Magnetic Stability Analysis for the Geodynamo

by

Philip William Livermore

Submitted in accordance with the requirements for the
degree of Doctor of Philosophy

The University of Leeds
School of Earth Sciences
December 2003

To my parents

Acknowledgements

First and foremost I would like to thank my supervisor, Andy Jackson, for his help, support and attentiveness to my work over my last three years. I am also grateful for his thorough reading of this manuscript.

Over the duration of my Ph.D. I have had enlightening discussions with Rich Kerswell, John Lister and Mike Proctor, which have greatly contributed to the path of my studies. In addition, I would like to acknowledge discussions on linear algebra with Brian Stewart, transient behaviour with David Gerard-Varet and Glenn Ierley and general geodynamo topics with Dave Gubbins, Ashley Willis, Chris Finlay and Nick Teanby.

I thank Steve Gibbons for supplying LEOPACK, an abundance of FORTRAN codes on which some of my routines are based, in particular those producing graphical outputs. I also used his kinematic dynamo codes to benchmark my results. I also thank Alan Thomson at BGS for supplying the Ørsted data on which Appendix A is based.

I would like to thank everyone in the School of Earth Sciences for making my time here so enjoyable, including James Wookey, Steve A, Trudi, Lindsey, Simon, Tanya, Dave, Ian, Kristof, Andy C, James While, Susan, Nicola, Bryony, Neil and Andy G. I would also like to acknowledge some very fine ales consumed at the Eldon on many a friday night post-seminar pub session.

I made many friends in the Ballroom and Latin American Dancesport Society, in particular I would like to thank the coaches John and Julie, and some of the dancers: Fiona, Tash, Lauren, Mat, Mary, Sophie and others for making my time so enjoyable. I have also been involved in various campanology activities, for which I thank the Burley and Leeds Parish Church ringers who are too numerous to list here.

I would like to thank Rob for his company when climbing Munros and on other long distance walks, and for often supplying transport to get us out there.

I thank my parents and my brother Stephen for their continual support, both financial where necessary and otherwise, over the whole duration of my education.

Last but by no means least, I give a special thank you to my fiancée Fiona, whose love and support over the last three years has been unfailing, particularly in the ‘dark’ months of writing up.

Abstract

Kinematic dynamo modelling addresses the growth of magnetic fields under the action of fluid flow, effected in the geophysical case by the action of the convecting electrically conducting outer core and magnetic diffusion. In this study, we prescribe the flow to be stationary and geophysically motivated by some large scale process, for example, differential rotation or convection. Historically, workers have applied eigenvalue stability analysis and although in some cases Earth-like solutions have been found, the results hinge critically on the precise choice of flow. We therefore cannot attribute physical mechanisms in such cases since it is rather case dependent.

Following the success of more generalised stability techniques applied to the transition to turbulence in various non-magnetic fluid dynamical problems, we investigate both the onset of and subsequent maximised transient growth of magnetic energy. These approaches differ from the eigenvalue methodology due to the non-normality of the underlying operator, which means that superposition of non-orthogonal decaying eigenmodes can result in sub-critical growth. The onset or instantaneous instability problem can be formulated using variational techniques which result in an equation amenable to a numerical Galerkin method. For the suite of flows studied, we find robust results that can be physically explained by field line stretching. Convectively driven flows exhibit the greatest instability, the field structures giving this maximal instability being axisymmetric. All flows indicate an apparent asymptotic dependence on the magnetic Reynolds number R_m, which is reached when $R_m = O(1000)$. For the flows studied, we find improved lower bounds on R_m for energetic instability of between 5 and 14 times, compared to that resulting from the analytic analysis of *Proctor* (1977a).

In all the flows studied, without exception the geophysically dominant axisymmetric dipole field symmetry is preferentially transiently amplified. The associated physical mechanisms are either shearing of poloidal field into toroidal field by differential rotation, or advection into locations of radial upwelling followed by field line stretching in the convective case. Transient energy growth of $O(1000)$, which can be obtained when $R_m = 1000$ is robust and may explain the recovery of field intensity after a magnetic reversal. Assuming the flow to be a stationary solution of the geostrophic balance equation where buoyancy, pressure and the Coriolis forces are in equilibrium, we computed the geophysically scaled ratio of the Lorentz to the Coriolis forces and found it to be $O(1)$ for flows with a large convective component. This indicates that transient growth, in particular of axisymmetric fields that are ostensibly precluded by the theorem of *Cowling* (1933), can explain the entry into the non-linear regime without the need for eigenmode analysis.

Contents

List of Figures

List of Tables

Abbreviations

BG	The *Bullard and Gellman* (1954) flow
CMB	Core-mantle boundary
DJ	The *Dudley and James* (1989) flows
IC	Inner core
MDJ	A modified DJ flow
MHD	Magnetohydrodynamics
KR	The *Kumar and Roberts* (1975) flow
STW	The Sarson Thermal Wind flow, after *Sarson* (2003)
VGP	Virtual geomagnetic pole

Chapter 1

Introduction

1.1 A brief history

From the study of ancient rocks, paleomagnetists know that the Earth has had a magnetic field for much of its 4.6 billion year history. The discovery and appreciation of this fact by human civilisation has been limited to somewhat more recent times however. The ancient Greeks knew that magnetised rocks (*lodestones*) could attract iron. In fact the word *magnet* is derived from Magnesia, a city in then Asia Minor, where archaeological excavations have uncovered such rocks (*Parkinson*, 1983). It is the Chinese (ca. 1000 AD) though who are credited with the discovery that a freely pivoted magnetised needle will always point northwards. The origin of this unknown (seemingly constant) attractive force was believed to be either celestial in nature (since the north star was a fixed point) or that there was a great mountain of lodestone somewhere near the north pole (*Stern*, 2002).

An important scientific breakthrough in Europe came with the publication of William Gilbert's *De Magnete* in 1600. After a series of experiments involving a spherical lodestone, he realised that "the Earth globe itself is a great magnet". Thus was born the idea that the magnetic field was of internal origin (*Malin*, 1987).

From about the late 16th century, sailors knew that compasses did not point exactly northwards. The deviation in angle is called the *declination* and over the next few centuries it was mapped by the many global voyages that furnished the era. It was realised that the declination was not constant, but varied depending on both the position and year of the reading. Halley came up with the ingenious solution of permanently magnetised concentric spherical shells inside the Earth, that could move relative to one another and thus explain the apparent fluctuations.

The idea of permanent magnetism (the only form known at that time) suffered a fatal

blow when it was realised that almost all of the Earth's interior was too hot (above its Curie temperature) to retain its magnetisation. *Larmor* (1919), with reference to the sun speculated that a self-sustaining dynamo process could be responsible for maintaining a magnetic field. It was soon after discovered that the Earth's field had reversed many times in its history. What process can then account for these spatially and temporally complex characteristics of our magnetic field?

It is now known through seismic studies that the Earth has a solid inner core, a liquid outer core and a solid mantle; a schematic picture depicting this is shown in figure 1.1 (e.g. *Fowler*, 1990). Today we believe that the magnetic field is generated and sustained by fluid motion in the outer core, driven by convection. It is almost certainly the influence of the Earth's rotation on the fluid outer core that causes the generated magnetic field to be preferentially dipolar, with magnetic poles almost coinciding with the geographic poles.

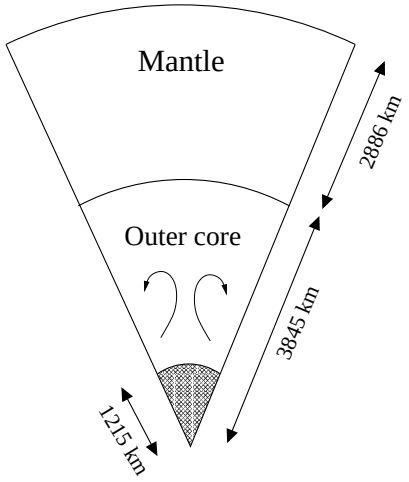


Figure 1.1: Schematic picture of the Earth. From the surface downwards: the solid mantle, liquid outer core and solid inner core.

The outer core consists mainly of liquid iron with a few dissolved light elements. Due to the action of pressure, the iron freezes preferentially at the centre of the Earth first (this is contrary to perhaps what might be expected since fluids generally freeze from the coldest point, in this case the outside, inwards). As the iron solidifies onto the inner core, the impurities are released. These are buoyant and so rise, creating a convective motion that stirs and stretches the magnetic field. Another principal source of energy is that of thermal convection generated by the latent heat of freezing at the inner core boundary. Fluid parcels are heated and become buoyant, rising in a manner similar to that of compositional convection. Although thermal convection is less efficient than compositional, it is important to understand how the Earth's dynamo (that

is, the mechanism by which the field is sustained) could operate without an inner core, that is, driven by heat alone. This is because thermodynamic calculations are still uncertain over whether the inner core has been in existence for the duration of the Earth's geomagnetic history (e.g. *Labrosse et al.*, 2001; *Gubbins et al.*, 2003).

1.2 Paleomagnetism

Detailed measurements of the Earth's magnetic field from the past 400 years or so (from sea voyages, fixed observatories and more recently satellites) comprise only about one ten millionth of the age of the field. Fortunately we know a lot more about its history from paleomagnetic measurements. When hot rock cools from above the Curie temperatures of its magnetic minerals, it records the magnetic field in which it is exposed at that time. In principle then, assuming the rock can be correctly dated and that sufficient samples are collected, the history of the field can be established. The main problem with this approach is that there is not enough data: volcanic eruptions have been too sporadic in time and do not offer good global coverage since they tend to cluster around tectonic plate margins or other hotspots. With insufficient data, the fitting of a global field to the local measurements is non-unique so resolution of large scale features is difficult. Another issue is that of dating accuracy: geologically fast field transitions (e.g. reversals) occur within the error bounds of the sample age estimation, so it is impossible to model anything but slow processes.

Other sources of paleomagnetic data are sedimentary cores taken from deep ocean floors. As the sediments are compressed into rock, they record the magnetic field that they were exposed to at that time. However, if the deposition rate is too slow the record tends to be filtered and smoothed (see *Merrill and McFadden*, 1999) and the data really only gives a time average. Nonetheless these are important as they give a contiguous data set of anything up to a few hundred thousand years back in time.

The Earth's magnetic field is around three billion years old and averaged over time it is well approximated by an axial dipole, aligned with the rotation axis. There is strong evidence that, for fast rotating bodies such as the Earth, rotational effects, manifested by the Coriolis force, have a significant effect on the flow patterns in the fluid core (see *Busse*, 2002, for example). Such a system favours axially aligned cylindrical structures, so perhaps this time averaged effect is not unexpected. The paleomagnetic data also shows that the average intensity (strength) of the field is slightly (around 30%) weaker over the past 0.8Ma than it is today, although significant deviations have occured during transitions of the field (*Dormy et al.*, 2000).

1.3 Reversals

Perhaps the most well known fact about the Earth's field is that it has reversed polarity many times over the course of its lifetime. The current reversal rate is about 4.5 per Ma, although this has changed over time (see e.g. *Constable*, 2000). On average, a reversal takes 4600 years to complete, making it geologically almost instantaneous. As a fraction, only 2% of the field's time has been spent in transitions, making it quite unlikely that lava flows of sufficient proliferation and of the correct age are stumbled upon, giving us some idea as to the processes that accompany such events. Nonetheless, this does occasionally happen — one of the highest profile data sets is that from Steens Mountain (Oregon, USA) which captures just such a phenomenon (*Prévot et al.*, 1985; *Mankinen et al.*, 1985). The ancient volcano erupted many times 16 Ma ago, producing a large array of lava flows, all of slightly different ages. Measurements from these flows reveal that a reversal took place during the period of volcanic activity. In particular, two flows recording the transition captured some very strange behaviour. The field recorded in the quickly cooled exterior of the flows was strikingly different from that of the slowly cooled interior. Over the days or weeks it would have taken to cool completely, the field changed by up to 6° per day, a huge change compared to current typical values of secular variation (field change over time) of around 0.5°/yr.

Although it has been argued that such results could be due to a material property of the rock and not of the field recorded, further studies (*Camps et al.*, 1995; *Coe et al.*, 1995; *Camps et al.*, 1999) have confirmed that this is not the case: the results still stand. External influences of magnetic storms have been called upon to explain these swift field changes (*Ultré-Guérard and Achache*, 1995) but this is discounted by *Jackson* (1995). Invoking an internal mechanism (seemingly the only explanation left) has its own set of problems however. In order to get such a rapid field movement, either the typical velocity in the outer core must show a strong increase (by at least 1000 times) or we need to have a spatially much more complex field. Neither of these are particularly appealing and are not manifested in the recent simulations showing reversals (e.g. *Glatzmaier and Roberts*, 1995). In any event, the semiconducting mantle will filter out signals of duration less than six days (*Coe et al.*, 1995) although this is a rather optimistic estimate based on a low value of lower mantle conductivity. Perhaps a more realistic timescale is that of about one year, consistent with observatory data from the Earth's surface (*Merrill and McFadden*, 1999). Hence even if the signal were to be of internal source one would not, on this basis, expect to be able to measure it. As an alternative, conductivity inhomogeneities could distort the magnetic signal as it passed through the mantle, causing localised apparent field movement. Whether or

not this is likely remains to be investigated.

Almost all documented field reversals are preceded by a period of low intensity, of about 10–20% of the pre-transitional field (*Merrill and McFadden*, 1990), presumably making the dynamo unstable. In many cases, for example in the data from Steens Mountain, Kauai (Hawaii) (*Bogue and Paul*, 1993) or La Palma (Canary Islands) (*Valet et al.*, 1999) where such transitions have been adequately recorded by lava flows, the post-transitional field intensity is higher than average. This state could either be a transient phenomenon associated with a reversal (*Bogue*, 2001) or the start of a slow decay over time (as in the controversial *sawtooth* pattern of *Valet and Meynadier* (1993); the higher the peak the longer the time until the next transition). In both cases there is evidence of swift field recovery after reversal in 1–10 Ka. In the sawtooth pattern, the post-transitional field is stable and slowly decays, presumably dominated by an axial dipole. This stability is also found in the data from Kauai and La Palma, although the high intensity lasts around 10Ka and 100Ka respectively (although only the latter time is long enough to be consistent with the sawtooth pattern). In contrast, records from Steens Mountain show that the field is more variable after the reversal than before; thus the newly established dynamo was unstable (see *Bogue*, 2001, for a comparison).

In virtually every case paleomagnetic measurements yield the direction of the ancient field at the sampling site. From values of the declination and inclination (the *dip* angle of the magnetic field below horizontal) the orientation of a simple dipole field that would reproduce such readings can be determined. If the field was purely dipolar then this process would exactly capture the true field. In addition, different locations would agree; the fact that such calculations produce different virtual geomagnetic dipoles depending on the sampling location means that the transitional field is highly complex and not dominated by the usual dipolar structure. Nonetheless, if there are sufficient readings at different points in time, we may track the virtual geomagnetic pole (VGP) during a reversal. The paths that the VGPs have taken have been linked to lower mantle heterogeneities (*Laj et al.*, 1991), although a recent study (*Valet and Herrero-Bervera*, 2003) suggests there is insufficient evidence to substantiate preferred longitudes.

The basic picture of reversals emerging is this:

1. The intensity before and during a reversal is low, about 10–20% of the pre-transitional field.

2. The reversals are highly complex, dominated by non-axial dipole fields and take on average 4600 yrs.

3. After the reversal the field intensity recovers very fast by some transient mechanism, in

about 1-10 Ka, although the stability of this post-transitional state seems variable.

1.4 The field at Epoch 2000

Satellite measurements of the Earth's magnetic field can be downwards continued to the core-mantle boundary (CMB) under various assumptions. In this way we can produce maps of the field at the edge of the dynamo region. The method is essentially that of a least squares fitting of the data to a model that gives the field at any point above the CMB. See Appendix A for details. Figure 1.2 shows a contour map of the radial component of field at the CMB on 1st January 2000 (orange is positive, blue negative; continents are plotted for reference). Strong colours indicate an increased density of field lines (field flux). The field is principally a dipole, with field lines leaving the core in the southern hemisphere and returning in the northern hemisphere. The picture is complicated somewhat by patches of reversed flux, for example, those beneath the southern Atlantic.

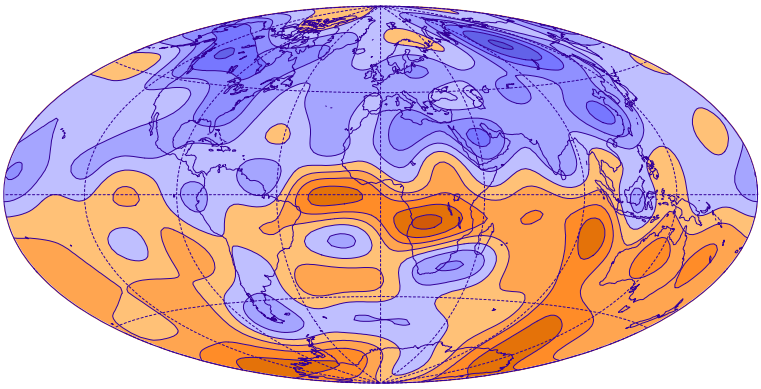


Figure 1.2: The radial component of the Earth's magnetic field at epoch 2000. Contour interval is 200 μT. Orange is positive, blue is negative.

The radial component of field is the only quantity that is guaranteed to be continuous across the CMB interface. Due to the possible presence of a current sheet, the other components will not necessarily be indicative of the field at the (inside) edge of the outer core (*Jacobs*, 1987).

The field at the Earth's surface can be approximated by the superposition of three mutually perpendicular dipoles. Table 1.1 shows the strength of each component, derived from the analysis in Appendix A.

The axial dipole component dominates by around an order of magnitude, and such values will be used in subsequent chapters to scale theoretically modelled fields to the geophysical case.

Dipole orientation	Associated north magnetic pole	Magnitude (nT)
Axial	$\theta = 0$	-29616.31
Equatorial	$\theta = \pi/2, \phi = \pi/2$	5185.72
Equatorial	$\theta = \pi/2, \phi = 0$	-1729.10

Table 1.1: Strength of each dipole component and the locations of the associated north magnetic pole. Coordinates are colatitude (θ) and longitude (ϕ).

1.5 Thesis outline

This thesis is concerned with the possible mechanisms of generating magnetic field inside the liquid outer core. In order to make progress, we use the kinematic assumption, whereby we impose some geophysically motivated flow structure in the outer core and determine whether or not magnetic fields are unstable, that is, will a small 'seed' field grow? We address this problem in three different ways: standard linear eigenvalue methods, the onset of magnetic energy growth and finite time growth of magnetic energy. These three schemes differ in general due to a property termed *non-normality* of the controlling system. In chapter two we detail the advances that have been made to date, mostly in the eigenmode formulation, which suffers from its sensitivity on the exact choice of the particular flow used, indicating that robust physical mechanisms describing such field growth are hard to find.

Chapter three discusses the more theoretical issues appropriate to our study, in particular the defining equations and subsequent manipulations that we shall base our numerical studies on. Chapter four details the numerical techniques that we use, including a fully spectral Galerkin method. Calculations relating to the onset of magnetic energy instability are presented in chapter five. We find excellently converged results that indicate a robust magnetic energy generation mechanism associated with the stretching of field lines. Whether or not such growth continues however is addressed in chapter six, where we investigate how large magnetic energy can become before its eventual decay, in the cases where it is not infinitely sustainable. We find that for a geophysical choice of flow, Earth-like axisymmetric dipolar fields may grow sufficiently large as to introduce non-linear effects, signalling the demise of the kinematic assumption. Although the theorem of *Cowling* (1933) prohibits infinite sustenance of such fields, their finite time behaviour could be of crucial importance in the geodynamo.

Chapter 2

Kinematic dynamo modelling

2.1 Introduction

> Dynamos exist. One can buy them in shops.
> "What", one may ask, "is all the fuss about?" (*Roberts*, 1992)

Dynamos can indeed be bought from shops (e.g. to power bicycle lights) and are essentially electrical motors run in reverse, turning mechanical energy into electric currents. Associated with every current is a magnetic field, for example, as manifested in concentric circular field lines surrounding a current carrying wire. Thus producing magnetic field and current are essentially one and the same, and it would therefore appear that generating magnetic fields is an easy task. Extending the argument, perhaps the same can be said of the Earth's core?

However, the geometry crucially differs between the two cases: in the motor-dynamo the current is forced to flow in directions dictated by the wires and moreover the circuit is not simply-connected, i.e. at all times there is not necessarily a path joining every point to every other in the circuit. In the Earth the situation is much more simple: currents may flow at all times in any direction that they so choose. It is quite possible however that these may short-circuit, thus completely ruling out dynamo action. It is certainly not obvious that a dynamo can operate in such a simple geometry.

2.2 Equations of Electromagnetism

2.2.1 Maxwell's equations

Maxwell's equations governing the behaviour of a magnetic field and its associated electric field in a moving conducting medium have been documented extensively elsewhere (see *Davidson*,

2001; *Roberts*, 1994, for example). In fact since the relevant velocities in the magnetohydrodynamic problem applicable to the Earth's core are very much less than the speed of light we may use the pre-Maxwell equations, a simpler subset. These describe the relation between the magnetic field **B**, the electric field **E**, the current density **J** and the time t:

$$\mu_0 \mathbf{J} = \boldsymbol{\nabla} \times \mathbf{B} \tag{2.1a}$$

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2.1b}$$

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \tag{2.1c}$$

The value $\mu_0 = 4\pi \times 10^{-7} H\, m^{-1}$ is the permeability of free space. Ampère's law (equation 2.1a) describes the magnetic field associated with an electric current. A simple example of this is the concentric circular field lines surrounding a straight current carrying wire. Faraday's law (equation 2.1b) relates a changing magnetic field with an induced electric field. For example, moving a magnet towards and away from a fixed wire will cause an electric field to be induced and therefore a current will flow if the circuit is closed. The solenoidal condition of equation (2.1c) can be expressed in integral form as

$$\int_{\partial V} \mathbf{B} \cdot \mathbf{dS} = 0 \tag{2.2}$$

for the boundary ∂V of any volume V. This law therefore excludes monopolar fields since they have no angular dependence and cannot satisfy relation (2.2) unless identically zero. The above three laws require a closure relation which is taken as Ohm's law. In general, the conductor will be moving with velocity **u** so it is written in the form

$$\mathbf{J} = \sigma \left(\mathbf{E} + \mathbf{u} \times \mathbf{B} \right) \tag{2.3}$$

where σ is the electrical conductivity. This can be simply understood as the current being proportional to the electric field but with an extra inductive term due to the movement of the conductor through the magnetic field.

The Earth's dynamo is called self-sustaining (it has no external energy sources) and homogeneous (it is simply connected and has isotropic conductivity). Currents flowing in the core generate heat due to Ohmic dissipation: this represents a drain on the system and without any generation process the magnetic field would decay. The mechanism by which the dynamo maintains the field can be understood in the following way. Motion in the outer core across

magnetic field lines induces currents (Ohm's law) that have their own field associated with them (Ampère's law). If the generated field reinforces the existing field then self-exciting dynamo action can occur (the process needs no external sources apart from a 'seed field' to get it started). If not then the dynamo will fail and the field will decay.

At first glance, given a 'seed field' this mechanism appears to be able to magnify the magnetic field indefinitely; indeed in theory this could occur. However, the increase in magnetic energy comes at the expense of the kinetic energy of the fluid; in particular the magnetic field exerts a *Lorentz force* on the conducting fluid so that not only has the fluid to do extra work against this magnetic 'resistance', but in general this force will also modify the flow pattern over time.

