \]  

(160)

where \( V \) is the magnetized source region. Then

\[ \mathbf{B} = -\nabla_r \Psi(\mathbf{r}) = -\frac{\mu_0}{4\pi} \nabla_r \int_V \mathbf{M}(\mathbf{s}) \cdot \nabla_s \frac{1}{|\mathbf{r} - \mathbf{s}|} d^3s \]

\[ = -\frac{\mu_0}{4\pi} \int_V \mathbf{M}(\mathbf{r}) \cdot \nabla_s \frac{1}{|\mathbf{r} - \mathbf{s}|} d^3s. \]  

(161)

You will easily verify that

\[ \nabla_r \nabla_s \frac{1}{|\mathbf{r} - \mathbf{s}|} = -\nabla_s \nabla_s \frac{1}{|\mathbf{r} - \mathbf{s}|}. \]  

(162)
Finally, if we have measured the total field anomaly and the regional core field has constant direction $\hat{y}$ in one direction, usually identified with the $x$ axis. Then (165) can be reduced to an integral of the form

$$\Delta T = \frac{\mu_0}{4\pi} \int_V \left[ 3 \mathbf{M} \cdot \mathbf{R} \frac{\mathbf{R}}{|\mathbf{R}|^3} - \frac{\mathbf{M} \cdot \mathbf{B}_0}{|\mathbf{R}|^3} \right] d^3s$$

(165)

Solving the forward problem thus comes down to evaluating the above integral for an appropriately shaped body. In practice this is usually done by making approximations to the shape of the body.

For compact regions such as seamounts (which are isolated submarine volcanoes) a favorite plan is to divide the region $V$ into a set of smaller elementary shapes, like cuboids, over which the integral can be done exactly, then sum. See Blakely’s book for some references to this approach. Sometimes it is permissible to take $\mathbf{M} = \text{constant}$ in space. Then one can apply identities and get (165) into a surface integral with the divergence theorem. The surface can be conveniently approximated by a set of triangular faces, called a tessellation; see Parker, Shure, and Hildebrand, *Rev. of Geophys.* 25, pp 17-40, 1987.

In the early days of marine magnetic survey work it was discovered, as you will know, that in many places the magnetic anomaly pattern takes the form of a series of stripes, caused by reversals of the ancient geomagnetic field and linear emplacement at the ocean ridges. This geometry allows a simplification by assuming that the magnetization $\mathbf{M}$ is constant in one direction, usually identified with the $y$ axis. Then (165) can be reduced to an integral of the $x$-$z$ plane, and is much easier to do; the formula for a polygon shape is quite simple.

Often it can be assumed that the magnetic layer is very thin, so that one can ignore variations of $\mathbf{M}$ in the vertical $z$ direction. You will appreciate that when the thin-layer assumption is made, (165) has the form of a convolution and then
Fourier methods can be invoked. This is very popular in marine magnetic work, both for the older single-profile data sets and the more modern surveys of an area. Again Blakely devotes a lot of space to this issue.

Exercise:

(a) As suggested, with \( M \) constant rewrite (165) in a form that you can apply the divergence theorem and reduce the integral to a surface form.

(b) Find out about Poisson’s relationship. Explain what it is and when its use might be appropriate.

5:3 The Magnetic Annihilator and Runcorn’s Theorem

At the risk of being repetitious, I want to restate that the object of the process of magnetic modeling is to learn about the state of the crust, first its magnetic state, and then, if we’re lucky, other things too. In the last section we assumed that \( M \) was known and we calculated \( \Delta \mathbf{B} \) from it, but in reality we don’t know the magnetization and do know the anomaly. We need to reverse the process, which is called solving the inverse problem. A major question in any inverse problem is that of uniqueness: does a (complete and exact) set of data determine the unknown magnetization, or is there ambiguity, even with ideal data? It turns out the inverse problem for magnetization is ill-posed, which means that there are infinitely many possible solutions to choose from, unless further restrictions or simplifying assumptions are brought in. You might perhaps have come to this conclusion from the equivalent source theorem for gravity, which applies equally well here: if \( \mathbf{B} = -\nabla \Psi \) and all magnetic sources are confined to compact region \( V \)