2.2.2 The induction equation

It is instructive (and sufficient) to consider only the evolution of the magnetic field since all other relevant quantities can be derived where needed using equations (2.1) above. By eliminating the electric field and current density, we arrive at the magnetic induction equation describing the rate of change of **B** in a conducting fluid moving with velocity **u**:

$$\frac{\partial \mathbf{B}}{\partial t} = \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}) - \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) \tag{2.4}$$

where the magnetic diffusivity is denoted $\eta = (\mu_0 \sigma)^{-1}$. In the case of constant finite magnetic diffusivity we may non-dimensionalise equation (2.4) by typical scales of length $\mathscr{L}$, of velocity $\mathscr{U}$ and (magnetic diffusion) time $\mathscr{L}^2/\eta$ to obtain:

$$\frac{\partial \mathbf{B}}{\partial t} = R_m \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}) + \nabla^2 \mathbf{B} \tag{2.5}$$

where we write $\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times = \boldsymbol{\nabla}\boldsymbol{\nabla} \cdot - \nabla^2$. The magnetic Reynolds number R_m is defined to be $\mathscr{U}\mathscr{L}/\eta$ and is the ratio between the action of the flow and diffusive terms. When $R_m = 0$ the equation remaining is analogous to a scalar diffusion equation: indeed in this case the magnetic field energy exponentially decays with a nondimensional e-folding time of $1/\pi^2$ (e.g. *Backus*, 1958). In the other extreme when $R_m \to \infty$, the magnetic field lines become frozen into the fluid (see e.g. *Roberts*, 1967). This is a useful picture to have: in this sense R_m is a measure of how much the fluid grips the magnetic field lines as it moves around. More precisely, it is the magnetic flux through any closed contour in the fluid that is conserved in this so called 'frozen flux' limit; this gives a useful way of describing how magnetic field is generated. Consider a

bundle of magnetic field lines having a cross sectional area A. Suppose the fluid motion stretches the bundle; if the fluid is incompressible then A must decrease and hence the field intensity must locally increase, since the flux must remain constant.

2.3 The general dynamo problem

2.3.1 The Navier-Stokes equations

In general the flow $\mathbf{u}$ of the conducting medium must be derived from the Navier-Stokes equations in a rotating frame of reference. The medium is usually, but not always, taken as incompressible except that we allow density differences to drive convection (the *Boussinesq* approximation). In addition, since the geodynamo is driven by chemical and thermal energy sources we must also solve the relevant equations for these quantities simultaneously. On the simplification that only heat is important in convection, the following formidable set of magneto-hydrodynamic (MHD) equations (*Fearn*, 1998) must be solved:

$$R_o\left(\frac{\partial \mathbf{u}}{\partial t}+(\mathbf{u}\cdot\boldsymbol{\nabla})\,\mathbf{u}\right)+\hat{\mathbf{z}}\times\mathbf{u}=-\boldsymbol{\nabla}\Pi+qRaT\,\mathbf{r}+E\nabla^2\mathbf{u}+(\boldsymbol{\nabla}\times\mathbf{B})\times\mathbf{B} \tag{2.6a}$$

$$\frac{\partial \mathbf{B}}{\partial t}=\boldsymbol{\nabla}\times(\mathbf{u}\times\mathbf{B})+\nabla^2\mathbf{B} \tag{2.6b}$$

$$\frac{\partial T}{\partial t}+(\mathbf{u}\cdot\boldsymbol{\nabla})\,T=q\nabla^2T+h \tag{2.6c}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B}=0 \quad \boldsymbol{\nabla}\cdot\mathbf{u}=0 \tag{2.6d}$$

The scalar function Π is the generalised pressure, the vector $\hat{\mathbf{z}}$ is a unit vector in the direction of the Earth's angular velocity, T is the temperature, $\mathbf{r}$ is the position vector and h are the thermal sources. The nondimensional parameters appearing give the strengths of various effects relative to the Coriolis force (written non-dimensionally as $\hat{\mathbf{z}}\times\mathbf{u}$) and are: the modified Rayleigh number R_a (the vigour of convection), the Rossby number R_o (the importance of inertial effects), the Ekman number E (the importance of viscosity) and the Roberts number q (a ratio of diffusivities). The definitions are

$$R_a=\frac{g_0\alpha\beta r_o^2}{2\Omega\kappa} \qquad R_o=\frac{\eta}{2\Omega r_o^2} \tag{2.7a}$$

$$E=\frac{\nu}{2\Omega r_o^2} \qquad q=\frac{\kappa}{\eta} \tag{2.7b}$$

where the gravitational force is $-g_0\,\mathbf{r}$, α is the thermal expansion coefficient, β is the maximum temperature gradient, r_o is the radius of the outer core, Ω is the Earth's rotation rate, κ is the

thermal diffusivity and ν is the viscosity. The parameters take on typical values of $R_o = 10^{-9}$, $q = 10^{-5}$, $E = 10^{-15}$ and $R_a = 10^8$ (although this is largely unknown).

Although at first sight the minuscule size of E would seem to simplify the problem since the viscous force could be neglected, it transpires that such a simplification does not work and we need to include very thin boundary layers, far beyond the resolution of any current simulations. It is for this reason that numerical work is currently limited by the size of the Ekman number. Typical simulations use $E = 10^{-3} - 10^{-6}$ corresponding to the viscosity of the liquid outer core being many orders of magnitude too high (for a review see *Kono and Roberts*, 2002). Notice that in this non-dimensionalisation the magnetic Reynolds number is determined *a posteriori* from the computed flow so that it does not appear as an adjustable parameter in the model. This is also true of the Elsasser number Λ, measuring the nondimensional size of the magnetic field.

2.3.2 The Lorentz force

In equations (2.6) the magnetic field is allowed to affect the flow via the Lorentz force, written non-dimensionally as $(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$. It is quadratic in the size of $\mathbf{B}$ and note that it is pointwise perpendicular to the field producing it. There is no reason to suppose that even though $\mathbf{B}$ might be large the same can be said of the Lorentz force, since for example, it might identically vanish if the electric current (proportional to $\boldsymbol{\nabla} \times \mathbf{B}$) happened to be parallel to the magnetic field. There is little *a priori* known about the size and effect of the Lorentz force associated with a kinematic dynamo; instead it must be computed from the generated magnetic field after the computation. It may be re-written in the following way:

$$(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B} = -\frac{1}{2}\boldsymbol{\nabla}(\mathbf{B}^2) + (\mathbf{B} \cdot \boldsymbol{\nabla})\,\mathbf{B} \tag{2.8}$$

The first term on the right hand side is a magnetic pressure and may be incorporated into the general pressure expression Π. The second term represents a tension in the magnetic field lines that exerts a force on the fluid.

2.3.3 A few comments on energy

The geometry of the system considered is that of a sphere V (the core) containing fluid of constant magnetic diffusivity that moves with a velocity $\mathbf{u}$; the external region (the mantle) denoted $\hat{V}$ is modelled as a quiescent electrical insulator. Taking the inner product of the non-dimensional induction equation (equation 2.6b) with $\mathbf{B}$ and integrating over all space yields the relation (see

chapter three for details)

$$\frac{1}{2}\frac{d}{dt}\int_{V+\hat{V}} \mathbf{B}^2\, dV = \int_V \mathbf{B}\cdot\boldsymbol{\nabla}\times(\mathbf{u}\times\mathbf{B})\, dV - \int_V |\boldsymbol{\nabla}\times\mathbf{B}|^2\, dV \tag{2.9}$$

When the flow is quiescent ($\mathbf{u}=\mathbf{0}$) the total magnetic energy of the system monotonically decays and is lost to Ohmic heating representing a drain on the system. The generation term (first on the right hand side) is responsible for matching this loss and may be rewritten in the case where the flow vanishes on the boundary of the region as

$$\int_V \mathbf{B}\cdot\boldsymbol{\nabla}\times(\mathbf{u}\times\mathbf{B})\, dV = -\int_V \mathbf{u}\cdot(\boldsymbol{\nabla}\times\mathbf{B})\times\mathbf{B}\, dV \tag{2.10}$$

This is identical to the work done in the Navier-Stokes equation by the flow against the Lorentz force. Hence no energy is lost in the system by the dynamo mechanism: energy is converted faithfully between magnetic and kinetic forms. One might expect because of Lenz's law [1] that the system will always oppose change of magnetic flux so that the two equivalent energy forms in (equation 2.10) are positive. That is, the fluid always does work on the magnetic field and always represents a drain on the kinetic energy. This is demonstrably true as calculations in chapter six indicate. Another noteworthy point is that the magnetic pressure does no work on the system (indeed neither does the general pressure) if the flow is incompressible and the flow has a vanishing normal component on the boundary.

$$\int_V \mathbf{u}\cdot\boldsymbol{\nabla}\Pi\, dV = \int_{\partial V} \Pi\,\mathbf{u}\cdot\mathbf{dS} = 0 \tag{2.11}$$

Hence only the maxwell tension effects the energy of the flow.

2.3.4 Rationale of the kinematic assumption

The general magnetohydrodynamical (MHD) problem of solving simultaneously for $\mathbf{u}$, $\mathbf{B}$ and T is formidable and has not yet been performed in a parameter regime comparable to the Earth. Indeed, resolving the Ekman boundary layers adequately may take many years of computational improvement to achieve. Despite full MHD simulations successfully showing many Earth-like characteristics such as the correct field strength and even magnetic reversals, there is no guarantee that these features will remain robust as the parameters tend towards their geophysical values. Indeed, there is concern that dynamo action may disappear as the magnetic Prandtl num-

[1] Lenz's law states that an electric current induced by a changing magnetic field is associated with an magnetic field that resists the change. This means that if $\mathbf{B}$ is increasing, the system will try to slow the change by creating a Lorentz force that acts against the flow thereby slowing the generation.

ber, $P_m = \nu/\eta$, decreases from typical simulation values of $O(1)$ to Earth-like values of $O(10^{-6})$ due to enhanced Ohmic dissipation.

In any event, the numerical solutions obtained are extremely complicated and are not easily understood in terms of known concepts. There is therefore a gap for simple models explaining not only why such a dynamo might exist in the Earth's core, but how it might operate. Kinematic models treat the flow **u** as prescribed and (typically) stationary so that the induction equation may be solved in isolation. This assumption makes the problem simpler because the Navier-Stokes equations no longer need to be solved; in addition the system is linear in **B** since the Lorentz force is neglected. The rationale behind this simplification is to study and learn how magnetic fields can be generated by simple flows; the fact that this has turned out harder than expected (see next section) leads to an increased justification in this isolated study. It is hoped that from these simple models some insight might be gained into generic dynamo processes with application not only to the Earth but to other planets and stars having some kind of dynamo mechanism. In addition, by comparing calculated fields with their geophysical counterparts we might be able to constrain the types of flow that generate Earth-like features without solving the full Navier-Stokes equations.

2.4 The kinematic dynamo problem

The kinematic dynamo problem that we shall consider is the following. A sphere V of radius $\mathscr{L}$ contains fluid of constant conductivity σ that moves with prescribed velocity **u** and is embedded in an infinite insulator $\hat{V}$. If the flow can be physically motivated we assume it to be a (usually steady) solution of the full (non-magnetic) Navier-Stokes equations perhaps with some kind of extra forcing term. The assumption that the Lorentz force should be neglected is only valid with a vanishingly small magnetic field. Fortunately the induction equation is linear in **B** so that any dynamo mechanism is independent of its magnitude. We therefore formulate the problem as one of an instability from a non-magnetic base state and pose the question: 'can a small adventitious seed field grow in such a flow?'.

Here we are more interested in whether an infinitesimal field can grow and not whether a finite steady field can be maintained. Such an extension is clearly possible although we would start to move away for the true physical system due to the finite Lorentz force.

Geophysically, the sphere of moving fluid represents the liquid outer core; in general a solid inner core may be modelled by prescribing a quiescent region in its centre. The assumption that the conductivity is constant stems from the belief that the outer core is well mixed. Although

the mantle is not a perfect insulator its conductivity is around four–five orders of magnitude less than that of the core, so this assumption is a fair approximation. Typical estimates are 10 Sm^{-1} for the lower mantle (*Shankland et al.*, 1993) compared to $5\times10^5\,Sm^{-1}$ for the outer core (*Gubbins and Roberts*, 1987).

The problem is then to solve the (non-dimensional) induction equation (equation 2.12)

$$\frac{\partial \mathbf{B}}{\partial t} = R_m \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}) + \nabla^2 \mathbf{B} \tag{2.12}$$

as previously discussed, subject to the conditions that the field is everywhere divergence-free, finite and perhaps most importantly in the geophysical case, that the field must be self-generated: all external sources of energy are excluded. We note that in this nondimensional system, the sphere V is of radius 1. For a given $\mathbf{u}$ and R_m we seek growing field solutions.

In the case of the Earth, $\eta \approx 2m^2/s$, $\mathscr{U} \approx 10^{-4} m/s$ and $\mathscr{L} = 3485\times10^3 m$ so that $R_m \approx$ 200. Clearly there is a vast amount of uncertainty in this figure which is attributed particularly to the poorly constrained velocity estimate. Geophysically relevant values of R_m are therefore probably in the range 0–1000.

2.4.1 Some notation

We now introduce some notation that will facilitate a review of kinematic dynamos to date; full details are presented in chapter three.

In order to satisfy the solenoidal condition, following *Elsasser* (1946) we typically expand the magnetic field in poloidal ($\mathbf{S}$) and toroidal ($\mathbf{T}$) form $\mathbf{B} = \mathbf{S} + \mathbf{T}$ where

$$\mathbf{S} = \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times [S(\mathbf{r})\,\hat{\mathbf{r}}] \tag{2.13}$$

$$\mathbf{T} = \boldsymbol{\nabla} \times [T(\mathbf{r})\,\hat{\mathbf{r}}] \tag{2.14}$$

We use spherical polar coordinates (r,θ,ϕ) and denote the position vector by $\mathbf{r} = r\hat{\mathbf{r}}$. Note that toroidal vectors have no radial component (see chapter three). It is immediately apparent that any field of this form will have zero divergence. In addition, the defining scalar functions of position, S and T, are typically expanded in terms of spherical harmonics defined as

$$Y_l^m(\theta,\phi) = P_l^m(\cos\theta) \left\{ \begin{array}{c} \sin m\phi \\ \cos m\phi \end{array} \right\} \tag{2.15}$$

The functions of colatitude, P_l^m, are the associated Legendre functions. For ease of notation we

use the Greek letters to denote a particular harmonic, corresponding to a choice of l, m and the sine or cosine function in azimuth. We can represent, for example, the poloidal part of the field in terms of 'vector spherical harmonics' where the scalar functions are unknown:

$$\mathbf{S} = \sum_{\alpha} \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times [Y_\alpha(\theta,\phi) S_\alpha(r) \hat{\mathbf{r}}] \tag{2.16}$$

In addition, when writing axisymmetric vectors such as $\mathbf{S}_1^0$, we omit the unnecessary superscript of 0. The flows that we study will be incompressible, an assumption reasonable for the Earth where the density difference across the fluid outer core is around 1.2 (from the Preliminary Earth Model, *Dziewonski and Anderson*, 1981). The flow $\mathbf{u}$ is therefore divergence free and we write $\mathbf{u} = \mathbf{s} + \mathbf{t}$ for the poloidal and toroidal components respectively (always using lower case). We refer to vector spherical harmonics in an identical way to those for the magnetic field. It transpires that the condition that the magnetic field is self-generating (due only to internal sources) may be represented easily in poloidal-toroidal form (see chapter three for details) as

$$\frac{dS_\alpha}{dr}(1) + l_\alpha S_\alpha(1) = 0 \qquad\qquad T_\beta(1) = 0 \tag{2.17}$$

The field in the exterior $\hat{V}$ is completely determined by the field on the boundary but only the poloidal part may escape the core to be observed at the Earth's surface; the toroidal field inside the core is unknown. Finiteness of the field follows automatically assuming that S_α and T_β are well behaved and that they satisfy a certain regularity condition at the origin (see chapter three).

2.5 Anti-dynamo theorems

After the proposal of *Larmor* (1919) that a magnetic field might be maintained by a self-sustaining dynamo mechanism, the subject suffered a severe blow with the publication of Cowling's celebrated theorem in 1933. It essentially states that a steady axisymmetric poloidal field cannot be maintained. Such a configuration is to be found in sunspots (the subject of the paper) where field lines, to a first approximation, lie in planes parallel to the axis of symmetry. That the Earth's field is also principally of this symmetry led to his conclusion:

> The theory proposed by Sir Joseph Larmor...is examined and shown to be faulty.

In subsequent years this theorem has been generalised by *Hide and Palmer* (1982) and *Ivers and James* (1984) to include time dependent fields, compressible media and spatially dependent diffusivity. There is however some ambiguity about how long these ill-fated axisymmetric fields

would take to decay in such cases. If the decay time greatly exceeds that due solely to Ohmic dissipation (around 20,000 years in the Earth) then these systems cannot be dismissed as being geophysically or astrophysically irrelevant. We summarise the so called *anti-dynamo* results as follows (as in *Roberts*, 1994, p.28)

- Axisymmetric fields may not be sustained by dynamo action.
- A magnetic field that is a function of only two spatial coordinates cannot be sustained.
- A toroidal flow may not maintain any magnetic field.
- An incompressible motion in which $u_z = 0$ for some Cartesian coordinate system (x, y, z) may not maintain any magnetic field.
- In a sphere, a velocity having only a radial component cannot maintain a magnetic field.

In summary, if the flow or field is too simplistic or has too much symmetry then a dynamo cannot operate. It is important however to realise what is not included in the above negative results. For example, non-axisymmetric fields may be maintained by axisymmetric (or non-axisymmetric) flows. There is also no poloidal counterpart to the toroidal flow theorem: a poloidal flow may generate a magnetic field (*Love and Gubbins*, 1996a). It is unfortunate though that working homogeneous dynamos are necessarily complex in structure and a simple model of the Earth cannot be constructed.

2.6 A review of kinematic modelling

2.6.1 After Cowling's theorem

It was some time after the publication of Cowling's theorem that any progress was made on the Earth's dynamo problem. Simple axisymmetric fields were ruled out and so was, perhaps, a self-sustaining mechanism in the Earth's core. It was not known whether a general anti-dynamo theorem might exist thus ruling out any kind of generative mechanism altogether. An outstanding problem though was that of explaining the observed secular variation: the Earth's field varied on a timescale of years, whereas it was known that the magnetic field would take tens of thousands of years to decay freely. *Elsasser* (1947) was the first to suggest that a complex mechanism might be taking place within the outer core, converting magnetic field between symmetries thus breaking the constraints of Cowling's theorem. He suggested that differential rotation could convert an axisymmetric dipole field, represented by the $\mathbf{S}_1$ vector harmonic, into $\mathbf{T}_2$. Figure 2.1

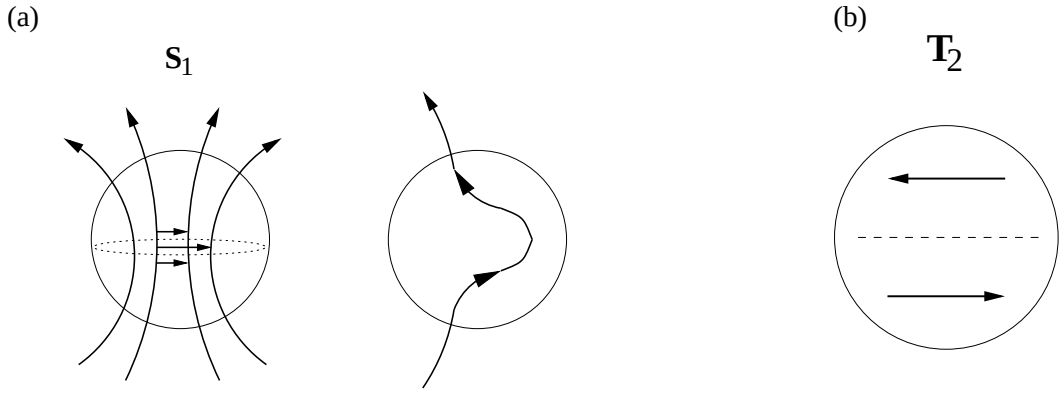


Figure 2.1: The effect of a $\mathbf{t}_1$ flow on an $\mathbf{S}_1$ axial dipole field. The sphere is viewed side on and the field lines are stretched by the differential rotation, being faster midstream than towards the boundaries (picture a). The field component $\mathbf{T}_2$ is created, amongst others (picture b).

shows a schematic diagram depicting this effect. The $\mathbf{S}_1$ field is sheared by differential rotation, that is, a flow with no radial component whose velocity differs from that of a rotating solid body so that over time fluid particles move apart from each other. In this case the flow is faster at the equator than towards the poles, so the axial field is stretched as shown in figure 2.1(a). This new field has a $\mathbf{T}_2$ component, shown in figure 2.1(b). The problem remained of how to close the loop, that is, how to change the newly created $\mathbf{T}_2$ field back into $\mathbf{S}_1$ so that the process might continue indefinitely. Note that reversing the flow will not work: a toroidal field can never be sheared by a toroidal flow to create poloidal field.

Of the two mechanisms he suggested, namely turbulence and tilted flows, the latter was pursued by *Bullard* (1949) who argued qualitatively that a flow component $\mathbf{s}_2^{2c}$ (describing a tilted convection roll) might be sufficient to regenerate the axial dipole field. Figure 2.2 shows a schematic picture of this process. Solid arrows show the effect on the field of the $\mathbf{t}_1$ flow; dashed arrows show the effect of the flow component $\mathbf{s}_2^{2c}$. Thus an $\mathbf{S}_1$ field, if acted upon in this manner, might be sustained creating byproducts of $\mathbf{T}_2$, $\mathbf{T}_2^{2s}$ and $\mathbf{T}_2^{2c}$ fields enroute. The emphasis at this time was on finding a mechanism whereby the principally observed geophysical field component $(\mathbf{S}_1)$ could be sustained; no attention was given to other field symmetries. Since this four component field is not axisymmetric, Cowling's theorem does not apply and the idea is plausible.

The choice of the $\mathbf{t}_1\,\mathbf{s}_2^{2c}$ flow was geophysically motivated by differential rotation and convection. *Bullard* (1949) suggested that radial motions stirred the outer core due to buoyant material rising. If a density difference existed between the upper and lower layers of the outer core then an angular momentum argument could be invoked to suggest that the inner sec-

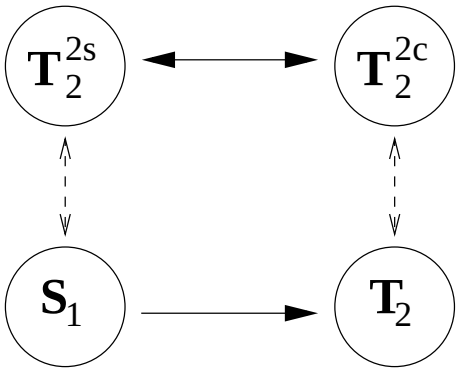


Figure 2.2: The interactions between the lowest degree harmonics involving the axial dipole $\mathbf{S}_1$ which forms a closed loop. Solid arrows denote coupling by $\mathbf{t}_1$ motion; dotted by $\mathbf{s}_2^{2c}$. If all links of this chain work at some non-zero rate then the axial dipole harmonic could be infinitely sustained by this process.

tion would be accelerated and the outer section decelerated relative to the mantle; in this way, differential rotation could arise.

2.6.2 The Bullard and Gellman model

The first quantitative model of a self-sustaining process taking place inside the Earth's core was proposed by *Takeuchi and Shimazu* (1953). This work was closely followed by *Bullard and Gellman* (1954) (henceforth BG); both papers extended the suggestions of Elsasser and Bullard that if a flow was chosen carefully, through a sequence of interactions the original field might be amplified and lead to a self-sustaining process. We introduce the terminology of BG below although it is largely similar to the Japanese work published a year previously.

In order to solve the induction equation, BG expanded the unknown field in vector spherical harmonics:

$$\mathbf{B} = \sum_{\gamma} \mathbf{S}_\gamma + \mathbf{T}_\gamma \tag{2.18}$$

In order to derive differential equations for the unknown defining scalar functions, they took the dot product of the induction equation (using the same vector harmonics) with scalar functions of unit dependence, and integrated over a spherical shell of radius r. If the velocity is chosen to be a finite sum of vector spherical harmonics (in this case $\mathbf{t}_1$ and $\mathbf{s}_2^{2c}$) with prescribed radial behaviour then

$$N_\gamma^2 \frac{\partial B_\gamma}{\partial t} = R_m \sum_{\alpha,\beta} \left(u_\alpha B_\beta B_\gamma \right) + N_\gamma^2 \left[\frac{\partial^2 B_\gamma}{\partial r^2} - \frac{l_\gamma (l_\gamma + 1) B_\gamma}{r^2} \right] \tag{2.19}$$

In this formula, B_γ denotes an arbitrary vector harmonic scalar of $\mathbf{B}$ (both toroidal and poloidal

parts); u_α denotes the same of $\mathbf{u}$. The value N_γ^2 is a normalisation factor associated with the vector spherical harmonics that are orthogonal (see chapter three). The interactions between the flow $\mathbf{u} = \sum_\alpha \mathbf{u}_\alpha$ and the magnetic field $\mathbf{B} = \sum_\beta \mathbf{B}_\beta$ are represented in the term

$$R_m \left(u_\alpha B_\beta B_\gamma \right) = R_m \int \boldsymbol{\nabla} \times \left(\mathbf{u}_\alpha \times \mathbf{B}_\beta \right) \cdot \mathbf{B}_\gamma \, d\Omega \tag{2.20}$$

and are fully evaluated in BG, involving the unknown field scalar functions and integrals between the three spherical harmonics involved. This leads to a large simplification for certain flows since the interaction relation decouples the field into disjoint sets. That is, no field component in one set can interact with the flow to produce any field component in another; therefore each set of field harmonics can be studied independently. BG listed selection rules that govern which interactions are non-zero; they are omitted here for brevity but detailed in chapter three.

In general a disjoint set containing a particular harmonic is an infinite subset of the complete set of harmonics. However, it is assumed (and hoped) that only the large scales are important and so the set is truncated in some way, keeping for example, only those harmonics with $l \leq L$. Both BG and Takeuchi and Shimazu initially truncated their set containing the axial dipole at $L = 2$, keeping only the harmonics $\mathbf{S}_1$, $\mathbf{T}_2$, $\mathbf{T}_2^{2c}$ and $\mathbf{T}_2^{2s}$.

After choosing a flow $\mathbf{u}$, (equation 2.19) defines the set of differential equations that must be solved. The original procedure followed was to seek steady fields, which converts these equations into a generalised eigenvalue problem for R_m; a steady solution was manifested by the existence of a real eigenvalue. Using the boundary conditions pertaining to the self-sustaining problem, the four resulting ordinary differential equations describing the four harmonic scalar functions were solved using the finite difference method, essentially solving for the unknown functions at certain grid points. The velocity chosen for both the Japanese and English studies was defined by

$$t_1(r) = \varepsilon r^2 (1-r) \qquad\qquad s_2^{2c} = r^3 (1-r)^2 \tag{2.21}$$

where ε is an adjustable parameter. No inner core was included in the flow pattern to make it as simple as possible. In both cases real eigenvalues were found for R_m which didn't appear to be affected greatly by changing the harmonic truncation, at least within the limited computer resources available at that time. Thus it had appeared that a working dynamo had been discovered: the combinations of the differential rotation $\mathbf{t}_1$ and the tilted convection roll $\mathbf{s}_2^{2c}$ was such that a (non-axisymmetric) steady field could be maintained with an Earth-like $\mathbf{S}_1$ component. However, 13 years later a repeat of the calculations (*Gibson and Roberts*, 1967) at higher resolution showed that the positive dynamo action found was spurious: the results were merely an artifact

of the inadequate numerical method used and not of the underlying physics. The question of the possibility of dynamo action in a homogeneous conducting sphere was again re-opened.

2.6.3 Non geophysical dynamos

The work of *Backus* (1958) and *Herzenberg* (1958) strengthened the belief that dynamo action was possible by studies of analytical models. The problem with the induction effect is that for any length scale, successively smaller scales will be created by the fluid-field interactions. Numerically it is hoped that on increasing the resolution, the solution converges sufficiently quickly to be captured with the available computing resources. Both of these dynamos use diffusive effects to kill off the small scale fields: the Backus dynamo by periods of stasis and the Herzenberg dynamo by spatial attenuation. Neither of these solutions would be a possible candidate for the Earth; this was evidence however that no generalised anti-dynamo theorem existed.

The study of the dynamo problem in a sphere was also complimented by calculations in other geometries. Analytical solutions of working dynamos in a cylindrical helical flow were published by *Lortz* (1968) and *Ponomarenko* (1973). *Roberts* (1970; 1972a) studied infinite spatially periodic dynamos, and found that almost all motions gave rise to dynamo action for almost all values of R_m. An important outcome of these works in both cylindrical and periodic geometries was that flows that depended only on either one or two spatial coordinates could maintain a magnetic field so long as the field itself depended on all three. This was real evidence that not only was it possible to find flows that worked as dynamos but that it was relatively easy to do so, completely ruling any generalised anti-dynamo theorem. Of course, such geometries were far from geophysical and the lack of spherical boundary conditions turned out to be very important.

2.6.4 Further work on spherical dynamos

The BG choice of flow harmonics was called into question by the analysis of *Braginsky* (1965) and *Tough* (1967) who studied the induction equation in a well defined nearly-axisymmetric limit. They showed that in order to obtain a working dynamo, any non-axisymmetric poloidal flow component would have to contain both sine and cosine functions of azimuth, although strictly this was only valid if the flow was a small perturbation to $\mathbf{t}_1$. This was addressed by *Lilley* (1970) who studied the flow $\mathbf{u} = \mathbf{t}_1 + \mathbf{s}_2^{2c} + \mathbf{s}_2^{2s}$ and found ostensibly converged growing magnetic field solutions. The fact that he had succeeded where BG had failed gave support to the perturbation analysis. Unfortunately, later workers yet again discovered his results to be

unconverged.

The first converged solutions to the kinematic dynamo problem were presented by *Gubbins* (1973), *Pekeris et al.* (1973) and *Roberts* (1972b). The method of solution switched from seeking steady solutions (leading to a generalised eigenvalue problem for R_m) to time-dependent eigenmode solutions. In this paradigm, the assumption that $\mathbf{B} = \hat{\mathbf{B}} e^{\lambda t}$ is made thereby reducing the induction equation (equation 2.12) to an eigenvalue problem for the complex growth rate λ. The adjustable parameter R_m was varied and the lowest value giving $\Re(\lambda) = 0$ sought. If such a value could be found (which was by no means certain) this was termed the critical magnetic Reynolds number (R_m^c). It is entirely possible that $\Im(\lambda) \neq 0$ in this critical state so that the dynamo would be oscillatory. Such behaviour cannot be captured assuming a steady field which might partially explain the convergence problems of previous authors. Gubbins studied the flow $\mathbf{u} = \mathbf{t}_n + \varepsilon \mathbf{s}_n$ where the radial scalar functions were chosen to have a cellular structure whose complexity was governed by the parameter n (also defining the degree of the vector spherical harmonics)

$$s_n(r) = \varepsilon t_n(r) = -r^2 \sin(n\pi r) \tanh\left(n\pi(1-r)\right) \tag{2.22}$$

He found that a greater flow complexity favoured dynamo action, i.e. the critical value of R_m was smaller. The flow component $\mathbf{s}_n$ represents an axisymmetric meridional overturn motion and the $\mathbf{t}_n$ component some kind of differential rotation. Since the flow is axisymmetric the field symmetries separate in azimuthal wave number by the selection rules of BG. The $m = 0$ (axisymmetric) fields must necessarily decay by Cowling's theorem so the growing fields he found were functions of some non-zero wave number m. In fact the solutions were dominated by an $m = 1$ dipole aligned with an axis through the equator (that is, perpendicular to the present-day geomagnetic configuration).

Flows comprising azimuthal wave numbers m of any multiple of 2 (including $m = 0$) are a popular choice for study since the field symmetry class that includes the axial dipole harmonic $\mathbf{S}_1$ also includes harmonics of $m = 2, 4, 6 \cdots$. This means that Cowling's theorem does not apply and geophysically relevant field solutions can be obtained, assuming that the $\mathbf{S}_1$ harmonic dominates the field. It was this fact that partially motivated the choice of BG and also that of *Pekeris et al.* (1973). They chose a flow that was quasi-geophysically motivated by satisfying the equation

$$(\mathbf{u} \cdot \boldsymbol{\nabla})\mathbf{u} = -\boldsymbol{\nabla} p + \mathbf{F} \tag{2.23}$$

which is the steady Navier-Stokes equation with no Lorentz, Coriolis or viscous forces but in-

stead a prescribed forcing $\mathbf{F}$. Solutions to this are given by flows of the form $\boldsymbol{\nabla} \times \mathbf{u} = \tau \mathbf{u}$, where τ describes the sign of helicity $h = \mathbf{u} \cdot \boldsymbol{\nabla} \times \mathbf{u}$ (the chirality of the flow spirals). The form of the defining scalar functions is

$$s_\alpha(r) = \tau r\, j_\alpha(\tau r) \qquad\qquad t_\alpha(r) = \tau s_\alpha(r) \tag{2.24}$$

where j_α is a spherical Bessel function of order α (see *Abramowitz and Stegun*, 1984). Their flow comprised only two components: one poloidal and one toroidal involving just one spherical harmonic. Interestingly, they found converged solutions when taking a Y_2^{2c} dependence in contrast with the perturbation analysis of Tough and Braginsky. They obtained growing eigenmode solutions containing the axial dipole which had energy roughly equally distributed between its toroidal and poloidal parts. Calculations involving a solid inner core showed very similar results to those without, thus mitigating the simplification of a full liquid sphere usually made. They also noted that there was a problem with the somewhat arbitrary choice of flow configuration, so they computed the flow that had a growing magnetic field with minimal Ohmic dissipation (from among those in their study), which might be naturally selected on geophysical grounds. The minimum value was 1.8 GW, well within the geophysical bounds of 1–2 TW (*Roberts et al.*, 2003) and 0.1–0.5 TW (*Buffett*, 2002a).

2.6.5 Work based on the flow of Kumar and Roberts

Kumar and Roberts (1975) (henceforth KR) studied the flow with four components:

$$\mathbf{u} = \mathbf{t}_1 + \varepsilon_1 \mathbf{s}_2 + \varepsilon_2 \mathbf{s}_2^{2c} + \varepsilon_3 \mathbf{s}_2^{2s} \tag{2.25}$$

where the values of $\varepsilon_1, \varepsilon_2$ and ε_3 are adjustable; this flow has been the basis of many studies of kinematic dynamo action to date. Following the previous study of Lilley both $\mathbf{s}_2^{2c}$ and $\mathbf{s}_2^{2c}$ terms representing tilted convection rolls were included. The $\mathbf{t}_1$ component again represented an azimuthal differential rotation and the $\mathbf{s}_2$ a meridional circulation. This choice of flow also precludes Cowling's theorem from applying to the symmetry class containing the axial dipole harmonic since it also includes all even azimuthal wave numbers. They found converged field solutions containing the axial dipole, and followed *Gubbins* (1973) in checking not only convergence of the eigenvalues but of the eigenvectors as well. Again, increasing the spatial complexity favoured more efficient dynamo action. They also found that the direction of the flow components (i.e. the sign of the values of ε_i) was crucial in determining the symmetry of the growing

magnetic field (a result similar to *Roberts* (1972b)). The geophysically scaled fields of dipole symmetry had an Ohmic dissipation of around 1TW which is physically plausible.

In an exhaustive study of the same flow by *Sarson and Gubbins* (1996), it was found that steady magnetic field solutions were favoured with strong meridional circulation. The inclusion of a solid inner core introduced significant changes in the field solutions (in contrast with the study of *Pekeris et al.* (1973)); the fact that different symmetries adapted in different ways led to the possibility that the inner core might play a role in field symmetry selection. Interestingly, little change was observed on altering the inner core from an insulator to a conductor, in contrast to recent simulations that suggest that a conducting inner core might stabilise the dynamo (e.g. *Hollerbach and Jones*, 1993). In an earlier study based on the same flow, *Gubbins and Sarson* (1994) attributed flux concentrations on the CMB and favoured VGP paths to downwelling regions possibly leading to inferences about the actual motion in the outer core from geophysical observations.

To some extent the KR flow resembles the spiralling convection rolls observed in non-magneto convection. *Sarson and Busse* (1998) extended previous studies and found that when the sense of the spiralling and zonal flow was that which is physically preferred, dipole symmetry was favoured. This supported all the earlier studies that indicated that dynamo solutions of this symmetry were easily found and perhaps going some way to explaining the Earth's magnetic field structure in a robust manner.

In the comprehensive parameter study of the KR flow by *Gubbins et al.* (2000a;b) and by *Gubbins and Gibbons* (2002), it was found that (a) there is a critical dependence of kinematic dynamo action on the exact the choice of flow (b) large scale meridional flow promotes large scale axisymmetric poloidal fields and that it decreases the Ohmic dissipation (c) there is some correlation between dynamo action and helicity although this is not conclusive (d) axial dipoles are the most commonly excited field symmetry (e) meridional circulation promotes steady solutions (f) if a flow manages to correlate radial and azimuthal field components then this might lead to oscillatory behaviour and reversals. Thus flows of KR type can explain the geomagnetic principally steady and dipolar field, by requiring strong, large scale meridional circulation.

2.6.6 Simple roll flows

Despite the correlation between increasing spatial complexity and dynamo efficiency, *Dudley and James* (1989) (henceforth DJ) found growing magnetic field solutions in very simple axisymmetric convection roll flows, of $\mathbf{t}_1\,\mathbf{s}_2$, $\mathbf{t}_2\,\mathbf{s}_2$ and $\mathbf{t}_1\,\mathbf{s}_1$ type. Figure 2.3 shows (a) streamlines of an $\mathbf{s}_2$ flow (comprising 2 convective cells: one above and one below the equator) and (b) con-

tours of u_ϕ (the azimuthal flow) supplying the differential rotation, being faster midstream than at the boundaries. No growing field solutions containing the geophysically relevant axial dipole

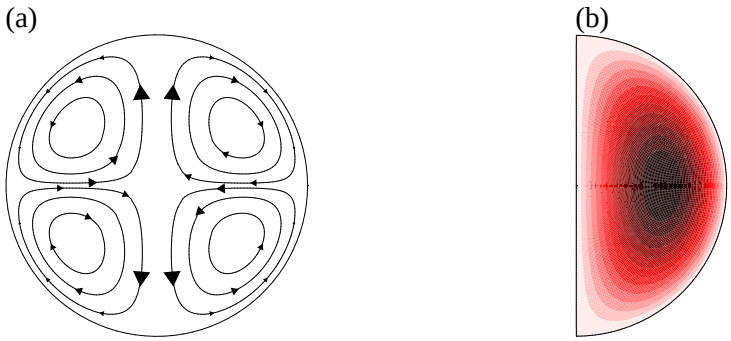


Figure 2.3: (a) Streamlines of an $\mathbf{s}_2$ flow in a meridian plane; (b) contours of u_ϕ in a meridian plane giving rise to differential rotation, the flow being stronger midstream than at the boundaries.

harmonic exist since the fields completely decouple in azimuthal wave number m and axisymmetric fields are precluded. DJ only found solutions in the first two models when the sense of meridional circulation was inwards along the equatorial plane and outwards towards the poles (as in figure 2.3(a)), possibly reducing the expulsion of toroidal field to the boundary and consequent increased Ohmic dissipation. The fact that a one cell convective flow $\mathbf{t}_1\mathbf{s}_1$ can support dynamo action dispelled beliefs that the flow had to be complex in nature, as results from the KR literature suggested. These simple flows were analysed with those of previous authors in the study of *Nakajima and Kono* (1991). They found increased dynamo efficiency with a greater hemispherically averaged helicity (similar to that found by *Gubbins et al.* (2000a)).

2.6.7 The influence of an insulating boundary

Although some flows had been found that show kinematic dynamo action, many more do not. It was not clear why a spherical geometry with an insulating exterior should inhibit growing magnetic fields so much, especially after *Roberts* (1972a) had showed that almost all infinite spatially periodic flows show dynamo action. This question was addressed by *Bullard and Gubbins* (1977) who showed that the insulating boundary condition trapped currents near the edge of the flow region and catastrophically increased Ohmic dissipation. They suggested that this might not be a problem if the flow near the CMB was quiescent, a topic taken up by *Hutchenson and Gubbins* (1994) and *Sarson and Gubbins* (1996). They found the addition of a static outer conducting region on top of the fluid core promoted dynamo action: either the critical magnetic Reynolds number was lowered or growing magnetic fields were found when all had previously decayed. The trapped currents could now flow outside the core region facilitating dynamo ac-

tion, geophysically motivated by the possibility of a stably stratified layer at the top of the outer core. It is not the case however that a monotonic increase in the static layer thickness is always beneficial: in general there is a finite optimal depth (*Kaiser and Tilgner*, 1999).

2.6.8 Motivating the choice of flow

Although one of the great strengths of kinematic theory is the ability to prescribe the flow, the lack of physical self-consistency is one of its great drawbacks. To partially mitigate this issue *Love and Gubbins* (1996b) following *Pekeris et al.* (1973), studied the problem whereby the dynamo efficiency, defined in terms of the critical magnetic Reynolds number and the Ohmic dissipation produced by the growing field, was minimised over a choice of flow (in this case the the values of ε in the KR flow). Such a method of choosing a flow has a strong basis on physical grounds because it might be energetically favoured. They found flows that supported dynamo action with geophysically scaled fields containing the axial dipole, producing a minimal Ohmic dissipation of $70GW$ (well within the allowed bounds). A similar minimisation was carried out by *Holme* (2003) who used B-splines to find the functional form of the scalars defining simple axisymmetric flows, based on the study of DJ. He found that out of the three roll flows proposed by DJ that the critical magnetic Reynolds number for the $\mathbf{t}_2\,\mathbf{s}_2$ flow could be minimised the most.

A partial solution to the full Navier-Stokes equations is another way of choosing the flow. *Pekeris et al.* (1973) had considered the steady state described by the balance of inertia, pressure and a prescribed driving force to derive the form of their velocity. *Sarson* (2003) considered the more geophysically relevant geostrophic approximation, balancing the Coriolis force, buoyancy and pressure as a first approximation to the full equations. He derived a fully consistent flow, and although the driving temperature profile was not geophysically plausible, it is certainly a step closer to attaining more Earth-like flows. He found growing field solutions of dipole symmetry in line with the studies of the KR flow.

2.6.9 Non-linear effects

In studying the linear induction equation in isolation we should be aware of as many of the non-linear effects as possible that might be of relevance in the full problem. Both *Brummell et al.* (1998) and *Fuchs et al.* (1999) studied the Navier-Stokes equation coupled with the induction equation. They neglected the Coriolis force but instead supplied a forcing, chosen such that a known kinematic dynamo flow was a steady solution when $\mathbf{B} = \mathbf{0}$.

There is no guarantee of stability of this magnetically favourable steady state, and with large forcing or a non-zero magnetic field the system may well move away from the desired flow.

The results may be summarised as follows:

1. An initial magnetic field was amplified as expected, until the Lorentz force became important. This non-linear feedback perturbed the flow and the system evolved towards another steady solution of the Navier-Stokes equations. The final state was not a kinematic dynamo and the magnetic field decayed.

2. Starting with a steady state that was not a kinematic dynamo, if the initial magnetic field was large enough, the associated Lorentz force perturbed the system into a flow regime that supported growing fields and the magnetic field grew.

Thus the magnetic field was observed to either kill or create a kinematic dynamo. This effect relies on multiple steady states of the Navier-Stokes equations and serves as a warning that a growing magnetic field solution of a kinematic dynamo is only valid until the Lorentz force becomes large.

2.6.10 Summary

In the above review, no attempt has been made to discuss the advances of mean field theory, a vast subject in its own right describing particularly the effects of small scale turbulence, since this thesis is only concerned with large scale flows. In any event, one might expect that since the Earth's field is principally large scale [2] that the underlying flow causing its maintenance is also large scale, especially considering the size of the magnetic Prandtl number $P_m = \frac{\nu}{\eta} \approx 10^{-6}$ in the Earth (the diffusive scale of the magnetic field is much smaller than that of the fluid).

It is possible in kinematic calculations given a choice of flow, only to either

- find a converged growing magnetic field solution and so label the flow as a kinematic dynamo.

- concede that with the computing resources available, no growing converged solution can be found.

Thus it is impossible to rule out dynamo action with finite computing resources; however, we can usually explore the geophysically relevant range of $R_m = 0$–1000 with sufficient resolution to achieve convergence.

Kinematic dynamo theory has had a lot of success, in particular showing:

[2] Models of the field at the core-mantle boundary inverted from observations can only be resolved up to degree $l = 14$ due to the effects of crustal noise, so the results are necessarily large scale. Nonetheless, decreasing power spectra (see e.g. *Langel*, 1987) suggest that this is the case.

- Self-exciting dynamo action in a conducting homogeneous sphere is possible.
- Earth-like field structures can be found with partially geophysically motivated flows.
- Some links have been made between the generated fields and various flow properties, e.g. strength of meridional circulation, locations of downwelling and helicity.
- A thin quiescent conducting layer on the top of the convecting region in the outer core would favour dynamo action.
- Reversal mechanisms have been proposed (e.g. *Sarson and Jones*, 1999; *Rikitake and Hagiwara*, 1968) typically appealing to a fall in meridional circulation leading to either oscillatory behaviour or field decay. In the first case, the field reversal is automatic and subsequent restarting of the motion perhaps would grow the field energy back to its pre-transitional value. In the latter case the field decays to zero; a recovery in the motion leads to kinematic dynamo action that grows the magnetic field but in the opposite polarity.

The main problems with kinematic dynamos are two-fold.

1. Assuming that the flow can be motivated physically there is still the issue of the neglected Lorentz force. For small fields this is justified but if growth occurs this assumption becomes increasingly worse. Thus kinematic dynamos can only describe magnetic field growth in a self-consistent manner until the Lorentz force becomes large. It may well be the case that this cutoff is reached before the fastest growing eigenmode can manifest itself, so that the eigenvectors computed may not be a good description of the geophysical field. The effect of the Lorentz force on the system is not well understood; however, it may well suppress differential rotation leading to enhanced heat transport efficiency across the outer core in the full non-linear system (*Busse*, 2002).