\[
\Psi(\mathbf{r}) = \frac{1}{4\pi} \int_{\partial V} d^2 s \left[ \frac{\Psi}{|\mathbf{r} - \mathbf{s}|} \frac{\partial}{\partial n} - \frac{1}{|\mathbf{r} - \mathbf{s}|} \frac{\partial \Psi}{\partial n} \right].
\] (166)

This equation tell us that we can mimic the potential of a magnetized body by poles and dipoles on the surface, whatever the true interior distribution of \( M \) may be. The need for poles is a bit disturbing, however.

Here is another, perhaps more startling example. In (165) let the magnetization be \( \mathbf{N} = \nabla q \) where \( q(s) \) is any smooth function that vanishes on the boundary \( \partial V \). It will be seen from (162) that \( \mathbf{G} = \nabla p \) and \( \nabla^2 p = 0 \) provided \( \mathbf{r} \) is outside \( V \), which it always is. Then

\[
\frac{4\pi}{\mu_0} \Delta T = \int_V \nabla p \cdot \nabla q \, d^3 s = \int_V [\nabla \cdot (q \nabla p) - q \nabla^2 p] \, d^3 s
\]

\[
= \int_{\partial V} q \nabla p \cdot \hat{n} d^2 s = 0.
\] (167)

Thus the anomaly caused by this whole family of functions is zero. Hence if we have a magnetization \( M \) that matches observation, then so does \( M + N \). The function \( N \) is an example of a magnetic annihilator, a function with no observable magnetic field, yet nonzero internal magnetization. Of course we have already encountered toroidal magnetic fields, whose currents in a sphere are also annihilator sources.

The existence of annihilators means that we cannot ever know what a magnetic source is based solely on its magnetic anomaly. The problem is ill-posed.
It was assumed in the marine world that if the magnetic layer was very thin, this problem would be eliminated. But Parker and Huestis (J. Geophys. Res., 79, pp 1587-93, 1974) showed there was ambiguity even then, in the form of a single function which can be added in arbitrary amounts to any solution, without disturbing the agreement.

Finally, an important example of an annihilator was discovered by Keith Runcorn in studies of the moon’s magnetic field (Nature 253, 1042, pp 701-3, 1975). Runcorn was convinced the moon had once had an internal dynamo which has ceased to operate because the lunar core became solid. Samples from the moon have proved to be magnetic, yet there appears to be almost no lunar magnetic field. Critics of Runcorn asked, If there is a general lunar magnetization from the ancient dynamo, why can’t we see the field from these fossil sources? Here is Runcorn’s surprising answer: If the magnetization of the moon is induced magnetization in a shell, no matter what form the internal dynamo field was like, the resulting magnetization would be an annihilator – no observable magnetic field even though the rocks could be strongly magnetic.

Here is a proof. Consider the induced lunar magnetization in the shell $c \leq r \leq b$: we suppose $\mathbf{M} = \kappa \mathbf{\nabla} \Phi$ where $\kappa = -\chi/\mu_0$ assumed constant, and so $\nabla^2 \Phi = 0$, since $\Phi$ is the scalar potential of the dynamo source in the moon’s core. Now consider the scalar potential $\Psi$ from the induced sources, outside the shell. Let $R = |\mathbf{r} - \mathbf{s}|$, then by (160):

$$\Psi(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_V \mathbf{M} \cdot \mathbf{\nabla} \frac{1}{R} \, d^3 \mathbf{s} = \frac{\mu_0}{4\pi} \int_V \kappa \mathbf{\nabla} \Phi \cdot \mathbf{\nabla} \frac{1}{R} \, d^3 \mathbf{s}. \quad (169)$$