2. Whether growing eigenmodes exist or not is critically dependent on the choice of flow. For example a typical KR type flow that supports growing magnetic fields can be shown to comprise 99% of $\mathbf{t}_1$ by root mean squared (rms) average over V, but $\mathbf{t}_1$ is excluded from supporting dynamo action. Thus two flows, only one of which is a kinematic dynamo, differ by just 1%. Sensitivity in the exact choice of flow is also manifested in the field symmetry selection (*Gubbins et al.*, 2000b) where a tiny perturbation in the flow may lead to a different preferred symmetry. This sensitivity can be understood by the fact that the eigenmodes must necessarily grow everywhere at the same rate (by the assumption $\mathbf{B} = \hat{\mathbf{B}}e^{\lambda t}$). Poloidal and toroidal shear act together to produce a growing field but their

effect must be the the same at all points and not just on average. This means that for example, in the case of the KR flow, that the balance between the chosen values of ε is very delicate and so only a small subset of the possible choices allow growing fields.

The concern here is that on changing the flow slightly, either by convective processes or by the Lorentz force, that the dynamo action would disappear. That many planets and stars have working dynamos does not support this; indeed the Earth has maintained its magnetic field for around 3Ga at an approximately constant magnitude and has showed no sign of failing.

Because of the sensitivity issue there is no robust method of explaining how a kinematic dynamo operates, for a small change may not effect the physical interpretation but it might switch the dynamo off. There is therefore a need for a robust way of characterising magnetic field growth and the associated physical processes. This is one of the aims of this thesis.

2.7 Dynamo experiments

2.7.1 Introduction

No introduction to kinematic dynamo theory would be complete without some comments on dynamo experiments. These can be seen as a testing mechanism for not only the ideas behind self-sustaining mechanisms but for the quantitative results from numerical calculations. The idea is to reproduce in a laboratory, as much as possible, the flow patterns already known to exhibit kinematic dynamo action to see if they really can be made to generate magnetic field.

Physical realisations of dynamos have many attractive characteristics over numerical models, for not only are the physical fluid properties automatically correct, but there can be no spurious results due to numerical inaccuracies, lack of proper boundary conditions or otherwise.

The experimental setup will be usually some variant of the following: liquid sodium is encased in a metal container of some geometry and made to flow in a certain manner by external forcing, usually propellers. A seed magnetic field is injected into the system by external coils and its subsequent time evolution is tracked by Hall probes. From kinematic theory it is known that in order to get dynamo action, the value of R_m must be sufficiently large, of $O(10)$. This is just about obtainable in experiments, recall that $R_m = \mathscr{U}\mathscr{L}/\eta$ and sodium has the lowest magnetic diffusivity of any liquid metal of $\eta = 0.092\ m^2 s^{-1}$ (at $150°C$, see *Nataf*, 2003). For a typical laboratory length scale $\mathscr{L}$ of $30cm$ this means that we need flows of around $\mathscr{U} = 3ms^{-1}$

to give $R_m = 10$. This is possible, although to get much higher values of R_m is clearly difficult, especially when considering the safety factors involved.

In contrast to the current parameter regimes addressed by current numerical simulations, experiments can achieve Ekman numbers of 10^{-8} (in the rotating case, see *Lathrop et al.*, 2001) and magnetic Prandtl numbers of around 10^{-6}, very similar to that predicted for the Earth's core. Unfortunately such a low value of P_m means that the kinematic Reynolds number $R_e = R_m P_m^{-1}$ is O(10^7) indicating strongly turbulent flow. This creates problems in visualising the flow, for significant time-dependent departures from the large scale forced flow will occur. This makes interpreting the results problematic since only the kinematic dynamo action of stationary spherical flows are at all well understood. The power required to drive a given non-magnetic turbulent flow scales as R_m^3, all of which is converted to heat by viscous processes. The main issue though is not putting the energy in, but rather extracting it sufficiently fast, for the conductivity of liquid sodium decreases (and hence so does R_m) as the temperature increases. For a good recent review of experimental work see *Nataf* (2003).

2.7.2 Two working dynamos

In 2000 two independent groups in Riga and Karlsruhe, working on completely different experimental dynamos, announced their success. The Riga group had worked on a setup based on the *Ponomarenko* (1973) helical screw dynamo, consisting essentially of a long cylindrical tube filled with liquid sodium driven at one end by a propeller. They reported a growing magnetic field that could be sustained in a saturation phase (where the Lorentz force was important) for several minutes (*Gailitis et al.*, 2000; 2001). In contrast the Karlsruhe group built a finite approximation to the infinite spatially periodic flow of *Roberts* (1972a) consisting of 52 juxtaposed vertical cylindrical cells. Liquid sodium was constrained to flow in a helical manner by blades in each cell and generated a magnetic field that was predominately horizontal (*Stieglitz and Müller*, 2001). Both of these experiments confirmed that dynamo action was physically possible and could be adequately described by kinematic theory.

2.7.3 Spherical dynamos

Although the Earth's dynamo is powered by convection, achieving this in laboratory dynamos is difficult simply because the likely velocities (and hence R_m) will be too small. For this reason investigators such as those in the Maryland (USA) group drive their spherical flow using propellers to achieve $R_m = O(10)$. In this case, the flow pattern they try to mimic is that of the $\mathbf{t}_2\,\mathbf{s}_2$ flow of DJ. The setup comprises two counter-rotating propellers mounted near the top and

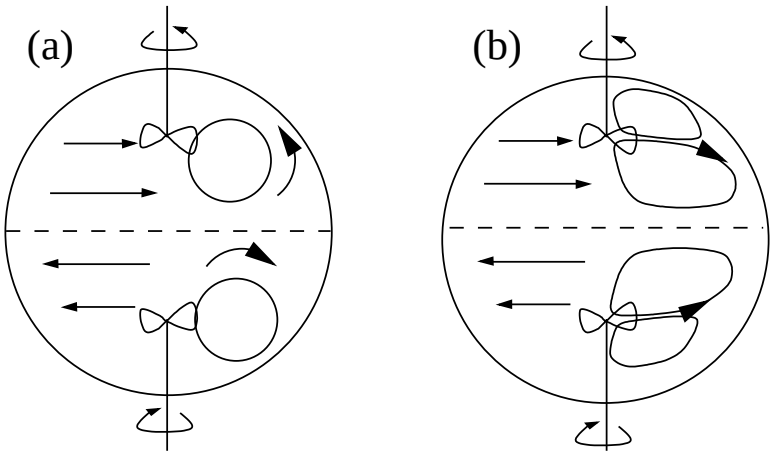


Figure 2.4: (a) Schematic picture of the $\mathbf{t}_2\,\mathbf{s}_2$ flow of *Dudley and James* (1989) on which one of the Maryland experiments is motivated (b) A more realistic situation: the propellers tend to fling fluid radially outwards, generating more of a $\mathbf{t}_2\,\mathbf{s}_2\,\mathbf{s}_4$ pattern.

bottom of the inside of the sphere on the axis of symmetry that not only push fluid up away from the 'poles' and outwards on the 'equator', but generate an equatorially antisymmetric toroidal flow. In fact, the propellers tend to fling fluid radially outward creating a flow more of a $\mathbf{t}_2\,\mathbf{s}_2\,\mathbf{s}_4$ configuration (see figure 2.4), which will no doubt alter the results of the kinematic theory. In addition the flow is highly turbulent and it is unknown whether such effects are important: it is unlikely that the steady kinematic dynamo solutions predicted by DJ will be manifested due to the strongly time-dependent flow.

In order to investigate the tendency towards dynamo action, the decay time of the initially injected field is monitored. If a generating mechanism is possible, this decay time should increase as a function of R_m towards infinity at the critical point. So far no dynamo action has been found; however, surprisingly there has been a trend of the field containing the $m = 0$ symmetries to increase towards self-generation in the manner described, more so than that containing the $m = 1$ symmetries (*Peffley et al.*, 2000a). Of course, since the flow is non-axisymmetric the field does not decouple in azimuth so Cowling's theorem does not rule out this effect; indeed, this is the kind of behaviour seen in kinematic models of KR type. The measurements of magnetic decay rates by Hall probes mounted on the sphere's exterior show a great dependence on spatial position (*Peffley et al.*, 2000b). In addition, in the current sub-critical regime, intermittent bursts of magnetic field energy due to the strong turbulence are observed (*Sweet et al.*, 2001). Such effects are not characteristic of theoretical kinematic solutions: if one eigenmode is obtained then its growth rate should be everywhere the same. *Tilgner* (2002) argues however, that multiple eigenmodes could be excited thus creating the spatial inhomogeneities

measured. It is not clear whether dynamo action will be physically realised in flows of this kind since the departures to the simple theory on which they are based are significant (recall that even a small perturbation in the flow can switch off dynamo action).

The interested reader should consult *Nataf* (2003) for a wider and more detailed overview.

Chapter 3

Theoretical considerations

3.1 The defining equations and geometry

In this thesis we will be interested in modelling the behaviour of a magnetic field **B** under the influence of a sphere V of conducting fluid (of conductivity σ and magnetic diffusivity $\eta = (\mu_0 \sigma)^{-1}$), that moves at prescribed flow velocity **u**. Because of the geometry, we use spherical polar coordinates (r, θ, ϕ) describing V by $r \leq R$ where the flow is non-zero, and its exterior $\hat{V}$ by $r > R$ where the medium is quiescent. The infinite region $\hat{V}$ will be taken to be an electrical insulator. We denote the boundary of a region by the ∂ symbol; for example, ∂V is equivalent to $r = 1$ although we must bear in mind that this surface is viewed from V i.e. from the inside.

The flow will be taken to be incompressible ($\boldsymbol{\nabla} \cdot \mathbf{u} = 0$) and non-slip ($\mathbf{u} = \mathbf{0}$ on ∂V). We denote the position vector by $\mathbf{r} = r\hat{\mathbf{r}}$ where $\hat{\mathbf{r}}$ is a unit vector.

3.1.1 The induction equation

The non-dimensional induction equation (equation 3.1) describes the effect of the moving conductor and diffusion on a magnetic field:

$$\frac{\partial \mathbf{B}}{\partial t} = R_m \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}) - \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) \tag{3.1}$$

If the diffusivity is constant and finite then this may be written

$$\frac{\partial \mathbf{B}}{\partial t} = R_m \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}) + \nabla^2 \mathbf{B} \tag{3.2}$$

The non-dimensional parameter above is the magnetic Reynolds number, $R_m = \mathscr{U}\mathscr{L}/\eta_0$, where the non-dimensional velocity, length and diffusivity are scaled relative to typical values of $\mathscr{U}$, $\mathscr{L}$ and η_0 (the diffusivity of V). We will take $\mathscr{L}$ to be the radius R of the sphere so that the non-dimensional radius of V is 1. The non-dimensional magnetic diffusivity η is unity in V and infinite in $\hat{V}$ (since $\sigma = 0$ there). Implicit in equations (3.1) and (3.2) is the non-dimensional time, scaled relative to the magnetic diffusion timescale $\mathscr{L}^2/\eta_0$. The choice of the typical velocity scale is somewhat open to interpretation since it could take on any number of forms: the maximum pointwise flow speed, the maximum rate of strain and the root mean squared (rms) average are all possible contenders. In this study we choose the latter (rms) scale. Consultation of equations (3.1) and (3.2) reveals that this is equivalent to normalising the flow to have unit rms, and then multiplying by the relevant value of R_m. Such a procedure is important since it allows us to compare the kinematic dynamo action of different flows meaningfully, especially when different sets of authors have different ways of defining their velocities.

Lastly we note that the pre-Maxwell equations are invariant under any rotation and it follows that this property is carried forward to the induction equation as well. Therefore, given any velocity field, we may add or subtract any solid body rotation to it without altering its dynamo effect; in particular, if V spins as a solid then the field will behave in the same manner as if V were quiescent: it decays.

3.2 Vector harmonics

3.2.1 Poloidal and toroidal form

Any divergence-free vector field, for example the magnetic field $\mathbf{B}$, can be expressed in poloidal $(\mathbf{S})$ and toroidal $(\mathbf{T})$ decomposition where $\mathbf{B} = \mathbf{S} + \mathbf{T}$ and

$$\mathbf{S} = \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times [S(\mathbf{r})\,\hat{\mathbf{r}}] \tag{3.3a}$$

$$\mathbf{T} = \boldsymbol{\nabla} \times [T(\mathbf{r})\,\hat{\mathbf{r}}] \tag{3.3b}$$

The literature is divided over whether the above expressions should be defined instead with $\mathbf{r}$ and not $\hat{\mathbf{r}}$; on comparing results with other authors care should be taken to determine which definition they have used.

The scalar fields S and T uniquely define $\mathbf{B}$ if on the surface of any origin-concentric spherical surface they have zero mean (see *Backus et al.*, 1996). Monopolar fields have no angular dependence, and if included these averages will be non-zero. Consequently this uniqueness

condition amounts to precisely the exclusion of monopoles that we demand anyway from the pre-Maxwell equations. We expand the scalar functions in spherical harmonics $Y_l^m(\theta,\phi)$ which form an orthogonal complete set over a spherical surface, each having zero mean if $l \geq 1$. They are defined as

$$Y_l^{ms/c}(\theta,\phi) = P_l^m(\cos\theta) \left\{ \begin{array}{c} \sin m\phi \\ \cos m\phi \end{array} \right\} \tag{3.4}$$

where P_l^m are the associated Legendre functions of degree l and order m (see Appendix B.1 or *Macrobert*, 1967). It is usual to take $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$ to define a spherical surface rather than any other description. For example, $0 \leq \theta \leq 2\pi$ and $0 \leq \phi \leq \pi$ is a viable alternative, although there is some ambiguity in using functions of $\cos\theta$, it having no inverse in this range.

Instead of tediously and sometimes confusingly defining each spherical harmonic by its l and m values and either $\sin m\phi$ or $\cos m\phi$ dependence, we shall sometimes use Greek subscript notation, for example:

$$S(r,\theta,\phi) = \sum_\alpha Y_\alpha(\theta,\phi)\, S_\alpha(r) \tag{3.5}$$

where each value of the summation variable α represents a different harmonic. The scalar functions $S_\alpha(r)$ describe the remaining radial dependence.

We may therefore express the magnetic field in terms of 'vector spherical harmonics', implicitly excluding the $l = 0$ monopolar terms:

$$\mathbf{B} = \sum_\alpha \mathbf{S}_\alpha + \mathbf{T}_\alpha = \sum_\alpha \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times [Y_\alpha(\theta,\phi)\, S_\alpha(r)\, \hat{\mathbf{r}}] + \boldsymbol{\nabla} \times [Y_\alpha(\theta,\phi)\, T_\alpha(r)\, \hat{\mathbf{r}}] \tag{3.6}$$

where the radial dependence is either given or assumed unknown and must be solved for. Sometimes we shall need to be explicit and write out the harmonic dependence in full. On such occasions we shall not write the unnecessary superscript of 0 for axisymmetric harmonics, thus $\mathbf{S}_2^0 = \mathbf{S}_2$ for example.

The components of the vectors $\mathbf{T}_\alpha$ and $\mathbf{S}_\alpha$ are

$$\begin{aligned} T_r &= 0 \\ T_\theta &= \frac{T_\alpha(r)}{r\sin\theta}\frac{\partial Y_\alpha}{\partial \phi} \\ T_\phi &= -\frac{T_\alpha(r)}{r}\frac{\partial Y_\alpha}{\partial \theta} \end{aligned} \tag{3.7}$$

$$\begin{aligned} S_r &= \frac{l_\alpha(l_\alpha+1)}{r^2}S_\alpha(r)Y_\alpha \\ S_\theta &= \frac{1}{r}\frac{dS_\alpha}{dr}\frac{\partial Y_\alpha}{\partial \theta} \\ S_\phi &= \frac{1}{r\sin\theta}\frac{dS_\alpha}{dr}\frac{\partial Y_\alpha}{\partial \phi} \end{aligned} \tag{3.8}$$

Toroidal vectors have no radial component so they are confined to spherical surfaces of constant radius; poloidal vectors have all three vector components. The curl of a toroidal vector is trivially poloidal although the converse is also true:

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times[S(\mathbf{r})\,\hat{\mathbf{r}}] = -\nabla^2\boldsymbol{\nabla}\times[S(\mathbf{r})\,\hat{\mathbf{r}}] \tag{3.9}$$

using the identity $\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{A} = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A}) - \nabla^2\mathbf{A}$ for any vector $\mathbf{A}$. The Laplacian operator can be taken inside the curl, leaving the vector in toroidal form, although inflicting some derivatives on the defining scalar function (see section 3.4.1 for details).

3.2.2 Orthogonality and normalisation

Spherical harmonics are orthogonal over any origin-concentric spherical surface S so that

$$\int_0^{2\pi}\int_0^{\pi} Y_\alpha(\theta,\phi)\,Y_\beta(\theta,\phi)\sin\theta\,d\theta\,d\phi = \oint_S Y_\alpha\,Y_\beta\,d\Omega = N_\alpha^2\delta_{\alpha\beta} \tag{3.10}$$

where $d\Omega = \sin\theta\,d\theta\,d\phi$ is the element of solid angle and the symbol $\delta_{\alpha\beta}$ is 1 only if $\alpha=\beta$ and zero otherwise. The constant N_α^2 depends on the specific normalisation chosen for the harmonics. Kinematic dynamo theory is plagued with differing normalisations used between authors, for example, *Bullard and Gellman* (1954) used unnormalised harmonics and *Sarson* (2003) used Ferrer normalisation. In this study, we shall always use Schmidt quasi-normalised harmonics, defined by

$$N_\alpha^2 = \frac{4\pi}{(2l_\alpha+1)} \tag{3.11}$$

Toroidal and poloidal vector harmonics are mutually orthogonal over S and are orthogonal with other vectors of the same type but of different spherical harmonic dependence:

$$\oint_S \mathbf{T}_\alpha \cdot \mathbf{S}_\beta \, d\Omega = 0 \tag{3.12}$$

$$\oint_S \mathbf{T}_\alpha \cdot \mathbf{T}_\beta \, d\Omega = \delta_{\alpha\beta} l_\alpha (l_\alpha + 1) N_\alpha^2 \frac{T_\alpha(r) T_\beta(r)}{r^2} \tag{3.13}$$

$$\oint_S \mathbf{S}_\alpha \cdot \mathbf{S}_\beta \, d\Omega = \delta_{\alpha\beta} N_\alpha^2 \frac{l_\alpha(l_\alpha+1)}{r^2} \left(\frac{l_\alpha(l_\alpha+1) S_\alpha(r) S_\beta(r)}{r^2} + \frac{dS_\alpha}{dr}\frac{dS_\beta}{dr} \right) \tag{3.14}$$

A closely related formula is that for the rms average of $\mathbf{B}$ within the sphere V. It may be written:

$$\frac{4\pi}{3} \mathbf{B}_{rms}^2 = \sum_\alpha l_\alpha (l_\alpha + 1) N_\alpha^2 \int_0^1 \left[\frac{l_\alpha(l_\alpha+1)}{r^2} S_\alpha^2 + \left(\frac{dS_\alpha}{dr} \right)^2 + T_\alpha^2 \right] dr \tag{3.15}$$

In deriving such relations use has been made of the following identities

$$\oint_S \left(\frac{\partial Y_\alpha}{\partial \theta}\frac{\partial Y_\beta}{\partial \phi} - \frac{\partial Y_\alpha}{\partial \phi}\frac{\partial Y_\beta}{\partial \theta} \right) d\theta d\phi = 0 \tag{3.16}$$

$$\oint_S \left(\frac{\partial Y_\alpha}{\partial \theta}\frac{\partial Y_\beta}{\partial \theta} + \frac{1}{\sin^2\theta}\frac{\partial Y_\alpha}{\partial \phi}\frac{\partial Y_\beta}{\partial \phi} \right) d\Omega = N_\alpha^2 \delta_{\alpha\beta} l_\alpha (l_\alpha + 1) \tag{3.17}$$

For derivations of these and further details on spherical harmonics see Appendix B.1.

3.3 Boundary conditions

3.3.1 Continuity conditions

The magnetic field $\mathbf{B}$, electric field $\mathbf{E}$ and current density $\mathbf{J}$ must satisfy the following continuity conditions across any surface with or without a conductivity discontinuity

$$[\mathbf{B}] = \mathbf{0} \tag{3.18}$$

$$[\hat{\mathbf{n}} \times \mathbf{E}] = \mathbf{0} \tag{3.19}$$

$$[\hat{\mathbf{n}} \cdot \mathbf{J}] = 0 \tag{3.20}$$

where [] denotes the jump and $\hat{\mathbf{n}}$ the normal to the surface in question. All other components may be discontinuous (*Gubbins and Roberts*, 1987).

The only discontinuity of σ we shall encounter is that at $r = 1$, between $\sigma = 0$ in $\hat{V}$ and σ being finite and non-zero in V. If the flow is non-slip there so that $\mathbf{u} = \mathbf{0}$, then from

the pre-Maxwell equations, $\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J} = \mu_0 \sigma \mathbf{E}$ and so equation (3.19) is equivalent to the continuity of the tangential components of $\mathbf{E} = \eta \boldsymbol{\nabla} \times \mathbf{B}$. Equation (3.20) supplies no additional information. In terms of the toroidal and poloidal defining scalars, these amount to

$$[S_\alpha] = 0 \qquad [S_\alpha^{'}] = 0 \qquad [T_\alpha] = 0 \tag{3.21a}$$

$$[\eta T_\alpha^{'}] = 0 \qquad [\eta \nabla_\alpha^2 S_\alpha] = 0 \tag{3.21b}$$

where the notation ∇_α^2 is defined in section 3.4.1 and $^{'}$ is the radial derivative. Note that if a quantity is continuous over the surface $r = 1$ then its spherical harmonic components must be also, hence it is justified to generalise to the above individual harmonic components. Equation (3.21a) is derived from the continuity of $\mathbf{B}$ and equation (3.21b) from consideration of the tangential components of $\eta \boldsymbol{\nabla} \times \mathbf{B}$. If large tangential current sheets occur at the discontinuity due either to boundary layer effects or a very large conductivity, then equations (3.21) must be reduced to merely the continuity of B_r, or equivalently S_α and the conditions of (3.21b). In the case where $\hat{V}$ is modelled as an electrical insulator, these reduce further since equation (3.21b) becomes meaningless (η being infinite in $\hat{V}$). We can now only guarantee the continuity of B_r, which is therefore the only component we may infer on the inside of the CMB from inverse models of observational data.

3.3.2 The magnetic field in an electrical insulator

In an electrical insulator, the conductivity σ vanishes and so $\boldsymbol{\nabla} \times \mathbf{B} = \mathbf{0}$ from the pre-Maxwell equations. Therefore $\mathbf{B}$ can be written as a potential

$$\mathbf{B} = -\boldsymbol{\nabla} \Phi \tag{3.22}$$

for some function Φ. Since $\mathbf{B}$ is solenoidal then

$$\nabla^2 \Phi = 0 \tag{3.23}$$

that can be solved to give

$$\Phi = \sum_{\alpha : l_\alpha \geq 1} G_\alpha r^{l_\alpha} Y_\alpha + H_\alpha r^{-(l_\alpha + 1)} Y_\alpha \tag{3.24}$$

identically to that presented in Appendix A. In the case we will consider, the electrical insulator is $\hat{V}$, being the region $r \geq 1$. Excluding external sources of field amounts to the condition that the

field must vanish as $r \to \infty$ so that $G_\alpha = 0$. The remaining solution then decays with increasing r, as consistent with an internally generated field. In particular, the potential $\Phi = O(r^{-2})$ and $\mathbf{B} = O(r^{-3})$ as $r \to \infty$. From the pre-Maxwell equations it follows also that $\mathbf{E} = O(r^{-2})$.