We can now apply our favorite vector identity and the divergence theorem:

$$\Psi = \frac{\mu_0 \kappa}{4\pi} \int_V d^3 \mathbf{s} \left[ \mathbf{\nabla} \cdot \left( \frac{1}{R} \mathbf{\nabla} \Phi \right) - \frac{1}{R} \mathbf{\nabla}^2 \Phi \right] = \frac{\mu_0 \kappa}{4\pi} \left[ \int_{S(b)} - \int_{S(c)} \right] d^2 \mathbf{s} \frac{1}{R} \partial_r \Phi = \Psi_b - \Psi_c \quad (170)$$

where we have used the fact $\Phi$ is harmonic; notice in (170) that the normals on $S(c)$ and $S(b)$ point in opposite directions and so the two surface integrals must be subtracted. Let us look at the contribution $\Psi_b$ from the outer surface $S(b)$; we introduce the SH expansion of $1/R$ from Part I (equation 138), and an expansion in $c_{lm}^m$ for the potential $\Phi$:

$$\Psi_b = \frac{\mu_0 \kappa}{4\pi} \int_{S(b)} d^2 \mathbf{s} \left[ \sum_{l,m} \frac{4\pi}{2l + 1} \mathbf{l}^l(\hat{\mathbf{r}}) Y^m_l(\hat{\mathbf{s}}) Y^*_{lm}(\hat{\mathbf{s}}) \right] \left[ \sum_{l',m'} \frac{b^l + 1}{b^{l'} + 1} \sum_{\ell'\ell} c_{\ell'}^{m'} Y^{\ell'}_{l'}(\hat{\mathbf{r}}) \right] \int_{S(1)} d^2 \mathbf{s} Y^{m'}_{l'}(\hat{\mathbf{s}}) Y^m_l(\hat{\mathbf{s}}) \right] \frac{1}{R} \, d^2 \mathbf{s}$$

$$= -\frac{\mu_0 \kappa}{b^2} \sum_{l,m} \sum_{l',m'} c_{l'}^{m'} \frac{b^l + 1}{2l + 1} \frac{1}{\ell'^{l'} + 1} Y^m_l(\hat{\mathbf{r}}) \int_{S(1)} d^2 \mathbf{s} Y^{m'}_{l'}(\hat{\mathbf{s}}) Y^m_l(\hat{\mathbf{s}}) \right] \frac{1}{R} \, d^2 \mathbf{s}$$

$$= -\mu_0 \kappa \sum_{l,m} \frac{l + 1}{2l + 1} c_l^m Y^m_l(\hat{\mathbf{r}}). \quad (171)$$

The remarkable thing about (171) is that it is independent of the radius of the surface $b$; so the same answer will be obtained for $\Psi_c$, the integral over $S(c)$. But $\Psi = \Psi_b - \Psi_c$, and so $\Psi(\mathbf{r})$ is identically zero for all $\mathbf{r}$.

This argument is applicable to the Earth to some degree. The induced crustal magnetization from a uniform medium in a shell will produce no observable anomaly. You will easily see this generalizes to any series of shells, so $\chi$ can vary with...
depth, and the annihilator property persists. Thus the crustal fields we see are due to lateral variability, changes in layer thickness, and other departures from uniformity. Of course the Earth’s crust is so heterogeneous we are not surprised to see very large anomalies almost everywhere. Why would one expect the upper regions of the moon to be less heterogeneous than the Earth’s crust?

5:4 Results – Magnetic Anomalies Everywhere

The major triumph for magnetic anomalies has been the discovery of the Vine-Mathews magnetic stripes, the discovery of sea-floor spreading, and the mapping of the ages of the ocean basins. Reversal of direction results in a very strong, short wavelength contrast in magnetization which gives rise to intense magnetic anomalies (200-2,000 nT) in the scale range 10-100 km at the sea surface. Traditionally these anomalies have been modeled by thin layers (500 - 1,000 m thick) with blocks of magnetization where the direction is constant, except for changes in sign, and the intensity is constant too. This kind of model never fits the observations exactly, but does a reasonably good job if geological ages are needed.