3.3.3 Electrically insulating boundary conditions

The solution for the magnetic field in the external insulator is of poloidal form:

$$\mathbf{B} = -\nabla \sum_{\alpha: l_\alpha \geq 1} H_\alpha r^{-(l_\alpha+1)} Y_\alpha = \sum_{\alpha: l_\alpha \geq 1} \nabla \times \nabla \times \left[\frac{H_\alpha}{l_\alpha} r^{-l_\alpha} Y_\alpha \hat{\mathbf{r}} \right] \tag{3.25}$$

By orthogonality there can be no toroidal field in $\hat{V}$ and by continuity T_α must vanish on the inside edge of $r = 1$. To derive a condition on the poloidal field, we use the continuity of the scalar function $S_\alpha(r)$ and its first derivative across $r = 1$. By eliminating the unknown scalar H_α we arrive at the following matching conditions:

$$\frac{dS_\alpha}{dr}(1) + l_\alpha S_\alpha(1) = 0 \qquad T_\alpha(1) = 0 \tag{3.26}$$

We can therefore extend any poloidal field $\mathbf{S}_\alpha$ into $\hat{V}$ to get:

$$\mathbf{B} = -l_\alpha S_\alpha(1) \nabla \left(\frac{Y_\alpha}{r^{l+1}} \right) \tag{3.27}$$

Note that the above analysis is made possible by applying continuity of the spherical harmonic components of the poloidal scalar and its derivative. A local finite element method for example, has no knowledge of the global harmonic coefficients so matching the solution to an external electrical insulator is far more problematic.

3.3.4 Regularity at the origin

The origin is a singular point of the spherical polar coordinate system and we must ensure that the quantities we use are regular there and infinitely differentiable (where appropriate). In order to do this, we need to be able to write the scalar functions and Cartesian components of the vectors as multinomials in (x, y, z). In the case of a field in poloidal and toroidal form, this is equivalent to using the $\mathbf{r}$ notation ($\mathbf{r}$ being infinitely differentiable) in the original definitions (equation 3.3), rather than $\hat{\mathbf{r}}$, and insisting that the poloidal and toroidal scalar functions themselves be multinomials of (x, y, z) everywhere.

To apply this to a spherical geometry, assume first that such a process has been carried

out for the vector or scalar function in question. We investigate what conditions it places on the behaviour of the defining harmonic scalar functions at the origin.

We first expand the scalar functions in the (complete) homogeneous multinomials in (x,y,z), f_n, of degree n (see *Backus et al.*, 1996, pg. 47). We can write each f_n as

$$f_n(x,y,z) = r^n \sum_{i=0}^{[n/2]} Y_{n-2i}(\theta,\phi) \tag{3.28}$$

where Y_l is a linear combination of spherical harmonics of degree l (see *Macrobert*, 1967) and here [ ] means the integer part.

The power of r multiplying Y_l is

- at least l
- differs from all other powers by an even number.

Hence the behaviour near $r = 0$ of the coefficient of Y_l is

$$r^l \left(a_0 + a_1 r^2 + a_2 r^4 + \cdots\right) \tag{3.29}$$

We might expect the expansion to be of this form since odd powers of r introduces singular behaviour at the origin. Reverting to the $\hat{\mathbf{r}}$ form of the toroidal-poloidal vectors means that we have the following necessary condition for regularity at the origin of any vector harmonic scalar function multiplying $Y_l^m(\theta,\phi)$:

$$r^{l+1}(a_0 + a_1 r^2 + a_2 r^4 + \cdots) \tag{3.30}$$

This condition is also sufficient, since $r^l Y_l^m(\theta,\phi)$ is a homogeneous multinomial in (x,y,z) of degree l and so is regular; any extra powers of r^2 are also analytic.

Infinite differentiability is not always required; for example, in the induction equation only the first derivative of the velocity is needed (which involves the second and first derivatives of the poloidal and toroidal scalars respectively). So long as the vectors in question have as many continuous derivatives as is necessary, no penalty should be placed on lack of infinite differentiability.

An additional point is that if a finite difference scheme is used to solve the equations, the relevant quantities are sampled only at discrete locations (the same also being true in general of a spectral scheme using quadrature). It could then be conceived that even if the poloidal and toroidal scalars did not conform to the behaviour given above for infinite differentiability but

nonetheless vanished at the origin, that they might conform on a scale smaller than that of the grid used. Therefore no discrete numerical procedure will be able to distinguish whether or not the above limiting form is obeyed.

Such an argument may be used to motivate geophysical flows that do not satisfy the criteria above.

3.4 The diffusion operator

3.4.1 Properties of the Laplacian

The Laplacian operator ∇^2 commutes with the curl vector operator $\boldsymbol{\nabla}\times$ in a Cartesian system and hence it must also do so in spherical polar coordinates. Additionally, ∇^2 annihilates the position vector $\mathbf{r}$ (but not $\hat{\mathbf{r}}$) so that we can write

$$\nabla^2\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times[S(\mathbf{r})\hat{\mathbf{r}}] = \boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\nabla^2[S(\mathbf{r})\hat{\mathbf{r}}] \tag{3.31}$$

$$= \boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\nabla^2[\frac{S(\mathbf{r})}{r}\mathbf{r}] \tag{3.32}$$

$$= \boldsymbol{\nabla}\times\boldsymbol{\nabla}\times[r\nabla^2\left(\frac{S(\mathbf{r})}{r}\right)\hat{\mathbf{r}}] \tag{3.33}$$

and similarly in the toroidal case. We may write, in spherical polar coordinates

$$\nabla^2 = \frac{1}{r}\frac{\partial^2}{\partial r^2}r - \frac{L^2}{r^2} \tag{3.34}$$

where L^2, the angular momentum operator, is defined as

$$L^2 = -\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right\} \tag{3.35}$$

The spherical harmonics Y_l^m are eigenfunctions of L^2 with eigenvalues $l(l+1)$. This means that

$$\nabla^2\left[f_l^m(r)Y_l^m(\theta,\phi)\right] = Y_l^m(\theta,\phi)\left[D_l^2 f_l^m(r)\right] \tag{3.36}$$

where we define the operator

$$D_l^2 = \frac{1}{r}\frac{d^2}{dr^2}r - \frac{l(l+1)}{r^2} = \frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2} \tag{3.37}$$

Hence from equations (3.33) and (3.36) the Laplacian of a poloidal (toroidal) vector harmonic is also poloidal (toroidal) and of the same harmonic dependence but defined by a radial scalar

function given by

$$r\left[\frac{1}{r}\frac{d^2}{dr^2}r-\frac{l(l+1)}{r^2}\right]\left(\frac{S_\alpha}{r}\right)=\left[\frac{d^2}{dr^2}-\frac{l(l+1)}{r^2}\right]S_\alpha \tag{3.38}$$

We define the operator associated with the Laplacian acting on toroidal and poloidal scalars as ∇^2_α or ∇^2_l (depending on the context) which we shall define as

$$\nabla^2_\alpha=\frac{d^2}{dr^2}-\frac{l_\alpha(l_\alpha+1)}{r^2} \tag{3.39}$$

or

$$\nabla^2_l=\frac{d^2}{dr^2}-\frac{l(l+1)}{r^2} \tag{3.40}$$

It should be noted that the Laplacian operator itself depends only on r and l (hence the subscript l used in the notation); it is independent of the azimuthal wave number m. Explicitly then:

$$\nabla^2\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times[Y_\alpha S_\alpha\hat{\mathbf{r}}]=\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\left[Y_\alpha\nabla^2_\alpha S_\alpha\hat{\mathbf{r}}\right] \tag{3.41}$$

$$\nabla^2\boldsymbol{\nabla}\times[Y_\alpha T_\alpha\hat{\mathbf{r}}]=\boldsymbol{\nabla}\times\left[Y_\alpha\nabla^2_\alpha T_\alpha\hat{\mathbf{r}}\right] \tag{3.42}$$

3.4.2 The diffusion problem

In benchmarking numerical code it is often invaluable to be able to make comparisons with known analytic results. In our case we can solve the diffusion problem $(R_m=0)$ exactly which will be used as a guide for numerical accuracy. The relevant equation is:

$$\frac{\partial\mathbf{B}}{\partial t}=\nabla^2\mathbf{B} \tag{3.43}$$

and we typically seek an exponential time-dependence of the form $\mathbf{B}(t)=\hat{\mathbf{B}}e^{-d^2t}$ that reduces the diffusion equation to a vector eigenvalue problem for $-d^2$. As has already been shown the toroidal and poloidal vector harmonics decouple under the Laplacian operator so the resulting equations are of a simple form:

$$-d^2_P S_l(r)=\nabla^2_l S_l(r) \tag{3.44}$$

$$-d^2_T T_l(r)=\nabla^2_l T_l(r) \tag{3.45}$$

where we have written $-d^2_P$ and $-d^2_T$ for the poloidal and toroidal decay rates respectively. Because the Laplacian is independent of m, the solutions for the defining scalar functions are

dependent only on l and r. The poloidal solutions are (and similarly in the toroidal case)

$$G\,r\,n_l(d_P r) + H\,r\,j_l(d_P r) \tag{3.46}$$

where j_l and n_l are spherical Bessel functions of order l and of the first and second kinds (see *Abramowitz and Stegun,* 1984); G and H are arbitrary constants. We discard n_l which is singular at the origin and are left to fit the relevant boundary conditions to determine the decay rates. Applying the electrically insulating boundary conditions at $r = 1$ we obtain

$$d_P\,j_l'(d_P) + (l+1)\,j_l(d_P) = 0 \tag{3.47}$$

$$j_l(d_T) = 0 \tag{3.48}$$

where the superfix $'$ denotes the derivative. Using standard recurrence relations the poloidal boundary condition amounts to $j_{l-1}(d_P) = 0$. Hence both sets of decay rates are simply related to zeros of relevant (but different) spherical Bessel functions. These form an infinite set (for any given l) but we typically are only interested in the most slowly decaying modes since these will dominate the solution at large times. The slowest decay rates are both associated with the $l = 1$ harmonics (being physically of largest scale) and are in the poloidal case: $-d_P^2 = -\pi^2$ with scalar function $S_1(r) = r\,j_1(\pi r)$ and in the toroidal case: $-d_T^2 = -4.4934^2 = -20.1907$ (approx) and $T_1(r) = r\,j_1(4.4934\,r)$.

The decay rates increase with l (i.e. as the spatial scales get more complex, the field decays faster). This means that the diffusive effect on field solutions containing harmonics with high values of l will be correspondingly larger. For a table of decay rates see *Gubbins and Roberts* (1987).

All decay modes are orthogonal over $V + \hat{V}$; since toroidal modes vanish in $\hat{V}$ however they are orthogonal over V (see Appendix B.4 for a proof). The decay modes also form a complete set, as follows from the self-adjointness (section 3.8.3) and then appealing to standard theory (e.g. *Lanczos,* 1961).

3.5 The magnetic energy equation

3.5.1 Magnetic energy

The induction equation governs the pointwise time-dependence of a magnetic field but how do we measure the global 'size'? We introduce the magnetic energy, M, by

$$M(t) = \frac{1}{2}\int_{V+\hat{V}} \mathbf{B}^2\, dV \tag{3.49}$$

which is a non-dimensional measure of the strength of the field and differs slightly from the true physical dimensional magnetic energy M/μ_0. However, we will be concerned only in relative energy growth and so such a linear factor is of no importance. We choose the region over which M is defined to be $V+\hat{V}$ and not simply the dynamo region V, because the magnetic field must be sustained everywhere.

3.5.2 Derivation of the magnetic energy equation

In the following derivation we shall assume that $\hat{V}$ has a finite but large magnetic diffusivity η that will eventually be allowed to tend to infinity (recall that $\eta = 1$ in V). This is to ensure clarity of certain limiting expressions that will appear.

Taking the dot product of the induction equation (equation 3.1) with the magnetic field $\mathbf{B}$ and integrating over all space gives

$$\frac{1}{2}\frac{d}{dt}\int_{V+\hat{V}} |\mathbf{B}|^2\, dV = R_m \int_V \mathbf{B}\cdot\boldsymbol{\nabla}\times(\mathbf{u}\times\mathbf{B}) - \int_{V+\hat{V}} \mathbf{B}\cdot\boldsymbol{\nabla}\times(\eta\boldsymbol{\nabla}\times\mathbf{B})\; dV \tag{3.50}$$

Since the flow $\mathbf{u}$ is only non-zero in V this is the only region that contributes to the velocity term above.

The first integrand on the right hand side is

$$\mathbf{B}\cdot\boldsymbol{\nabla}\times(\mathbf{u}\times\mathbf{B}) = \mathbf{B}\cdot(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{u} - \mathbf{B}\cdot(\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{B} \tag{3.51}$$

and integrating over V:

$$\begin{aligned}\int_V \mathbf{B}\cdot(\mathbf{B}\cdot\boldsymbol{\nabla})\,\mathbf{u}\, dV &= \int_V B_i B_j \frac{\partial u_i}{\partial x_j}\, dV \\ &= \int_V \mathbf{B}\cdot\mathsf{e}\,\mathbf{B}\, dV\end{aligned} \tag{3.52}$$

where $\mathsf{e}_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$ is the rate of strain tensor, representing the local stretching of the fluid (see section 5.2). The notation A_i denotes the i^{th} Cartesian component of the vector $\mathbf{A}$.

The 2nd contributed term from the velocity represents only advection and does no work on the system:

$$\begin{aligned}\int_V \mathbf{B}\cdot(\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{B}\,dV &= \int_V B_i u_j \frac{\partial B_i}{\partial x_j}\,dV = \frac{1}{2}\int_V \boldsymbol{\nabla}\cdot(\mathbf{B}^2\mathbf{u})\,dV \\ &= \oint_{\partial V} \mathbf{B}^2\mathbf{u}\cdot\mathbf{dS} = 0 \end{aligned} \tag{3.53}$$

since the flow $\mathbf{u}$ is incompressible and has no normal component on the boundary ∂V.

We now turn our attention to the diffusive term. Using the standard vector identity

$$\boldsymbol{\nabla}\cdot(\mathbf{F}\times\mathbf{G}) = \mathbf{G}\cdot\boldsymbol{\nabla}\times\mathbf{F} - \mathbf{F}\cdot\boldsymbol{\nabla}\times\mathbf{G} \tag{3.54}$$

for any vectors $\mathbf{F}$ and $\mathbf{G}$ we can write

$$\begin{aligned} &-\int_{V+\hat{V}} \mathbf{B}\cdot\boldsymbol{\nabla}\times(\eta\,\boldsymbol{\nabla}\times\mathbf{B})\,dV \\ &= -\int_{V+\hat{V}} \boldsymbol{\nabla}\cdot(\eta[\boldsymbol{\nabla}\times\mathbf{B}]\times\mathbf{B})\,dV - \int_{V+\hat{V}} \eta|\boldsymbol{\nabla}\times\mathbf{B}|^2\,dV \end{aligned} \tag{3.55}$$

Using the divergence theorem we may turn the first term on the second line into two surface integrals. Also recall from the pre-Maxwell equations that on $r = 1$, $\eta\,\boldsymbol{\nabla}\times\mathbf{B} = \mathbf{E}$ (section 3.3.1) so that

$$-\int_{V+\hat{V}} \boldsymbol{\nabla}\cdot(\eta[\boldsymbol{\nabla}\times\mathbf{B}]\times\mathbf{B})\,dV = \oint_{\partial V} \mathbf{B}\times\mathbf{E}\cdot\mathbf{dS} + \oint_{\partial\hat{V}} \mathbf{B}\times\mathbf{E}\cdot\mathbf{dS} \tag{3.56}$$

The surface $\partial\hat{V}$ has two parts: one at $r = 1$ and one at $r = \infty$. At $r = 1$ the two above surface integrals cancel by continuity of $\mathbf{B}$ and $\mathbf{E}\times\mathbf{dS}$ (we may permute the terms as $\mathbf{B}\times\mathbf{E}\cdot\mathbf{dS} = \mathbf{B}\cdot\mathbf{E}\times\mathbf{dS}$). At infinity we must be a little careful since although $\boldsymbol{\nabla}\times\mathbf{B} = \mathbf{0}$, η will be allowed to become infinite there. However, their product, $\mathbf{E}$, is finite and $O(r^{-2})$; also $\mathbf{B} = O(r^{-3})$ (see section 3.3.2) and so the integrand is $O(r^{-3})$ and therefore vanishes.

Lastly, $\eta|\boldsymbol{\nabla}\times\mathbf{B}|^2 = \mathbf{E}^2/\eta$ in $\hat{V}$ and so vanishes when $\eta\to\infty$ since $\mathbf{E}$ must remain finite and independent of η.

We therefore have derived the equation for the time evolution of magnetic energy:

$$\frac{dM}{dt} = R_m\int_V \mathbf{B}\cdot\mathsf{e}\,\mathbf{B}\,dV - \int_V |\boldsymbol{\nabla}\times\mathbf{B}|^2\,dV \tag{3.57}$$

Energy increases by the stretching action of the flow, and decreases by Ohmic dissipation that stems from the diffusion term. Flow advection has no global effect at all, being integrated out in the analysis.

3.6 Symmetry

Consideration of symmetry properties is a powerful technique that can be used to simplify a problem greatly, for we might be able to vastly reduce the necessary representation of the field. In addition, the different types of solution might behave in a different fashion, perhaps aiding the understanding of the underlying physical processes. Since the Earth is rotating, the symmetry transformations that are most relevant are reflections in the equatorial plane and rotations about its rotation (z) axis.

3.6.1 Symmetry of vectors

Following *Gubbins and Zhang* (1993) and *Sarson* (1994) we define a scalar to be equatorially symmetric (E^S) if

$$s(r,\pi-\theta,\phi) = s(r,\theta,\phi) \tag{3.58}$$

and antisymmetric (E^A) if

$$a(r,\pi-\theta,\phi) = -a(r,\theta,\phi) \tag{3.59}$$

Note the nomenclature: a for antisymmetric and s for symmetric. The same attributes can be given to symmetric (E^S) vectors $\mathbf{S}$ and antisymmetric E^A vectors $\mathbf{A}$:

$$[S_r,S_\theta,S_\phi](r,\pi-\theta,\phi) = [S_r,-S_\theta,S_\phi](r,\theta,\phi) \tag{3.60}$$

$$[A_r,A_\theta,A_\phi](r,\pi-\theta,\phi) = [-A_r,A_\theta,-A_\phi](r,\theta,\phi) \tag{3.61}$$

The θ-component is antisymmetric since the direction of increasing θ becomes inverted on reflection. Scalars and vectors that are symmetric or antisymmetic with respect to a rotation by

$\frac{2\pi}{M}$ about the z-axis are similarly denoted P_M^S or P_M^A and satisfy

$$s(r,\theta,\phi+\frac{M}{2\pi}) = s(r,\theta,\phi) \tag{3.62}$$

$$a(r,\theta,\phi+\frac{M}{2\pi}) = -a(r,\theta,\phi) \tag{3.63}$$

$$[S_r,S_\theta,S_\phi](r,\theta,\phi+\frac{M}{2\pi}) = [S_r,S_\theta,S_\phi](r,\theta,\phi) \tag{3.64}$$

$$[A_r,A_\theta,A_\phi](r,\theta,\phi+\frac{M}{2\pi}) = -[A_r,A_\theta,A_\phi](r,\theta,\phi) \tag{3.65}$$

since there are no coordinate inversions associated with such rotations. If M is unspecified it means it may take on any value.

3.6.2 Symmetry of operators

Suppose a quantity has a particular symmetry with respect to a transformation (e.g. it could be E^S for example) and we apply an operator to it. If both the resulting and original quantities have the same symmetry then the operator is said to be symmetric; if they have opposite symmetries then the operator is antisymmetric. Specifically, the following differential operators have the symmetries:

$$\frac{\partial}{\partial r} : E^S P_M^S \tag{3.66}$$

$$\frac{\partial}{\partial \theta} : E^A P_M^S \tag{3.67}$$

$$\frac{\partial}{\partial \phi} : E^S P_M^S \tag{3.68}$$

which means that the vector operators: $\boldsymbol{\nabla}$, $\boldsymbol{\nabla}\cdot$ and ∇^2 are all of $E^S P_M^S$ symmetry whereas $\boldsymbol{\nabla}\times$ is $E^A P_M^S$. That the curl operator differs in symmetry is not surprising since it involves the anti-symmetric alternating tensor ε_{ijk} that also effects all vector cross-products:

$$[\mathbf{a}\times\mathbf{b}]_i = \varepsilon_{ijk} a_j b_k \tag{3.69}$$

If $\mathbf{a}$ and $\mathbf{b}$ are both either E^A or E^S then $\mathbf{a}\times\mathbf{b}$ is E^A; otherwise, if $\mathbf{a}$ and $\mathbf{b}$ differ in symmetry then it is E^S. The opposite is true of the dot product $\mathbf{a}\cdot\mathbf{b}$ that is E^S if both $\mathbf{a}$ and $\mathbf{b}$ have the same symmetry and E^A if they differ.

3.6.3 Symmetry of harmonics

Spherical harmonics have the property that

$$Y_l^m(\pi-\theta,\phi) = (-1)^{(l-m)} Y_l^m(\theta,\phi) \tag{3.70}$$

so those for which $(l-m)$ is even are E^S and those for which $(l-m)$ is odd are E^A. For the toroidal and poloidal vector harmonics, since $\boldsymbol{\nabla}\times$ essentially switches the symmetry and noting that the unit vector $\hat{\mathbf{r}}$ is E^S, then $\mathbf{T}_l^m$ has the opposite parity to its Y_l^m ingredient and $\mathbf{S}_l^m$ has the same.

3.6.4 Separation of harmonics

We now consider the effect of symmetry on the induction equation. The time derivative $\frac{\partial}{\partial t}$ and ∇^2 are both E^S operators and so is the velocity operator $\boldsymbol{\nabla}\times\mathbf{u}\times$ if $\mathbf{u}$ is E^S. In this case the symmetry with respect to the equatorial plane of a magnetic field $\mathbf{B}$ would be unaffected by these operators, and the E^S and E^A parts of the field would then be independent. Given a flow without E^S symmetry no such separation can occur. Many flows studied in the literature, for example the *KR* flow

$$\mathbf{u} = \mathbf{t}_1 + \varepsilon_1\mathbf{s}_2 + \varepsilon_2\mathbf{s}_2^{2c} + \varepsilon_3\mathbf{s}_2^{2s} \tag{3.71}$$

are E^S (which can easily be checked). In addition, this flow is P_2^S, since the azimuthal dependence is either $\cos 2\phi$ or $\sin 2\phi$. Hence the full set of harmonics is divided into four sets: two divisions into either E^S or E^A and two further into either P_2^S or P_2^A. These sets are

$$E^S P_2^S : \mathbf{T}_1, \mathbf{S}_2, \mathbf{S}_2^{2c}, \mathbf{S}_2^{2s}, \mathbf{T}_3, \mathbf{T}_3^{2s}, \mathbf{T}_3^{2c}, \mathbf{S}_4 \cdots \tag{3.72a}$$

$$E^S P_2^A : \mathbf{S}_1^{1c}, \mathbf{S}_1^{1s}, \mathbf{T}_2^{1c}, \mathbf{T}_2^{1s}, \mathbf{S}_3^{1c}, \mathbf{S}_3^{1s}, \mathbf{S}_3^{3c}, \mathbf{T}_3^{3s}, \cdots \tag{3.72b}$$

$$E^A P_2^S : \mathbf{S}_1, \mathbf{T}_2, \mathbf{T}_2^{2s}, \mathbf{T}_2^{2c}, \mathbf{S}_3^{2c}, \mathbf{S}_3^{2s}, \mathbf{T}_4, \mathbf{T}_4^{2c}, \mathbf{T}_4^{2s}, \cdots \tag{3.72c}$$

$$E^A P_2^A : \mathbf{T}_1^{1c}\mathbf{T}_1^{1s}, \mathbf{S}_2^{1c}, \mathbf{S}_2^{1s}, \mathbf{T}_3^{1c}, \mathbf{T}_3^{1s}, \mathbf{T}_3^{3c}, \mathbf{T}_3^{3s}, \cdots \tag{3.72d}$$

Historically these sets of harmonics have been referred to by their largest scale poloidal member. For example, the set $E^A P_2^S$ is sometimes called the 'axial dipole' symmetry since it contains the dipole $\mathbf{S}_1$. Similarly $E^S P_2^A$ is the 'equatorial dipole' symmetry since it contains $\mathbf{S}_1^{1c}$. The other sets that do not contain a (poloidal) dipole are referred to as quadrapolar, since their largest scale poloidal harmonic has four poles. Sometimes this nomenclature is misleading, for if there is no symmetry separation then the 'axial dipole' and 'equatorial quadrapole' sets could in fact be the

same, both being the whole set. It should hopefully be clear by the context which harmonic set is being referred to; in general however, we shall use the more explicit $E^{S/A}P_2^{S/A}$ or otherwise to specify a particular symmetry.

Other choices of flow may not be E^S so the separation detailed above will not be appropriate. However, other symmetries may apply in these cases leading to a different separation. In practice there is little point detailing the symmetry classes for every choice of flow since this procedure can be automated using interaction selection rules that are detailed in the following section. It is useful though to understand how such separation arises from a mathematical point of view.

3.6.5 Selection rules

Bullard and Gellman (1954) studied the properties of the operator $\boldsymbol{\nabla}\times\mathbf{u}\times$, being the only term in the induction equation that couples field harmonics together (the time derivative and the Laplacian operator preserving the harmonics they act upon). If the velocity is chosen expediently, the magnetic field may partition into independent symmetry classes. *Bullard and Gellman* (1954) detailed a set of selection rules applying when the flow is expanded as a simple sum of vector spherical harmonics, determining whether or not a particular field harmonic belongs to the same symmetry class as another. We introduce the terminology

$$\left(u_\alpha B_\beta B_\gamma\right) = \int_V \boldsymbol{\nabla}\times\left(\mathbf{u}_\alpha\times\mathbf{B}_\beta\right)\cdot\mathbf{B}_\gamma\, dV \tag{3.73}$$

The symmetry classes are effectively determined by the interactions $\left(u_\alpha B_\beta B_\gamma\right)$ and are non zero if:

1. $|l_\alpha - l_\gamma| \le l_\beta \le l_\alpha + l_\gamma$
   That is, the l's form a (possibly degenerate) triangle.

2. At least one of $m_\alpha \pm m_\beta \pm m_\gamma = 0$.

3. For an interaction involving an even number of toroidal harmonics:

   (a) $l_\alpha + l_\beta + l_\gamma$ is even.

   (b) there are an even number of $\sin m\phi$ harmonics ($m = 0$ counts as a cosine).

4. For an interaction involving an odd number of toroidal harmonics:

   (a) $l_\alpha + l_\beta + l_\gamma$ is odd.

(b) there are an odd number of $\sin m\phi$ harmonics.

(c) all the harmonics are different.

5^*. $\left(t_\alpha T_\beta P_\gamma\right) = 0$

To determine the symmetry classes numerically, we start off with $\mathbf{B}_\beta$ as a 'seed symmetry' and use it to generate other harmonics $\mathbf{B}_\gamma$ in the class using the selection rules. For example, if we were interested in generating Earth-like solutions we would begin with $\mathbf{S}_1$ since then we would be guaranteed of its inclusion. We keep on reapplying the selection rules on every harmonic in the class until all the interactions have been accounted for. In general, the classes are infinitely large, so that we must restrict our search to a truncated set of harmonics. In general we will truncate in l, so that $l \leq l_{\max}$ for some prescribed l_{max}.

The inclusion of 5^* leads to an asymmetric set of interaction rules. That is, although one field symmetry may generate another, the converse is not necessarily true. For example, in a toroidal flow, a toroidal field may be generated from a poloidal field (rule 3); however no poloidal field may be created from a toroidal field (rule 5^*). This means, depending on the choice of the 'seed symmetry' either the whole set or a subset of a symmetry class (as predicted by section 3.6.4) may be generated. To alleviate this difficulty and to make the interactions symmetric, we remove the offending rule from the list. In particular, if the term involving the velocity is symmetric (section 3.9.2) then this necessitates such a choice of selection rules.

This removal may lead to the inclusion of more harmonics than is actually necessary in a solution. Of course in this case, numerically the solutions will separate and those symmetries that were falsely incorporated will have no effect. The selection rules merely provide a convenient way of studying different sets of harmonics independently; the final result must be produced by putting all the findings from each symmetry class together since in general it is not *a priori* obvious which will be relevant to the problem.

One of the most important separation of symmetries is that which occurs with an axisymmetric flow. In this case, field harmonics of different azimuthal wave number m are completely independent. In addition, if the flow is E^S then the decoupling is as follows:

$$E^S m = 0 : \mathbf{T}_1, \mathbf{S}_2, \mathbf{T}_3, \mathbf{S}_4, \cdots \tag{3.74a}$$

$$E^A m = 0 : \mathbf{S}_1, \mathbf{T}_2, \mathbf{S}_3, \mathbf{T}_4, \cdots \tag{3.74b}$$

$$E^S m = 1 : \mathbf{S}_1^{1c}, \mathbf{S}_1^{1s}, \mathbf{T}_2^{1c}, \mathbf{T}_2^{1s}, \mathbf{S}_3^{1c}, \mathbf{S}_3^{1s}, \cdots \tag{3.74c}$$

$$E^A m = 1 : \mathbf{T}_1^{1c}, \mathbf{T}_1^{1s}, \mathbf{S}_2^{1c}, \mathbf{S}_2^{1s}, \mathbf{T}_3^{1c}, \mathbf{T}_3^{1s}, \cdots \tag{3.74d}$$

and so on. For a given truncation in l, we therefore need many fewer harmonics in the solution. Although all the different groups above must be studied individually and their results compiled at the end, this is much quicker than performing one large computation, since the time taken per run will scale as some (super linear) power law in the number of harmonics.

In addition, in the range of $R_m = 0$–1000 of geophysical interest, the harmonic sets of large m also have correspondingly large l (since $m \leq l$). Ignoring the fact that we would have to increase the truncation level in order to keep the number of harmonics in each set the same, higher values of l lead to a much higher Ohmic dissipation rate (see section 3.4.2) so dynamo action is more difficult. For very large R_m this is no obstacle, but in the range under consideration this encumbrance is significant and kinematic dynamo action is seldom seen for $m > 2$.

A further point on axisymmetric flows is that in general, fields of a single wave number m represent solid body rotations; this perhaps is to be expected since the flow itself has no preferred azimuthal orientation and can be rotated at will. We may show this by writing the ϕ dependence of $\mathbf{B}$ as a complex Fourier mode $e^{im\phi}$ where $i^2 = -1$. If we also have a dependence $\mathbf{B} \propto e^{\lambda t}$ then

$$e^{im\phi}\, e^{\lambda t} = e^{i(m\phi + \Im(\lambda))t}\, e^{\Re(\lambda)t} \tag{3.75}$$

so that the field rotates in time with angular velocity $\Im(\lambda)/m$ and has an amplitude that grows or decays depending on the sign of $\Re(\lambda)$. Note that this may only occur if the field symmetry contains both sine and cosine harmonics in azimuth, both being necessary to span the relevant rotations.

Any eigenvector of the induction equation is determined only up to an arbitrary multiplicative constant; if the vector is complex this introduces a degeneracy since the argument of this constant is unknown. If only a single wave number m is present then this amounts to a rotation of the field about the z-axis. In general we combine the two eigenvectors forming a complex conjugate pair into a new eigenvector such that its largest scale poloidal mode attains $1 + 0\,i$ on $r = 1$.

3.7 Existing bounds

Theoretical bounds for the dynamo problem are very useful, for they provide rigorous boundaries to the relevant quantities and parameters involved. Most of the bounds surveyed below are lower limits on the value of R_m needed for dynamo action, that serve as not only a useful comparison for geophysical estimates and for numerical models, but also for designing experiments since they place constraints on their size and typical flow velocities required.

3.7.1 Backus's bound on R_m

If we define dynamo action to be the manifestation of growing magnetic energy (at least initially) as in the definition given in section 3.5.2, this places constraints on the size of R_m. The bound of *Backus* (1958) is derived as follows. If M increases, then from equation (3.57) we obtain the necessarily condition that:

$$R_m \int_V \mathbf{B} \cdot \mathsf{e}\,\mathbf{B}\, dV \geq \int_V |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV \tag{3.76}$$

If μ_{max} is the pointwise maximum rate of strain of e, that is, its largest pointwise eigenvalue then

$$R_m\, \mu_{max} \int_V \mathbf{B}^2\, dV \geq R_m \int_V \mathbf{B} \cdot \mathsf{e}\,\mathbf{B}\, dV \tag{3.77}$$

We may then write

$$R_m\, \mu_{max} \geq \frac{\int_V |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV}{\int_V \mathbf{B}^2\, dV} \tag{3.78}$$

The field $\mathbf{B}$ is unknown, so in order to derive a bound we replace the right hand side by its minimum value.

Backus (1958) facilitated this minimisation by replacing the domain of the denominator by $V + \hat{V}$ although thereby reducing the bound's sharpness. It becomes

$$R_m\, \mu_{max} \geq \min \frac{\int_V |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV}{\int_{V+\hat{V}} \mathbf{B}^2\, dV} \tag{3.79}$$

As in section 3.5.2 we now consider the electrically insulating $\hat{V}$ in its limiting form, i.e. where η is large but finite. From equation (3.55), adding in $\eta |\boldsymbol{\nabla} \times \mathbf{B}|^2 = O(\eta^{-1})$ in $\hat{V}$ we can write, allowing $\eta \to \infty$

$$\int_V |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV = \int_V |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV + \int_{\hat{V}} \eta |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV \tag{3.80}$$

$$= \int_{V+\hat{V}} \mathbf{B} \cdot \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B})\, dV \tag{3.81}$$

$$= -\int_V \mathbf{B} \cdot \nabla^2 \mathbf{B}\, dV - \int_{\hat{V}} \mathbf{B} \cdot \frac{\partial \mathbf{B}}{\partial t} \tag{3.82}$$

where we have used $\frac{\partial \mathbf{B}}{\partial t} = -\boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B})$ in $\hat{V}$. Expanding as a linear sum of decay modes: $\mathbf{B} = \sum_\alpha c_\alpha \mathbf{B}_\alpha$ and recalling that their time dependence is $e^{-d_\alpha^2 t}$ and that in V they are eigenvectors of

∇^2 with eigenvalue $-d_\alpha^2$, then

$$\min \frac{\int_V |\nabla \times \mathbf{B}|^2 \, dV}{\int_{V+\hat{V}} \mathbf{B}^2 \, dV} = \min \frac{\sum_\alpha \, d_\alpha^2 c_\alpha^2 \int_{V+\hat{V}} \mathbf{B}_\alpha^2 \, dV}{\sum_\alpha \, c_\alpha^2 \int_{V+\hat{V}} \mathbf{B}_\alpha^2 \, dV} \tag{3.83}$$

by orthogonality of $\mathbf{B}_\alpha$ over all space. The minimum required is d_α^2 where $-d_\alpha^2$ is the slowest decay rate of $-\pi^2$, being associated with the $l = 1$ poloidal decay mode. Hence Backus showed that

$$R_m \mu_{max} \geq \pi^2 \tag{3.84}$$

3.7.2 Proctor's bound

Proctor (1977a) improved on this bound by directly minimising equation (3.78) using a Lagrange multiplier method. A direct decay mode expansion now can only find the minimum in the toroidal-field only case, since the poloidal modes are not orthogonal over V. This provides an upper bound of $4.4934^2 = 20.1907$ for the minimum sought, being associated with the $l = 1$ toroidal decay rate. However, Proctor found that a $l = 1$ poloidal field gave the minimum bound of

$$R_m \mu_{max} \geq 12.29 \tag{3.85}$$

Figure 3.1 shows how the two $l = 1$ poloidal scalar functions compare. Solid is that of Proctor's bound: $S_1(r) = r\, j_1(\alpha r) + j_1(\alpha) r^2/2$ where $\alpha^2 = 12.29$; dashed is that of Backus: $S_1(r) = r\, j_1(\pi r)$. Each function is determined within an arbitrary linear factor that will typically be chosen by some form of normalisation. However, the plotted S_1 scalar functions have not been scaled, for example so that they attain a maximum value of unity, since the curves would be graphically indistinguishable from each other. It is remarkable how similar these two curves are given the approximate 25% difference in the related bounds. A derivation of Proctor's bound is given in section 3.9.6

3.7.3 Other bounds on R_m

Childress (1969) also produced a similar bound although based on the maximum flow rate u_{max} rather than the maximum strain rate:

$$R_m \, u_{max} \geq \pi \tag{3.86}$$

Due to the invariance of the pre-Maxwell equations under a rotation (e.g. *Davidson*, 2001), we may transform to a co-rotating frame where u_{max} is least to give the best bound. This cannot be applied to measures involving the strain rate which is independent of the rotational frame of

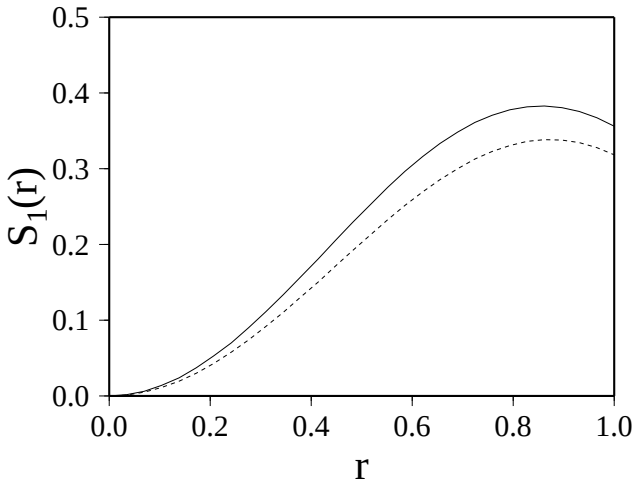


Figure 3.1: Comparison between the minimising $l = 1$ poloidal field scalar functions from the bounds of *Proctor* (1977a) (solid) and *Backus* (1958) (dashed). The curves are not normalised and are subject to arbitrary linear scaling factors.

reference.

3.7.4 A note on the trivial bound

It is worth pointing out why 0 is not the minimum of (3.78). If this were to be the case then the field would have to be curl free everywhere; such a field could not be everywhere continuous. Let us consider a potential field: $\mathbf{B} = \boldsymbol{\nabla}[f_\alpha(r)Y_\alpha]$ where the solenoidal condition of $\mathbf{B}$ gives that pointwise f_α is either of the form r^{l_α} or $r^{-(l_\alpha+1)}$ (see section 3.3.2); regularity requires that near the origin it must be of the former type, and exclusion of external fields requires the latter towards infinity. At at least one point in between there must be a boundary at which the behaviour changes. This introduces a discontinuity in $\mathbf{B}$ since we cannot match both $f_\alpha(r)$ (as required by the θ and ϕ components) and $f_\alpha^{'}(r)$ (required by the radial component) across the boundary. Hence in order that $\mathbf{B}$ be everywhere continuous we must have either a zero field, or one that is not everywhere of potential form.

3.7.5 Bounds on motion

Toroidal motions cannot sustain magnetic fields (one of the antidynamo theorems of section 2.5) so some amount of poloidal flow is needed. *Busse* (1975) considered the generated poloidal

field to get a lower bound on u_r necessary for dynamo action. Instead of dotting the induction equation with $\mathbf{B}$ before integrating, he instead dotted with $\mathbf{r}$ and multiplied by $(\mathbf{B}\cdot\mathbf{r})$, producing

$$\frac{1}{2}\frac{d}{dt}\int_V (\mathbf{r}\cdot\mathbf{B})^2\,dV = \int_V (\mathbf{r}\cdot\mathbf{B})\,\mathbf{r}\cdot\left(R_m\boldsymbol{\nabla}\times(\mathbf{u}\times\mathbf{B})+\nabla^2\mathbf{B}\right)\,dV \tag{3.87}$$

He was able to manipulate this into the form

$$R_m(\mathbf{u}\cdot\mathbf{r})_{max} \geq \left(\frac{2E_P}{E_T+E_P}\right)^{\frac{1}{2}} \tag{3.88}$$

where E_P and E_T are the poloidal and toroidal field energies respectively, taken over all space. Given a field $\mathbf{B}$ this can be used to calculate the minimum radial motion required. Unfortunately without knowing $\mathbf{B}$, replacing the right hand side with its minimum gives little information since it may be made arbitrarily small by choosing the field to have a vanishingly small poloidal component. Nonetheless, this bound may be seen as evidence that some poloidal motion is necessary for dynamo action and that the outer core could not be everywhere stably stratified ($u_r \equiv 0$).

3.7.6 Upper bounds on time-dependence

We may bound the rate of growth of magnetic energy by discarding the Ohmic dissipation term of equation (3.57). In a similar argument to that used above in the bounds of *Backus* (1958) and *Proctor* (1977a) we can then write

$$\frac{dM}{dt} \leq R_m\mu_{max}\int_V \mathbf{B}^2\,dV \leq R_m\mu_{max}\int_{V+\hat{V}} \mathbf{B}^2\,dV \tag{3.89}$$

This inequality may be re-written:

$$\frac{1}{\sqrt{M}}\frac{d}{dt}\sqrt{M} \leq R_m\mu_{max} \tag{3.90}$$

This bounds the normalised instantaneous rate of magnetic energy growth by the maximum strain rate of the flow. If we define λ_E to be the maximum value of the left hand side over all possible magnetic fields (see section 3.9.1), this also is similarly bounded. In section 3.9.3 we show λ_E to be an upper bound on the eigenvalue growth rate $\Re(\lambda)$, where the eigenmode ansatz is $\mathbf{B} = \hat{\mathbf{B}}e^{\lambda t}$, so that

$$\Re(\lambda) \leq \lambda_E \leq R_m\mu_{max} \tag{3.91}$$

The equality in bound (3.89) is attained only when (a) the maximal rate of strain is achieved everywhere in the flow, (b) **B** is aligned with the direction of maximal straining at all points in V and (c) if Ohmic dissipation is negligible (none of which will occur in reality). Physically this means that the larger the ability of the flow to stretch field lines, the larger the possible growth.

The above bound on the growth rate, and indeed most of the lower bounds on R_m presented above depend on pointwise maximum properties of the flow: either maximum strain rates or flow velocities. Such results are clearly sub-optimal, since the flow could be almost entirely at rest with just a small vigorously moving section. These bounds then would not reflect the fluid as a whole, but only tiny fraction. A better situation, especially given our definition of R_m based on the rms average of the flow, would be to formulate a bound in terms of the fluid kinetic energy. Unfortunately *Núñez* (2002) showed that no bound on λ can be formulated in terms of the quadratic norm $\int_V \mathbf{u}^2 \, dV$. However, it can be shown that (Núñez, personal communication, 2003):

$$\Re(\lambda) \leq \frac{1}{2} R_m^2 u_{max}^2 \tag{3.92}$$

3.8 Adjoints

Roberts (1972b) noted a striking similarity with the eigenvalue growth rates of E^A and E^S fields under the change of sign $\mathbf{u} \rightarrow -\mathbf{u}$. That is, by changing the direction of the flow the favoured symmetry switched from that of a dipole to a quadrapole. This result had important ramifications for the geodynamo since the fact that the Earth's field was principally dipolar could perhaps lead to inferences about the direction of flow. This mathematical peculiarity was traced to the so called adjoint equation, that has the property of having the same eigenvalues and biorthogonal eigenvectors as the induction equation. *Gibson and Roberts* (1967) showed that the adjoint equation could be constructed to have the same form as the induction equation but with a flow of reversed direction, although the boundary conditions were non physical. *Proctor* (1977b) studied the 'comparison problem' where the poloidal and toroidal fields were made to vanish on the boundaries and found that the similarity observed by *Roberts* (1972b) was exact in this case: E^A and E^S fields had precisely the same growth rate under change of velocity direction. If the boundary conditions were not important to the problem, that is, if they were pushed away from the dynamo region by for example, a quiescent conducting layer on top of the outer core, then this result might be expected to hold approximately (*Kaiser and Tilgner*, 1999). However, as has already been discussed (section 2.6.7) the spherical boundary conditions are of fundamental importance to the kinematic problem so this result is unlikely to hold in the physical Earth. In

any case, the non-linear Navier-Stokes equations are not invariant under a directional change of flow, so that the sign of **u** is not a free parameter but it inherent in the dynamical system.

3.8.1 The magnetic energy adjoint

The adjoint we will consider here is that relevant to the magnetic energy problem and differs from other formulations in the literature (see *Sarson*, 1994, for an overview). Consider a vector differential operator **L** that acts upon a magnetic field **B**. Solutions of the equation $\mathbf{L}\,\mathbf{B} = \mathbf{0}$ can also be represented as

$$\int_{V+\hat{V}} \mathbf{B}^{\dagger} \cdot \mathbf{L}\,\mathbf{B}\, dV = 0 \tag{3.93}$$

for all vectors $\mathbf{B}^{\dagger}$ provided that the space spanned by them is sufficiently large. The operator **L** has an associated adjoint operator called $\mathbf{L}^{\dagger}$ and is defined as (e.g. *Lanczos*, 1961)

$$\int_{V+\hat{V}} \mathbf{B}^{\dagger} \cdot \mathbf{L}\,\mathbf{B}\, dV = \int_{V+\hat{V}} \mathbf{B} \cdot \mathbf{L}^{\dagger}\mathbf{B}^{\dagger}\, dV \tag{3.94}$$

To find $\mathbf{L}^{\dagger}$ we typically successively integrate by parts, transferring the spatial derivatives of **B** onto $\mathbf{B}^{\dagger}$ and imposing suitable boundary conditions to kill off the terms than cannot be expressed in the above form. The adjoint operator $\mathbf{L}^{\dagger}$ may be shown to have the same eigenvalues as **L** although having different eigenvectors in general. These form a biorthogonal set, that is, each eigenvector of $\mathbf{L}^{\dagger}$ is orthogonal to all eigenvectors of **L** except that which has an eigenvalue being the complex conjugate of that of the adjoint eigenvector.