If one wishes to get more detail, and allow more flexibility in the models the annihilator remains a problem. The next step in sophistication is to allow the models to vary in intensity with \( x \), and to make the layer follow the topographic variations. The magnetic annihilator then has a large scale component and another that varies like the bathymetry. Adding or subtracting this function can make the magnetization change sign and appear more-or-less correlated with topographic relief. When one is seeking geomagnetic intensity histories, the ambiguity is troublesome.

One of the happy exercises carried out on profile data is to apply spatial linear filters to change the apparent dip of the magnetization vector. It can be shown that there is a filter which after application to an anomaly results in the profile that would have been observed at the north pole, with vertical magnetization. This activity is called reduction to the pole (see Blakely for more). The idea here is that magnetizations measured or acquired in different latitudes yield anomalies that differ widely in appearance even if the underlying block pattern is similar; reduction to the pole makes profiles much easier to compare, and also offers some information about the latitude of formation of that piece of crust.

Another (but less successful) application of geomagnetism to marine anomalies is the analysis of the anomalies from seamounts. The major piece of information required here is the average direction of magnetization, which gives the usual paleomagnetic clues about where the volcano was formed. To overcome a serious ambiguity problem the first studies simply assumed the magnetization vector within the seamount was quite constant, both in direction and in intensity. Now with only three parameters to fit nonuniqueness disappears. Of course the constant-\( M \) models don’t fit the observations very well at all, but marine geologists have generally ignored this difficulty. Various methodologies have been invented for improving the fits and estimating the error in the resulting direction. One approach that worked reasonably well recognized the observation from drilling the oceanic crust that direction of \( M \) is much more nearly constant than \( |M| \), which varies by more than one order of magnitude in a single body. So we ask, What uni-directional magnetizations fit the given anomaly pattern, if any? It turns out that in some cases, quite a close clustering of directions will fit the anomaly, thus providing a direction and uncertainty. For details see Parker, R. L., A theory of ideal bodies for seamount magnetism, *J. Geophys.*
Magnetic anomalies observed near the surface over land are generally smaller in amplitude than the marine ones. This is assumed to be because continental rocks are on the whole comprised of much less magnetic types: granitic, metamorphic and sedimentary rocks are orders of magnitude less magnetizable, and have no thermoremanent component. But aeromagnetic surveys are relatively cheap and are routinely used to locate and delineate ore bodies. At the longest wavelength (> 1,000 km) however, as mapped by satellite, the continents have larger magnetic anomalies than the oceans. It is assumed that this is because on the longest scales induced magnetization is at work, and the extra thickness of the continental crust (or greater depth to Curie isotherm) gives the continents an advantage.
6. ELECTROMAGNETIC INDUCTION IN EARTH’S CRUST AND MANTLE

In the late 19th century Schuster discovered that a minor part of the diurnal variation field originated within the earth: this part of the variation is due to eddy currents induced within the conducting earth by the larger external field. The response of Earth to time variations in the externally generated geomagnetic field can be interpreted in terms of electrical conductivity variations with depth. The estimation of geomagnetic response functions and their interpretation in terms of mantle electrical conductivity structure dates from the end of the last century.

There are two different approaches: magnetotelluric sounding and geomagnetic deep sounding. The first uses relatively high frequency variations (sometimes as high as 1 kHz) and requires the measurement of electric as well as magnetic fields; it is suitable mainly for shallow structures, in the upper crust or above. As the name suggests the magnetic sounding employs only magnetic fields usually measured by arrays of instruments, for probing the mantle.