3.8.2 The adjoint of the induction operator

We define the induction operator **L** (of which we might want to find the eigenvalues) by

$$\mathbf{L} = R_m \boldsymbol{\nabla} \times \mathbf{u} \times - \boldsymbol{\nabla} \times \eta \boldsymbol{\nabla} \times \tag{3.95}$$

Therefore the equation defining $\mathbf{L}^{\dagger}$ is

$$\int_{V+\hat{V}} \mathbf{B}^{\dagger} \cdot \left[R_m \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}) - \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) \right] dV = \int_{V+\hat{V}} \mathbf{B} \cdot \mathbf{L}^{\dagger}\mathbf{B}^{\dagger}\, dV \tag{3.96}$$

where $\eta = 1$ in V and is infinite in $\hat{V}$. The velocity term can be written

$$R_m \int_V \left\{ \boldsymbol{\nabla} \cdot \left[(\mathbf{u} \times \mathbf{B}) \times \mathbf{B}^{\dagger} \right] + (\mathbf{u} \times \mathbf{B}) \cdot (\boldsymbol{\nabla} \times \mathbf{B}^{\dagger}) \right\} dV \tag{3.97}$$

and if the flow $\mathbf{u}$ is non-slip on the boundary $r = 1$, then it is simplified to

$$-R_m \int_V \mathbf{B} \cdot (\mathbf{u} \times \boldsymbol{\nabla} \times \mathbf{B}^\dagger)\, dV \tag{3.98}$$

The diffusion term may be manipulated into the form

$$\begin{aligned} &\int_{V+\hat{V}} -\mathbf{B}^\dagger \cdot \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B})\, dV \\ &= \int_{V+\hat{V}} -\mathbf{B} \cdot \boldsymbol{\nabla} \times \left(\eta \boldsymbol{\nabla} \times \mathbf{B}^\dagger\right) dV + \oint_{\partial V+\partial \hat{V}} \left[\mathbf{B}^\dagger \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) - \mathbf{B} \times \left(\eta \boldsymbol{\nabla} \times \mathbf{B}^\dagger\right)\right] \cdot \mathbf{dS} \end{aligned} \tag{3.99}$$

If we choose $\mathbf{L}^\dagger = -R_m \mathbf{u} \times \boldsymbol{\nabla} \times - \boldsymbol{\nabla} \times \eta \boldsymbol{\nabla} \times$ and the boundary conditions on $\mathbf{B}^\dagger$ usual for a magnetic field, then this equation is satisfied by the usual continuity properties of section 3.3.1. In this case, the adjoint equation differs from the induction equation by having a different flow-field interaction but possesses the same diffusive term and boundary conditions. Notice that when $R_m = 0$ the adjoint and original equations coincide.

3.8.3 Self adjoint operators

In certain cases the adjoint operator with its boundary conditions is exactly the same as those for the original problem. In this case, the operator is said to be self-adjoint. Such operators possess real eigenvalues and mutually orthogonal eigenvectors. Such an example is the diffusion operator, a special case of the above when $R_m = 0$. Thus the decay modes (eigenmodes of this problem) form an orthogonal (and complete) set over all space (see also section B.4) so that any solution $\mathbf{B}(t)$ to this equation may be expressed as a sum over these modes:

$$\mathbf{B}(t) = \sum_\alpha c_\alpha e^{-d_\alpha^2 t} \mathbf{B}_\alpha \tag{3.100}$$

where the spatial component of each mode $\mathbf{B}_\alpha$ is normalised to have unit magnetic energy. The magnetic energy of the total field is

$$\frac{1}{2} \int_{V+\hat{V}} \mathbf{B}(t)^2\, dV = \sum_\alpha c_\alpha^2 e^{-2d_\alpha^2 t} \tag{3.101}$$

by the normalisation and orthogonal properties. Therefore the evolution of the energy is completely determined by the eigenvalues of the original system. Indeed in this case, to maximise the field energy at any time $t > 0$ given the restriction of having unit initial energy, we would put all of the energy into the $l = 1$ poloidal mode that decays at (the slowest) rate of $-\pi^2$.

3.8.4 Non self-adjoint operators

If the operator is non self-adjoint (e.g. when $R_m \neq 0$) then the above behaviour does not apply: in general eigenvectors are not mutually orthogonal and the energy has a different time-dependence from the linear system. Eigenvectors of the induction operator are complete (*Browder*, 1952; 1953) so any solution may be written as a sum

$$\mathbf{B}(t) = \sum_{\alpha} c_\alpha \mathbf{B}_\alpha \, e^{\lambda_\alpha t} \tag{3.102}$$

where again $\mathbf{B}_\alpha$ is the spatial part of the eigenmode, but now the energy is

$$M = \frac{1}{2}\int_{V+\hat{V}} \mathbf{B}^2 \, dV = \frac{1}{2}\sum_{\alpha\beta} c_\alpha c_\beta e^{(\lambda_\alpha+\lambda_\beta)t} \int_{V+\hat{V}} \mathbf{B}_\alpha \cdot \mathbf{B}_\beta \, dV \tag{3.103}$$

The non-orthogonality of the eigenvectors means that the resulting non-zero cross products introduce a time dependence not related to any of the individual eigenvalues.

For a simple example, consider a $\mathbf{t}_1$ flow acting upon a poloidal field. Although no dynamo may operate (section 2.5) so that all the linear eigenvalues have negative real parts, much toroidal field can be produced by differential rotation before the eventual decay. Hence a period of transient growth occurs that is not predicted by the eigenmode formulation.

This is important for kinematic theory since we are interested in magnetic fields that grow only until the Lorentz force becomes important. Finite time transience rather than exponentially growing solutions may therefore be of relevance in the physical system.

3.9 Magnetic energy instability

3.9.1 Measures of magnetic energy growth

Historically, magnetic field growth has been linked exclusively to the eigenvalues of the induction operator, when we make the assumption that each eigenmode is of the form $\mathbf{B}_i = \hat{\mathbf{B}}_i e^{\lambda_i t}$. Such an expansion is complete (*Browder*, 1952; 1953) so we can write any solution of the problem as

$$\mathbf{B} = \Re\left[\sum_i A_i \mathbf{B}_i \, e^{\lambda_i t}\right] \tag{3.104}$$

where we take only the physically relevant real part. In general the eigenmode formulation allows complex eigenvectors and eigenvalues that indicate a rotation about the z-axis at a frequency depending on the azimuthal wave number m of each mode. Any initial field will almost

certainly have a projection onto all of the eigenvectors, and so the time-evolution of the field at large times is controlled by the most unstable mode, i.e. that which is the fastest growing or most slowly decaying, associated with the eigenvalue of maximal real part. We could take the view that measuring the growth of magnetic fields by this mode is very much related to a physical process, since it is naturally selected by the system, albeit possibly after only times $t \gg 1$. Numerically to find out if a growing eigenmode exists, we increase R_m from 0 and to track the eigenvalue with maximal real part. We denote the lowest value of R_m for which there exists an eigenvalue with $\Re(\lambda) = 0$ as R_m^c, the linear critical magnetic Reynolds number. Whether or not there exists such a value however is crucially dependent on the choice of flow studied, and is therefore arguably not a robust way of measuring field growth. Indeed, geophysically we are interested in magnetic field growth only until the Lorentz force becomes important, which can have any time-dependence, not necessarily exponential.

Another way of characterising magnetic energy growth is instead to look at very short times, and pose the question, 'what is the magnetic field structure that grows instantaneously, energetically the fastest?'. This may be a good indicator of growth over short time periods and in general will not coincide with that predicted by eigenmode analysis since the induction operator is not self-adjoint. We may formulate this as

$$\begin{aligned} \lambda_E &= \max_{\mathbf{B}} \frac{1}{\sqrt{M}} \frac{d}{dt} \sqrt{M} \\ &= \max_{\mathbf{B}} \frac{1}{2M} \frac{dM}{dt} \end{aligned} \tag{3.105}$$

If $\lambda_E > 0$ then at least one magnetic field structure will grow instantaneously; if $\lambda_E < 0$ the flow is stable for all magnetic perturbations (since the maximum growth rate is negative). The value of R_m for which $\lambda_E = 0$ gives the energetically critical state and will be called R_m^E; it is the greatest lower bound (infimum) on R_m for energetic instability. Note that $R_m^E \leq R_m^c$ since it is always easier to grow fields instantaneously than for long periods (for a proof of this see section 3.9.3); equality is obtained if the underlying operator is self-adjoint. In every case considered in this thesis, R_m^E exists; by contrast, in most cases, R_m^c does not. A simple example where $R_m^E = R_m^c$ is in the diffusion problem, where the induction operator is self-adjoint. In such cases, the linear eigenmodes are orthogonal and so no transient effects can be caused by mutual superposition. Thus the time dependence is controlled completely by the eigenvalues, and in particular so is the instantaneous growth.

A further alternative method of characterising magnetic field growth is to solve the fully time-dependent kinematic problem. In this manner, we can answer such questions as,

'what is the maximum relative growth of field energy over a time t?' That is

$$\max_{\mathbf{B}(0)} \frac{M(t)}{M(0)} \tag{3.106}$$

where the maximisation is taken over all possible initial fields $\mathbf{B}(0)$. When t is small we expect agreement with the instantaneous analysis; when t is large we expect to mimic the eigenmode formulation. In the latter case, such a method alleviates the worry about whether or not the fastest growing eigenmode has had a chance to dominate and in general incorporates effects from all the eigenmodes, growing or not. We may track the evolution of the field and physically monitor how it changes over time: this should give far greater insight than looking at just the final state (for large t), governed by the most unstable eigenmode of the system.

3.9.2 The onset of magnetic energy instability

We will now investigate the instantaneous onset problem of magnetic energy growth. Recall that we measure this by λ_E:

$$\begin{aligned} \lambda_E &= \max_{\mathbf{B}} \frac{1}{\sqrt{M}} \frac{d}{dt} \sqrt{M} \\ &= \max_{\mathbf{B}} \frac{1}{2M} \frac{dM}{dt} \end{aligned} \tag{3.107}$$

The maximisation is taken over all possible magnetic fields, constrained to be solenoidal and to satisfy the continuity and boundary conditions at $r = 1$. Let us suppose that the maximum of equation (3.105) is attained by field $\mathbf{B}$. In order to find an equation that λ_E must satisfy, we perturb $\mathbf{B}$ by a relatively small field $\varepsilon\delta\mathbf{B}$, whose magnitude is $O(\varepsilon) \ll 1$ where $\mathbf{B} = O(1)$. At such a stationary point, λ_E must be at least quadratic in ε and using equations (3.105) and (3.57), terms of first order give:

$$\lambda_E \int_{V+\hat{V}} \delta\mathbf{B} \cdot \mathbf{B}\, dV = R_m \int_V \delta\mathbf{B} \cdot \mathsf{e}\mathbf{B}\, dV - \int_V (\boldsymbol{\nabla} \times \delta\mathbf{B}) \cdot (\boldsymbol{\nabla} \times \mathbf{B})\, dV \tag{3.108}$$

The symmetry of each term makes λ_E necessarily real, since if $\mathbf{B}$ were complex, we could take our arbitrary perturbation $\delta\mathbf{B}$ to be proportional to its complex conjugate: $\delta\mathbf{B} = \varepsilon\mathbf{B}^*$, forcing λ_E to be real.

Using a standard identity, we may recast this in terms of the Laplacian operator:

$$\lambda_E \int_{V+\hat{V}} \delta\mathbf{B} \cdot \mathbf{B}\, dV = R_m \int_V \delta\mathbf{B} \cdot \mathsf{e}\mathbf{B}\, dV + \int_V \delta\mathbf{B} \cdot \nabla^2\mathbf{B}\, dV - \oint_{\partial V} \delta\mathbf{B} \times [\boldsymbol{\nabla} \times \mathbf{B}] \cdot \mathbf{dS} \tag{3.109}$$

Ideally we would now like an additional simplification to remove the unwanted perturbation $\delta\mathbf{B}$ to leave an equation for just the maximising $\mathbf{B}$ and λ_E. However, due to the presence of the boundary integral this is intractable, and we are left with the slightly unsatisfactory position of an equation for λ_E and $\mathbf{B}$ but containing any arbitrary perturbation $\delta\mathbf{B}$. In the case where $\hat{V}$ has a non-zero electrical conductivity, the relevant equations can be written down exactly (see section 3.9.8) and in a modified situation they can be analytically solved (section 3.9.6). Here however, progress is only possible numerically using a Galerkin method as detailed in chapter four.

The term involving the velocity is now symmetric in $\delta\mathbf{B}$ and $\mathbf{B}$, and so any selection rules used to separate the harmonics into disjoint sets (section 3.6.5) must be symmetric. It is for this reason that we deleted the asymmetric rule so that the results could be applied here.

3.9.3 Using λ_E to bound the eigenvalue growth rate

We now show that if

$$\lambda_E = \max_{\mathbf{B}} \frac{1}{2M}\frac{dM}{dt} \tag{3.110}$$

and

$$\lambda\,\mathbf{B} = \frac{\partial \mathbf{B}}{\partial t} = \mathbf{L}\,\mathbf{B} \tag{3.111}$$

then the following bound holds:

$$\Re(\lambda) \leq \lambda_E \tag{3.112}$$

Physically this intuitive since one might expect that it is easier to grow magnetic fields instantaneously than over a long period of time. We give a proof of the result here, courtesy of Dr. Richard Kerswell (personal communication, 2001). It immediately follows that $R_m^E \leq R_m^c$. It is first necessary to introduce the energy inner product:

$$(\mathbf{B}_1, \mathbf{B}_2) = \int_{V+\hat{V}} \mathbf{B}_1 \cdot \mathbf{B}_2 \, dV \tag{3.113}$$

We write the eigenvalue as $\lambda = \lambda_r + i\lambda_i$ that has associated eigenvector $\mathbf{B} = \mathbf{B}_r + i\mathbf{B}_i$. It follows that, since $\mathbf{B}$ is real:

$$\lambda_r \mathbf{B}_r - \lambda_i \mathbf{B}_i = \mathbf{L}\,\mathbf{B}_r \tag{3.114a}$$

$$\lambda_r \mathbf{B}_i + \lambda_i \mathbf{B}_r = \mathbf{L}\,\mathbf{B}_i \tag{3.114b}$$

Taking the energy inner product of equation (3.114a) with $\mathbf{B}_r$ and of equation (3.114b) with $\mathbf{B}_i$ and adding, gives

$$\lambda_r = \frac{(\mathbf{B}_r, \mathbf{L}\mathbf{B}_r) + (\mathbf{B}_i, \mathbf{L}\mathbf{B}_i)}{|\mathbf{B}_r|^2 + |\mathbf{B}_i|^2} \tag{3.115}$$

It remains to show that this may be bounded above by

$$\lambda_E = \max_{\mathbf{B}} \frac{(\mathbf{B}, \mathbf{L}\mathbf{B})}{|\mathbf{B}|^2} \tag{3.116}$$

Writing $(\mathbf{B}, \mathbf{L}\mathbf{B}) = g(\mathbf{B})$ and $|\mathbf{B}|^2 = h(\mathbf{B})$ this amounts to showing that

$$\frac{g(\mathbf{B}_1) + g(\mathbf{B}_2)}{h(\mathbf{B}_1) + h(\mathbf{B}_2)} \leq \max_{\mathbf{B}} \frac{g(\mathbf{B})}{h(\mathbf{B})} = \frac{g(\mathbf{B}_M)}{h(\mathbf{B}_M)} \tag{3.117}$$

where $\mathbf{B}_M$ is the field at which the maximum is attained. It follows that

$$\frac{g(\mathbf{B}_1)}{h(\mathbf{B}_1)} \leq \frac{g(\mathbf{B}_M)}{h(\mathbf{B}_M)} \tag{3.118}$$

so that

$$g(\mathbf{B}_1)\, h(\mathbf{B}_M) - g(\mathbf{B}_M)\, h(\mathbf{B}_1) \leq 0 \tag{3.119}$$

and similarly for $\mathbf{B}_2$. It is then straightforward to show that

$$\begin{aligned} &\frac{g(\mathbf{B}_1) + g(\mathbf{B}_2)}{h(\mathbf{B}_1) + h(\mathbf{B}_2)} - \frac{g(\mathbf{B}_M)}{h(\mathbf{B}_M)} \\ =&\frac{h(\mathbf{B}_M)\,[g(\mathbf{B}_1) + g(\mathbf{B}_2)] - g(\mathbf{B}_M)\,[h(\mathbf{B}_1) + h(\mathbf{B}_2)]}{[h(\mathbf{B}_1) + h(\mathbf{B}_2)]\, h(\mathbf{B}_M)} \leq 0 \end{aligned} \tag{3.120}$$

3.9.4 An adjoint derivation of magnetic energy instability

An alternative derivation of the equation for instantaneous magnetic energy instability can be obtained as follows. Let us write the induction equation in the following form:

$$\frac{\partial \mathbf{B}}{\partial t} = \mathbf{L}\mathbf{B} \tag{3.121}$$

Taking the dot product with $\mathbf{B}$ and integrating over all space gives the equation for the rate of change of magnetic energy that we can write as

$$\frac{1}{2M}\frac{dM}{dt} = \frac{\int_{V+\hat{V}} \mathbf{B}\cdot\mathbf{L}\mathbf{B}\, dV}{\int_{V+\hat{V}} \mathbf{B}^2\, dV} \tag{3.122}$$

If we maximise this expression we shall obtain λ_E as defined in section 3.9.1. Applying the variational methodology we obtain:

$$2\lambda_E \int_{V+\hat{V}} \delta\mathbf{B}\cdot\mathbf{B}\,dV = \int_{V+\hat{V}} [\delta\mathbf{B}\cdot\mathbf{L}\,\mathbf{B} + \mathbf{B}\cdot\mathbf{L}\,\delta\mathbf{B}]\,dV \tag{3.123}$$

where we have collected all first order terms together; $\delta\mathbf{B}$ is an $O(1)$ field perturbation (as before). This can be rewritten as

$$\lambda_E \int_{V+\hat{V}} \delta\mathbf{B}\cdot\mathbf{B}\,dV = \int_{V+\hat{V}} \delta\mathbf{B}\cdot\frac{(\mathbf{L}+\mathbf{L}^\dagger)}{2}\mathbf{B}\,dV \tag{3.124}$$

The operator $\frac{(\mathbf{L}+\mathbf{L}^\dagger)}{2}$ is self-adjoint and so has a complete set of orthonormal eigenvectors $\mathbf{v}_i$ with corresponding real eigenvalues μ_i. Writing $\mathbf{B} = \sum_i c_i\mathbf{v}_i$ and $\delta\mathbf{B} = \sum_i d_i\mathbf{v}_i$ we get

$$\lambda_E = \max\frac{\sum_i \mu_i c_i d_i}{\sum_i c_i d_i} = \max\{\mu_i\} \tag{3.125}$$

Therefore λ_E is the maximum eigenvalue of $\frac{(\mathbf{L}+\mathbf{L}^\dagger)}{2}$ and is attained when $\mathbf{B}$ is the associated eigenvector.

In particular, if the problem is self adjoint (for example in the case of the decay problem when $R_m = 0$), the instantaneous growth rates coincide with the linear eigenvalues. This is to be expected since the eigenvectors are orthogonal and no transient superposition effects are possible.

3.9.5 Finite time growth of magnetic energy

Since the induction equation is linear, we can in fact write down exact solutions in the following way. Writing it again in the form

$$\frac{\partial\mathbf{B}}{\partial t} = \mathbf{L}\mathbf{B} \tag{3.126}$$

where $\mathbf{L}$ is the induction operator, we can regard it as an ordinary differential equation in time and so the solution is

$$\mathbf{B}(t) = e^{\mathbf{L}t}\mathbf{B}(0) \tag{3.127}$$

The exponential function can be made sense of by a power series expansion:

$$e^{\mathbf{L}t} = 1 + \mathbf{L}t + (\mathbf{L}t)^2/2! + (\mathbf{L}t)^3/3! + \cdots \tag{3.128}$$

Suppose that we would like to calculate the maximum relative growth of magnetic energy over some time t. We use the energy inner product

$$(\mathbf{B}_1, \mathbf{B}_2) = \int_{V+\hat{V}} \mathbf{B}_1 \cdot \mathbf{B}_2 \, dV \tag{3.129}$$

This problem can be formulated as a maximisation over all initial fields $\mathbf{B}(0)$ with unit energy ($M(0) = 1$):

$$\max M(t) = \max \left(e^{\mathbf{L}t} \mathbf{B}(0), e^{\mathbf{L}t} \mathbf{B}(0) \right) \tag{3.130}$$

$$= \max \left(\left(e^{\mathbf{L}t} \right)^{\dagger} e^{\mathbf{L}t} \mathbf{B}(0), \mathbf{B}(0) \right) \tag{3.131}$$

$$= \max \left(e^{\mathbf{L}^{\dagger} t} e^{\mathbf{L}t} \mathbf{B}(0), \mathbf{B}(0) \right) \tag{3.132}$$

since we may take the adjoint operation inside any linear sum (of which the exponential is a limiting form) and it is independent of t, involving only spatial derivatives. The methodology just derived can be applied to the instantaneous maximised energy growth problem, which is equivalent to maximising the growth of $M(t)$ for vanishingly small t. If $t \ll 1$ then to first order $e^{\mathbf{L}t} = 1 + \mathbf{L}t$ so again if $M(0) = 1$:

$$\max M(t) = \left((1 + \mathbf{L}^{\dagger} t)(1 + \mathbf{L}t) \mathbf{B}(0), \mathbf{B}(0) \right) \tag{3.133}$$

Taking the derivative with respect to time gives that

$$\max \frac{dM}{dt} = \max \left((\mathbf{L}^{\dagger} + \mathbf{L}) \mathbf{B}(0), \mathbf{B}(0) \right) \tag{3.134}$$

This is of a similar form to that in the previous section; again λ_E is the largest eigenvalue of the operator $(\mathbf{L}^{\dagger} + \mathbf{L})/2$. (Note the extra factor of 2 needed in the above equation to make it conform to the definition of λ_E).

3.9.6 A variational method applied to Proctor's bound

Recall that *Proctor* (1977a) addressed the problem

$$\lambda_P = \min \frac{\int_V |\boldsymbol{\nabla} \times \mathbf{B}|^2 \, dV}{\int_V \mathbf{B}^2 \, dV} \tag{3.135}$$

by a Lagrange multiplier method. Here, we are able to use the variational method as above to solve it analytically. In order to do so however, like Proctor we must allow the external region $\hat{V}$

to be finitely electrically conducting and subsequently let its magnetic diffusivity tend to infinity in a well defined limit. We are at liberty to approach this limit in whatever way that we choose: therefore we only consider $\eta = \eta(r)$ since this facilitates the analysis. From the derivation of the magnetic energy equation (equation 3.55), we see that the only difference that an electrically conducting exterior makes is the inclusion of the Ohmic dissipation term in $\hat{V}$. The relevant minimisation is

$$\lambda_P = \min \frac{\int_{V+\hat{V}} \eta |\boldsymbol{\nabla} \times \mathbf{B}|^2 \, dV}{\int_V \mathbf{B}^2 \, dV} \tag{3.136}$$

which agrees with (3.135) in the electrically insulating limit where $\eta \boldsymbol{\nabla} \times \mathbf{B}$ remains finite but $\boldsymbol{\nabla} \times \mathbf{B} \to \mathbf{0}$ in $\hat{V}$.