6:1 Skin Depth

We shall ignore magnetic permeability and electric polarizations, fairly good approximations in the deep crust and mantle. Then the equations we must work with are the same as those we studied in 4.3, derived from pre-Maxwell’s equations in a stationary conductor; we get the vector diffusion equation for $\mathbf{B}$. There are three differences in the approach here. First, we cannot so readily ignore the boundary conditions, which in our discussion of the core had no major effect on the physics. The interface of the conductor with the insulating atmosphere is critical because it is there that we make our measurements. The second, perhaps slightly related, question concerns the temporal behavior of the magnetic field, which we now think of as typically periodic, not exponentially decaying with time. Finally, we cannot treat the conductor as uniform for any but the most superficial analysis because in the mantle alone $\sigma$ increases by at least four orders of magnitude.

Despite the last remark, we will perform a simple study of a uniform conductor. In our present problem energy is being supplied by a fluctuating external field, created above the atmosphere by the solar wind and its interplay with the magnetosphere. So we write (118), the vector diffusion equation

$$\nabla^2 \mathbf{B} = \mu_0 \sigma \partial_t \mathbf{B}. \quad (172)$$

Later we will return to spatially variable $\sigma$, but to get started we will make $\sigma$ constant. We erect a traditional Cartesian coordinate system with $z$ positive downward. Then above ground, $z < 0$, the atmosphere is an insulator; below in $z > 0$, is a uniformly conducting halfspace with conductivity $\sigma$. The wavelengths of the external fields are very long (the ring current is typically at 10 Earth-radii) and so for a local problem we may consider a uniform magnetic field in the atmosphere, in the $x$ direction, varying in time as $e^{i\omega t}$:

$$\mathbf{B} = \hat{x} B_0 e^{i\omega t}, \quad z \leq 0. \quad (173)$$

Below ground, to match continuity at $z = 0$, we have a horizontal field too, but it can vary in the $z$ direction:

$$\mathbf{B} = \hat{x} b(z) e^{i\omega t}, \quad z \geq 0. \quad (174)$$
Plugging (174) into (172) gives

\[ i\omega\mu_0\sigma\dot{b}(z)e^{i\omega t} = \ddot{x}e^{i\omega t}\frac{d^2b}{dz^2} \tag{175} \]

or

\[ \frac{d^2b}{dz^2} = i\omega\mu_0\sigma b(z). \tag{176} \]

This is the simplest second-order ordinary differential equation (it is just \( y'' = k^2y \) dressed up), and its general solution is

\[ b(z) = Ae^{(1+i)z/z_0} + Be^{-(1+i)z/z_0} \tag{177} \]

where

\[ z_0 = \sqrt{2/\omega\mu_0\sigma} = 1/\sqrt{\pi\mu_0 f\sigma}. \tag{178} \]

The first term in (177) is a growing (and oscillating) exponential function, a term that increases with depth: this term is analogous to the magnetic fields of internal origin we saw in the ordinary SH expansion. Since the source of the energy for the system is above ground, we conclude that \( A = 0 \). The remaining term dies away exponentially with depth, and the characteristic e-folding scale is \( z_0 \) which is the skin depth. Matching with the atmospheric field at \( z = 0 \) gives us

\[ B(z) = \hat{x}B_0e^{-z/z_0}[\cos z/z_0 - i\sin z/z_0]. \tag{179} \]
It is easily seen from (178) that the high-frequency variations correspond to very shallow disturbances below ground, while the long period fields penetrate more deeply. You will easily verify that both the electric field and electric currents (both horizontal, in the \( y \) direction) fall off at the same rate as the magnetic field.

### Table of Skin Depths

<table>
<thead>
<tr>
<th>material</th>
<th>( \sigma ), S/m</th>
<th>1 year</th>
<th>1 month</th>
<th>1 day</th>
<th>1 hour</th>
<th>1 sec</th>
<th>1 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>core</td>
<td>( 3 \times 10^5 )</td>
<td>4 km</td>
<td>770 m</td>
<td>200 m</td>
<td>40 m</td>
<td>71 cm</td>
<td>23 mm</td>
</tr>
<tr>
<td>L. mantle</td>
<td>10</td>
<td>900 km</td>
<td>170 km</td>
<td>46 km</td>
<td>9 km</td>
<td>160 m</td>
<td>5 m</td>
</tr>
<tr>
<td>seawater</td>
<td>3.4</td>
<td>1500 km</td>
<td>300 km</td>
<td>80 km</td>
<td>16 km</td>
<td>270 m</td>
<td>9 m</td>
</tr>
<tr>
<td>sediments</td>
<td>0.1</td>
<td>9000 km</td>
<td>1700 km</td>
<td>460 km</td>
<td>95 km</td>
<td>1.6 km</td>
<td>50 m</td>
</tr>
<tr>
<td>U. mantle</td>
<td>0.001</td>
<td>( 10^4 ) km</td>
<td>( 10^4 ) km</td>
<td>( 4600 ) km</td>
<td>( 950 ) km</td>
<td>16 km</td>
<td>500 m</td>
</tr>
<tr>
<td>igneous</td>
<td>( 1 \times 10^{-5} )</td>
<td>10^5 km</td>
<td>10^5 km</td>
<td>10^4 km</td>
<td>9500 km</td>
<td>160 km</td>
<td>5 km</td>
</tr>
</tbody>
</table>

The table shows skin depths for a variety of typical Earth environments and a large range of frequencies. The first line shows that for practical sounding the core is effectively a perfect conductor into which the external magnetic field does not penetrate. The first column shows that 1-year variations (which are not the lowest frequency external fields – there is an 11-year cycle associated with solar sunspot activity) can penetrate into the lower mantle and even to the bottom of the mantle. The very large range of conductivities found in the crust points to the need for a corresponding large range of frequencies in electromagnetic sounding of that region.

### 6:2 Variable \( \sigma \) – Magnetotelluric Sounding

As we already noted, the magnetotelluric (MT) method relies on simultaneous measurements of time series of the electric and magnetic fields at the surface. The traditional development, which we will follow, works with the electric field \( E \) rather than with \( B \). As in the previous section we use a flat-Earth approximation and a uniform exciting field above the ground; the only difference is that now the conductivity is allowed to vary with \( z \). The idealized experimental arrangement to measure two horizontal components, \( B_x \) and \( E_y \) at a single site. We will see in a moment that if the ground is truly vertically stratified the electric and magnetic fields will be mutually perpendicular. Then think of the magnetic field as causing eddy currents in the ground, currents that are evident from the electric fields seen at the surface. The governing equations are linear, and so a periodic magnetic field with frequency \( f \) excites currents and a responding electric field at the same frequency. In the real world the fluctuations of magnetosphere are far from sinusoidal, but that difficulty is overcome by Fourier transforming the actual magnetic and electric time series, thereby isolating a particular frequency by extracting a particular Fourier component. Fourier decomposition is done at a series of frequencies motivated by the skin-depth analysis: the long-period variations penetrate deep into the ground and give some kind of average measure of the deep conductivity, while the short-period energy probes the near-surface structure.

Since the amplitude of the source field is of no interest to us in this problem, our objective must be to find a measure
that removes dependency on properties of the source. In magnetotelluric sounding one divides the electric field complex Fourier amplitude at a given frequency, by the complex amplitude of the magnetic field. The resulting complex number carries the electromagnetic response of the ground; it is complex because the currents induced into the ground lag behind the forcing magnetic field, and the phase lag is contained in the complex response, as well as the amplitude. Let us analyze the simplest physical system.

We assume all the fields vary periodically in time as $e^{i\omega t}$; then we write two of the pre-Maxwell equations and Ohm’s law

$$\nabla \times \mathbf{E} = -i\omega \mathbf{B} \quad (180)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \quad (181)$$

$$\mathbf{J} = \sigma \mathbf{E} \quad (182)$$

We take the curl of (180), then substitute from (181) and (182):

$$\nabla \times \nabla \times \mathbf{E} = -i\omega \mu_0 \sigma \mathbf{E} \quad (183)$$

In the geometry of the previous section nothing varies in the $x$ or $y$ directions, and that remains true here also. We look for solutions in this form, with variations in the $z$ direction only. Again we set the $\mathbf{B}$ field in the $\hat{x}$ direction. Then it follows from (181) that the vectors $\mathbf{J}$ and hence $\mathbf{E}$ (through Ohm’s law) have no $x$ or $z$ components; they are vectors in the $y$ direction. Thus $\nabla \cdot \mathbf{E} = \partial_y E_y$; but there is no gradient in $y$ and so we must have that $\nabla \cdot \mathbf{E} = 0$. Hence the familiar vector identity shows that $\nabla \times \nabla \times \mathbf{E} = -\nabla^2 \mathbf{E}$; from this, bearing in mind there are no $x$ or $y$ gradients, we conclude that (183) simplifies to

$$\partial^2_z E_y = i\omega \mu_0 \sigma E_y \quad (185)$$

which is an ordinary differential equation governing $E_y$. We will concentrate on solving this equation. To find $B_x$ we use (180)

$$B_x = \frac{1}{i\omega} \partial_z E_y \quad (186)$$

In the previous section we obtained one boundary condition from the fact that of two linearly independent solutions, one grows with depth. That works here too, but particularly for numerical solutions, a boundary condition applied at $z = \infty$ is awkward. It is easier to work in a system with finite $z$ extent; then a natural boundary condition is to set $E_y = 0$ at the base, which we will put at $z = h$. Physically this is equivalent to placing a perfect conductor there; a small amount of energy does flow upward now – it is the energy reflected back from the perfect conductor. Obviously if $h$ is many skin depths (at the lowest frequency), solutions to the finite problem will be indistinguishable from those in a halfspace.

To get the second boundary condition recall that the measurements we will be using are normalized by the source field. In fact a common type of measurement is the magnetotelluric admittance function, defined by

$$e = -\frac{E_y(0)}{i\omega B_x(0)} = -\frac{E_y(0)}{E_y'(0)} \quad (187)$$
where the prime means derivative on \( z \). Recall that \( E_y(0) \) and \( B_x(0) \) are complex functions of frequency, derived from the field time series by Fourier analysis. Thus (187) shows that \( c \) is a complex function of frequency available from observation, and it is also computable from a model. We can solve (185) with a conductivity profile \( \sigma \) choosing an arbitrary nonzero value for \( E_y \) at \( z = 0 \); whatever value is used for the top boundary condition, disappears in the normalization performed in (187). Alternatively, and more precisely, \( c = E_y(0) \) is found by solving (185) with the boundary conditions \( E'_y(0) = -1 \) and \( E_y(h) = 0 \).

This prescription allows us to compare observed values of \( c \) with predictions from a model \( \sigma \), and then of course we can attempt to adjust the model until its predictions match the measurements: this is another example of an inverse problem. After the magnetic annihilator and the equivalent source theorem we have become used to the situation where having good agreement between a model and data does not guarantee any degree of approximation to the true structure in the Earth. However, in this case there is a uniqueness theorem – a function \( \sigma(z) \) that achieves exact matching of \( c \) for all \( \omega \) is the only possible solution.

There is an alternative to the admittance which provides some insight into the Earth’s response. We ask, How does \( c \) behave over a uniform halfspace? You will readily verify that, when \( \sigma = \sigma_0 = \text{constant}, \)

\[
c(\omega) = \frac{1 - i}{\sqrt{2\omega\mu_0\sigma_0}} = \frac{1}{2} z_0(1 - i)
\]  

(188)

where you may recall \( z_0 \) is the skin depth in (178). Thus the magnitude of the admittance \( |c(\omega)| \) is \( 1/\sqrt{2} = 0.707 \) times the skin depth. Approximating the system as though it were a uniform halfspace, we can calculate its effective resistivity directly from (188), a quantity called the apparent resistivity:

\[
\rho_a = \omega \mu_0 |c(\omega)|^2.
\]  

(189)

By plotting \( \sqrt{2}|c| \) on the depth axis and \( \rho_a \) as a resistivity, we can obtain a very crude picture of the electrical structure of the ground directly from admittance data.

As a method of model construction the use of apparent resistivity is obviously too primitive. The MT problem in one dimension is probably the best-understood nonlinear inverse problem in geophysics. We could devote a whole quarter to exploring the various approaches. For one powerful technique see Chapter 5 of my book, Geophysical Inverse Theory.

**Exercise:**

Solve (185) for a non-constant function \( \sigma \) of your choice. Plot \( c \) and \( \rho_a \) as functions of frequency and then plot \( \rho_a \) against \( \sqrt{2}|c| \); also plot \( 1/\sigma(z) \) on the second plot.

**6:3 Global Electromagnetic Sounding - Results**

In the huge frequency range of \( 3 \times 10^{-8} \) to 0.001 Hz, corresponding to periods of one year to 15 minutes, the fluctuating geomagnetic field at the Earth’s surface is driven by currents circulating within the magnetosphere, a giant loop called the
**ring current.** The main constituent is a small westward drift of protons whose velocity direction at any instant is mainly north or south as the particles bounce between mirror points at high latitudes. To an observer on the Earth’s surface the field appears to be almost uniform, or a \( Y_0 \) spherical harmonic in the external potential (with the \( z \) coordinate aligned with the main dipole). When a radially stratified conductor is subjected to a varying uniform field, the toroidal currents that flow have a \( Y_0 \) geometry, and the internally generated response is \( Y_0 \) also. For a radially stratified conductor, an externally generated field with \( Y_m \) geometry produces an internal part with the same symmetry. The ratio of the internal to external parts of the SH expansion at a particular frequency can be used as an electromagnetic response measure as we will now sketch.

Considering only axisymmetric fields, those with \( m = 0 \), we write the SH expansion:

\[
\Psi(r, \theta) = a \sum_l \left[ a_l \left( \frac{r}{a} \right)^l + b_l \left( \frac{r}{a} \right)^{l+1} \right] P_l(\cos \theta)
\]  

where \( a_l \) and \( b_l \) are the external and internal Gauss coefficients for some frequency \( \omega \). As we have seen it is possible to obtain \( a_l, b_l \) by Fourier analysis of surface observatory time series The complex ratio \( Q_l = b_l/a_l \) is a response measure sensitivity to the Earth’s conductivity. Weidelt (Zeitschrift für Geophysik, 38, pp 257-89, 1972) showed that \( Q_l \) could be mapped into a function that is the 1-dimensional admittance of an equivalent flat-Earth conductor with

\[
c = \frac{a_l}{l(l+1)} \frac{l - (l + 1)Q_l}{1 + Q_l}.
\]  

This response measure has become popular. Constable (J. Geomag, Geoelec., 45, pp 707-28, 1993) has performed an analysis of world-wide data in this form. Figure 6.3.1 shows the global admittance over the period range 3 days to half a year. The next figure, Figure 6.3.2, shows a range of models compatible with these admittances. Here is the caption for this figure, quoting from Constable’s paper: *Curve 1 has the minimum 2-norm of the first derivative of log \( \sigma \) vs log depth; curve 2 the minimum-norm of second derivative. Curve 3 is the same as 1 but with the penalty on the derivative relaxed at 660 km, and curve 4 is the same as 3 except with a prejudice for \( 4 \times 10^{-3} \) S/m applied between 200 and 410 km. All models terminate with an infinite conductor at 3,000 km. Measurements of lower mantle materials by Shankland: \( pw = \) perovskite plus wustite, \( p = \) perovskite, both with 11% iron.*
Figure 6.3.1

Figure 6.3.2