To form the variational equation, assume that the minimum value, λ_P, is attained with some field $\mathbf{B}$. At this stationary point, perturbations to this field $\varepsilon \delta \mathbf{B} = O(\varepsilon) \ll 1$ where $\mathbf{B} = O(1)$ will affect the minimum at least quadratically in ε. Collecting terms linear in ε gives

$$-\int_{V+\hat{V}} \eta \, (\boldsymbol{\nabla} \times \mathbf{B}) \cdot (\boldsymbol{\nabla} \times \delta \mathbf{B}) \; dV + \lambda_P \int_V \mathbf{B} \cdot \delta \mathbf{B} \, dV = 0 \tag{3.137}$$

Using a standard identity equation (3.137) may be written, noting that $\eta = 1$ in V:

$$\begin{aligned} &-\oint_{\partial V + \partial \hat{V}} \eta \, \delta \mathbf{B} \times (\boldsymbol{\nabla} \times \mathbf{B}) \cdot \mathbf{dS} + \int_V \delta \mathbf{B} \cdot \left(\nabla^2 \mathbf{B} + \lambda_P \mathbf{B} \right) dV \\ &-\int_{\hat{V}} \delta \mathbf{B} \cdot \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) \, dV = 0 \end{aligned} \tag{3.138}$$

The surface integrals vanish due to the usual continuity conditions imposed and the behaviour at infinity. Using the notation $\chi_V(\mathbf{r}) = 1$ if $\mathbf{r} \in V$ and 0 otherwise, we rewrite the remaining equation in implicit form:

$$\int_{V+\hat{V}} \delta \mathbf{B} \cdot \left\{ \chi_V(\mathbf{r}) \left[\nabla^2 \mathbf{B} + \lambda_P \mathbf{B} \right] - \chi_{\hat{V}}(\mathbf{r}) \left[\boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) \right] \right\} dV = 0 \tag{3.139}$$

Using the Lemma in Appendix C, we obtain the explicit equations for any stationary point of (3.136), in particular its minimum:

$$\nabla^2 \mathbf{B} + \lambda_P \mathbf{B} + \boldsymbol{\nabla} \xi = \mathbf{0} \qquad \text{in } V \tag{3.140a}$$

$$-\boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) + \boldsymbol{\nabla} \xi = \mathbf{0} \qquad \text{in } \hat{V} \tag{3.140b}$$

where ξ is continuous at $r = 1$.

These equations may be solved (see Appendix C for a derivation) to give that $\mathbf{B}$ is of

$l = 1$ poloidal form with

$$S_1(r) = r\, j_1(\alpha r) + j_1(\alpha) r^2/2 \tag{3.141}$$

in the insulating limit: $\eta \to \infty$ in $\hat{V}$. The constant α solves

$$j_1(\alpha) + \frac{2}{3}\, j_0(\alpha) = 0 \tag{3.142}$$

that numerically can be evaluated to be $\lambda_P = \alpha^2 = 12.29$ (to 2 decimal places).

3.9.7 The singular electrically insulating limit

To obtain the limiting form of the equations in the electrically insulating limit, if one naively ignores equation (3.140b) and instead imposes the relevant boundary conditions at $r = 1$, this demonstrably leads to the wrong answer. In fact, on taking the curl of equation (3.140a) we arrive back at the diffusion problem which would give $\lambda_P = \pi^2$. So what has gone wrong and why is this limit ostensibly singular?

The answer is of course that it is not singular, and we have made an error en route in the derivation. The increase of the domain of the numerator from V in equation (3.135) to $V + \hat{V}$ in equation (3.136) makes no difference in the electrically insulating limit. The extra addition splits up into two exactly cancelling integrals:

$$\int_{\partial \hat{V}} \eta \mathbf{B} \times (\boldsymbol{\nabla} \times \mathbf{B}) \cdot \mathbf{dS} + \int_{\hat{V}} \mathbf{B} \cdot \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B})\, dV = \int_{\hat{V}} \eta |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV \to 0 \tag{3.143}$$

in the limit as $\eta \to \infty$, and all parts stay finite and non-zero. If we ignore the integral over $\hat{V}$ then we must also ignore the boundary integral. This happens to cancel out a similar boundary term from V which is why in the conducting case the problem involves only volume integrals. If we take the limit properly then equation (3.138) should lead to the following equation in V:

$$\int_V \delta \mathbf{B} \cdot \left[\nabla^2 \mathbf{B} + \lambda_P \mathbf{B}\right]\, dV - \oint_{\partial V} \delta \mathbf{B} \times (\boldsymbol{\nabla} \times \mathbf{B}) \cdot \mathbf{dS} = 0 \tag{3.144}$$

We therefore have two options:

(a) treat the exterior as finitely conducting, solve in both V and $\hat{V}$, match at $r = 1$ and then let $\eta \to \infty$ in $\hat{V}$ or

(b) treat the exterior as an electrical insulator and include the extra (now non-cancelling) boundary term in the analysis when solving in V. We may use the usual boundary conditions at $r = 1$.

Incorporation of the boundary term makes an analytic approach intractable which is why option (a) was taken by *Proctor* (1977a) and is done so here. Option (b) may be treated by numerical Galerkin methods that are detailed in chapter four.

3.9.8 The differential equations of instantaneous magnetic energy growth

One might think that, following the success of the variational approach to Proctor's bound that perhaps an analytic treatment of the equation for instantaneous energy growth is possible; such an attempt is outlined here. Including the finitely electrically conducting region $\hat{V}$ in the diffusive term of equation (3.108) as in the previous treatment gives the following equation:

$$R_m \int_V \delta\mathbf{B}\cdot\mathsf{e}\mathbf{B}+\mathbf{B}\cdot\mathsf{e}\,\delta\mathbf{B}-\int_{V+\hat{V}}\eta\,(\boldsymbol{\nabla}\times\mathbf{B})\cdot(\boldsymbol{\nabla}\times\delta\mathbf{B})\,dV=\lambda_E\int_{V+\hat{V}}\delta\mathbf{B}\cdot\mathbf{B}\,dV \qquad (3.145)$$

On transforming the diffusive terms and noting that in this case the boundary integrals cancel out, we end up with

$$\int_V \delta\mathbf{B}\cdot\left(R_m\,\mathsf{e}\mathbf{B}+\nabla^2\mathbf{B}-\lambda_E\,\mathbf{B}\right)\,dV-\int_{\hat{V}}\delta\mathbf{B}\cdot(\boldsymbol{\nabla}\times(\eta\boldsymbol{\nabla}\times\mathbf{B})+\lambda_E\,\mathbf{B})\,dV \qquad (3.146)$$

Using the Lemma in section C.1.1 these are equivalent to:

$$R_m\,\mathsf{e}\mathbf{B}+\nabla^2\mathbf{B}-\lambda_E\,\mathbf{B}=\boldsymbol{\nabla}\xi \qquad \text{in } V \qquad (3.147)$$

$$\boldsymbol{\nabla}\times(\eta\boldsymbol{\nabla}\times\mathbf{B})+\lambda_E=-\boldsymbol{\nabla}\xi \qquad \text{in } \hat{V} \qquad (3.148)$$

for some continuous function ξ. The action of e couples vector harmonics together in the same fashion as the operator $R_m\boldsymbol{\nabla}\times\mathbf{u}\times$ in the induction equation. In order to formulate a solution, one could try extract the toroidal and poloidal harmonic components of each equation and use a finite difference scheme in radius, but that leaves the question of how to let $\eta\to\infty$ in $\hat{V}$ in a well defined way. As has already pointed out, simply solving in V and using the usual electrically insulating boundary conditions is incorrect: either we solve in $\hat{V}$ as well or we are forced to include a boundary integral earlier on in the analysis which stops us getting this far. Notice that the unknown eigenvalue λ_E appears in both domains which inhibits any kind of local solution, for example in $\hat{V}$, which could then be matched to that in V to determine λ_E.

Chapter 4

Numerical methods

In this chapter we detail the numerical procedures used in this thesis, coded in FORTRAN 90, and their convergence. In particular, we introduce the use of a fully spectral Galerkin method to solve ostensibly intractable sets of equations. We motivate this method by solving a simple example outlined below. Some references that the reader may turn to for more detail are *Boyd* (2001), *Canuto et al.* (1988) or *Fornberg* (1998).

4.1 A simple example

Consider the simple 1D diffusion equation for $u = u(x,t)$:

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \tag{4.1}$$

subject to the boundary conditions

$$\frac{\partial u}{\partial x} + \frac{u}{x} = 0 \qquad x \geq 1, t \geq 0 \tag{4.2}$$

$$u(0,t) = 0 \tag{4.3}$$

Let us seek solutions of the form $u(x,t) = P(x)\, e^{\lambda t}$. The boundary condition implies that the solution in $x \geq 1$ is $u(x,t) = \frac{A e^{\lambda t}}{x}$ for some constant A, and so its behaviour is completely determined by its value on $x = 1$. The decay modes are then:

$$u(x,t) = \begin{cases} e^{-\nu_n^2 t} \sin \nu_n x & 0 \leq x \leq 1 \\ e^{-\nu_n^2 t} \frac{\sin \nu_n}{x} & x \geq 1 \end{cases} \tag{4.4}$$

where ν_n is the n^{th} strictly positive root of the equation $\tan k + k = 0$. The slowest decaying mode is associated with the value $\nu_1 = 2.0288$ (or equivalently $-\nu_1^2 = -4.1159$).

4.1.1 A range of numerical approaches

Since we may solve equation (4.1) analytically it is interesting to discuss how this problem might be tackled numerically. We would start by formulating the equation as an eigenvalue problem

$$\lambda P(x) - \frac{d^2P}{dx^2} = 0 \tag{4.5}$$

and discretise the differential operator in some way to give it a finite representation. For example, we might use a finite difference scheme whereby we solve for $P(x)$ on some grid, and express the derivatives at each location as linear combinations of P evaluated on neighbouring grid points (with some degree of error), using a Taylor expansion. As we increase the number of grid points and the accuracy of the differencing scheme, we hope that converged solutions are found. An alternative method is to use a so called spectral scheme, where we express $P(x)$ as a finite sum over known functions, for example

$$P(x) = \sum_{n=0}^{N} A_n x^n \tag{4.6}$$

In this case the differential operator is exact and in fact the diffusion problem amounts to solving

$$\lambda \sum_{n=0}^{N} A_n x^n - \sum_{n=0}^{N-2} (n+1)(n+2) A_{n+2} x^n = 0 \tag{4.7}$$

Equating the first $n-2$ powers of x gives $n-2$ constraints, and adding in the two boundary conditions produces n equations for n unknowns. Alternatively we may use a collocation scheme to derive the $n-2$ constraints on A_i from the equation above. This is where, instead of equating powers of x, we use the fact that equation (4.7) holds everywhere in $[0,1]$ and we simply evaluate it at $n-2$ locations (at our disposal). Again by increasing the number of basis functions N and concomitantly the number of collocation points, we hope to find converged solutions. We may expand $P(x)$ over any basis that we choose, for example, Fourier series, Bessel functions, power series, Chebyshev polynomials are all possibilities, the choice of which will be motivated by examining the particular properties of the problem. For example, if odd and even solutions separate then we might either use a sine or cosine Fourier series or odd and even polynomials to address each symmetry separately.

A clever way of approaching spectral problems is to use a 'recombined basis' (*Boyd*,

2001). This is where each basis function individually satisfies all the relevant boundary conditions and so we do not need to explicitly include them when we solve the system. For example, in the above problem the following is a basis:

$$U_n(x) = \begin{cases} T_n(x) - \frac{1+n^2}{2}x & 0 \leq x \leq 1 \\ \frac{1-n^2}{2x} & x \geq 1 \end{cases} \tag{4.8}$$

where $T_n(x)$ is the n^{th} Chebyshev polynomial with $n \geq 2$ (see section 4.4.2). It is a simple exercise to check that each $U_n(x)$ satisfies both boundary conditions. We can then expand $P(x)$ in this basis and obtain the full quota of n equations using a collocation method without 'wasting' two equations for the boundary conditions. In addition, these problems are often numerically better conditioned since the boundary conditions are implicit in each matrix element and not represented by a few rows in what could be very large matrices.

An important property of spectral schemes is that very few basis functions may be required to represent the unknown solution extremely well. Indeed, these methods often exhibit so called 'spectral accuracy', that is, they converge exponentially fast. This relies on the properties of the basis functions and it certainly holds in the case of Chebyshev polynomials.

4.1.2 A Galerkin method

Recall that the methodology of the collocation method using a recombined basis is as follows. We expand u in the spectral basis, substitute this into the differential expression, equate it to zero and evaluate it at the collocation points. In a Galerkin method, instead of equating the differential expression to zero, we make it orthogonal to some set of functions. In general we choose this set of functions to be the basis we first thought of. As we increase the number of basis functions N, not only do we improve the representation of the unknown function $P(x)$ but also the representation of 0, being orthogonal to all functions. Writing $P(x) = \sum_i A_i U_i(x)$ the Galerkin formulation is then

$$\int_0^1 U_i \left[\lambda \sum_j A_j U_j - \sum_j A_j \int_V \frac{d^2 U_j}{dx^2} \right] dx \tag{4.9}$$

$$= \lambda \sum_j A_j \int_0^1 U_i U_j \, dx - \sum_j A_j \int_V U_i \frac{d^2 U_j}{dx^2} \, dx = 0 \tag{4.10}$$

for $i = 1, 2, 3 \ldots, N$. This gives N equations in $(A_1, A_2, \ldots, A_N)$ that we solve as an eigenvalue problem

$$\lambda \, \mathsf{B}\,\mathbf{v} - \mathsf{D}\,\mathbf{v} = 0 \tag{4.11}$$

where the matrices are

$$\mathsf{B}_{ij} = \int_0^1 U_i U_j \, dx \qquad \mathsf{D}_{ij} = \int_0^1 U_i \frac{d^2 U_j}{dx^2} \, dx \tag{4.12}$$

and $\mathbf{v} = (A_1, A_2, A_3, \ldots, A_N)$ is the vector of the unknown coefficients. We might be able to formulate these integrals analytically from knowledge of the expressions for U_i. Indeed, in general we can integrate by parts, resulting in a "weak form" of the system with lower spatial derivatives than the original. If analytic methods are intractable we could use an exact quadrature scheme, that essentially relates the integral to samples of the integrand at a number of judiciously chosen points. This is then equivalent to a collocation method using the same grid.

There is some flexibility when choosing the region of integration of this Galerkin scheme. Clearly we must include the region $[0,1]$ but since the behaviour of u in $x > 1$ is completely determined from $u(1)$ we are at liberty to choose any region $[0,a]$ where $a \geq 1$. This will modify the matrices although the method will still converge to the same solution. If $a > 1$, no comparison with the collocation method can now be made, and the defining equation here is

$$\lambda P(x) = \frac{\partial u}{\partial t}(x) = \frac{\partial u}{\partial t}(1)/x = \frac{\partial^2 u}{d x^2}(1)/x \tag{4.13}$$

Therefore the contributions to B and D are respectively:

$$\int_1^a \frac{U_i(1) U_j(1)}{x^2} \, dx \qquad \int_1^a \frac{U_i(1) U_j''(1)}{x^2} \, dx \tag{4.14}$$

Table 4.1 shows exponential convergence in N to the slowest decay rate of -4.1159 (approx) using the package *Maple*, numerically calculating terms to 30 decimal places. Since the solution sought is odd we restrict the basis to its odd subset: U_3, U_5, U_7, etc. On increasing N by 2, the logarithmic error approximately halves for both choices of a. Choosing $a = 1$ in this case exhibits the best convergence although both choices clearly work.

	$\log_{10}(\lambda - \nu_1)$	
N	$a = 1$	$a = \infty$
1	-1.602	0.055
2	-4.522	-1.533
4	-11.90	-5.141
6	-20.62	-9.456

Table 4.1: Exponential convergence towards the analytic decay rates in N, for the two regions of integration $[0,a)$ where $a = 1$ and $a = \infty$.

4.2 Comparison of various numerical methods

The Galerkin paradigm has been outlined above, and we detail its relevant applications later. For now, we present a comparison of the different numerical methods that could be employed to solve scalar differential equations.

1. Finite difference:
   (a) Expand the unknown scalar on a grid.
   (b) Impose that the equation be satisfied at these points.
   (c) Handle the derivatives by central/forward/backward "differences" of neighbouring grid points. This is not exact and depends on the 'order' of the scheme.
   (d) Each grid point will only depend on those immediately around it, leading to banded matrices and more efficient storage.
   (e) The inaccuracies in representing the unknown function stem from the finite grid used and the order of the differencing scheme.
2. Collocation using a re-combined basis:
   (a) Expand the unknown scalar as a linear sum of N known basis functions, each of which satisfies the boundary conditions.
   (b) Impose that the equation be satisfied at a set of N points (which are in general expediently chosen).
   (c) The derivatives can be handled exactly by differentiating the known functions.
   (d) The resulting system is dense (that is, almost all matrix entries are non-zero), although it is exact for the finite basis representation used.
3. Galerkin
   (a) Expand the unknown scalar as a linear sum of N known basis functions.
   (b) Introduce some integral measure of orthogonality.
   (c) Impose that the equation be orthogonal to each basis function, instead of being identically zero.
   (d) Derivatives are usually handled using integration by parts, yielding a "weak form" of the equation. The resulting integrals involve lower order derivatives than those of the original equation.

(e) The resulting system is dense, although exact for the finite basis representation used.

If it clear that finite difference schemes are very different in structure to Galerkin and re-combined collocation methods, having between them many similarities. Consider a spectral scheme using a recombined basis where we expand $\mathbf{B} = \sum_{i=1}^{N} A_i \mathbf{B}_i$ in order to numerically solve the equation $\mathscr{D}\mathbf{B} - \lambda\mathbf{B} = \mathbf{0}$ where $\mathscr{D}$ is some vector differential operator. The residual is defined to be

$$[\mathscr{D} - \lambda\mathbf{I}] \sum_{i=1}^{N} A_i \mathbf{B}_i = \mathbf{R}_N(\mathbf{r}, A_0, A_1, A_2, \ldots, A_N) \tag{4.15}$$

that is the error between the finite approximation **B** and the true (unknown) solution that solves the equation exactly; in general it will also be spatially dependent so is a function of the position vector **r**. Both Galerkin and re-combined collocation schemes seek to minimise the residual: collocation by imposing that it be zero at certain (expediently chosen) grid points, and Galerkin by imposing that it be orthogonal to the basis used. If we expand **R** in terms of the same basis, then these schemes amount to imposing that the first N spectral coefficients vanish. Should we use a basis expansion that converges exponentially (e.g. Fourier or Chebyshev series) then the residual will converge to zero exponentially fast, i.e. $I(\mathbf{R}) \propto c^{-N}$ for some integral measure I of the residual (e.g. the energy) and some constant $c > 1$. This is so called "spectral convergence".

A Galerkin method numerically integrating by a quadrature scheme using the set of Gaussian-abscissas (discrete sampling points) gives precisely the same matrix formulation as a re-combined collocation method on the same grid (see *Boyd*, 2001, section 4.4). Thus in certain cases these two schemes are exactly equivalent. In general, collocation methods refer to a spectral expansion, for example in Chebyshev polynomials, but imposing the boundary conditions explicitly rather than implicitly, using up one or more rows of the matrix system. Another method that may be used is a so called tau-method (e.g. *Fornberg*, 1998), similar to the general collocation scheme but instead of imposing that the equation be pointwise zero, it uses the Galerkin orthogonality condition instead. Thus tau is akin to Galerkin, but imposes the boundary conditions explicitly.

4.3 Application of the Galerkin method

4.3.1 The induction equation

We now detail the application of a Galerkin method to solve the induction equation eigenvalue problem. Expanding the unknown magnetic field as a linear sum over a basis of vectors so that

$$\mathbf{B} = \sum_{i=1}^{N} A_i \mathbf{B}_i \tag{4.16}$$

we may formulate the Galerkin problem as a matrix equation:

$$\lambda \, \mathsf{B}_V \, \mathbf{v} = (R_m \mathsf{U} + \mathsf{D}_V) \, \mathbf{v} \tag{4.17}$$

where the vector $\mathbf{v} = (A_1, A_2, \cdots, A_N)$ represents the unknown coefficients, and we have imposed orthogonality over the unit sphere V. The matrix entries are:

$$\begin{aligned} [\mathsf{B}_V]_{ij} &= \int_V \mathbf{B}_i \cdot \mathbf{B}_j \, dV \qquad \mathsf{U}_{ij} = \int_V \mathbf{B}_i \cdot \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}_j) \, dV \\ [\mathsf{D}_V]_{ij} &= \int_V \mathbf{B}_i \cdot \nabla^2 \mathbf{B}_j \, dV \end{aligned} \tag{4.18}$$

This is a generalised eigenvalue problem that may be solved using standard methods. In addition we may determine the secular variation of any given field similarly. Suppose that **B** (and therefore **v**) is known; since $\frac{\partial \mathbf{B}}{\partial t}$ satisfies the same boundary conditions as **B** we can also expand it as a sum over the basis $\mathbf{B}_i$, represented by the vector, **s**, say. Substituting these expressions into the induction equation and integrating against each basis function gives

$$\mathbf{s} = \mathsf{B}_V^{-1} (R_m \mathsf{U} + \mathsf{D}_V) \mathbf{v} \tag{4.19}$$

The matrix $\mathsf{B}_V^{-1}(R_m \mathsf{U} + \mathsf{D}_V)$ is a discretisation of $\mathbf{L} = R_m \boldsymbol{\nabla} \times \mathbf{u} \times \; - \boldsymbol{\nabla} \times \eta \boldsymbol{\nabla} \times$, the induction operator, and can either be used to time-propagate magnetic fields or to find the eigenvalue growth rates of linear modes (this being equivalent to finding the generalised eigenvalues as above).

4.3.2 Variational equations

In chapter three we derived two equations that could not be solved analytically and appeared to be intractable because of the presence of boundary integrals. These were, the equation describing

the instantaneously most unstable magnetic field, growing in energy at a rate λ_E (section 3.9.2)

$$\lambda_E \int_{V+\hat{V}} \delta\mathbf{B}\cdot\mathbf{B}\,dV = R_m \int_V \delta\mathbf{B}\cdot\mathsf{e}\mathbf{B}\,dV + \int_V \delta\mathbf{B}\cdot\nabla^2\mathbf{B}\,dV - \oint_{\partial V} \delta\mathbf{B}\times[\boldsymbol{\nabla}\times\mathbf{B}]\cdot\mathbf{dS} \qquad (4.20)$$

and a simpler equation coming from the application of the variational principle to the bound of *Proctor* (1977a), section 3.9.7:

$$\int_V \delta\mathbf{B}\cdot\left[\nabla^2\mathbf{B}+\lambda_P\mathbf{B}\right]dV - \oint_{\partial V}\delta\mathbf{B}\times(\boldsymbol{\nabla}\times\mathbf{B})\cdot\mathbf{dS} = 0 \qquad (4.21)$$

In both of these equations, $\delta\mathbf{B}$ is any perturbation field. Galerkin methods offer a powerful way of treating such problems numerically: we expand $\mathbf{B}$ over some basis and then take $\delta\mathbf{B}$ to be each basis function in turn, forming a linear system of equations. In this sense, the above equations really represent an orthogonality relation written in some kind of weak form. Thus the numerical approach to the variational equations is no different from the application of a Galerkin method to the induction equation, even with the inclusion of ostensibly troublesome boundary integrals.

We then therefore formulate equation (4.20) as a generalised eigenvalue problem:

$$\lambda_E\,\mathsf{B}_{V+\hat{V}}\,\mathbf{v} = (R_m\mathsf{E}+\mathsf{D}_{V+\hat{V}})\,\mathbf{v} \qquad (4.22)$$

where λ_E is the largest such eigenvalue, and

$$\begin{aligned}
[\mathsf{B}_{V+\hat{V}}]_{ij} &= \int_{V+\hat{V}} \mathbf{B}_i\cdot\mathbf{B}_j\,dV \qquad \mathsf{E}_{ij} = \int_V \mathbf{B}_i\cdot\mathsf{e}\mathbf{B}_j\,dV \\
[\mathsf{D}_{V+\hat{V}}]_{ij} &= \int_V \mathbf{B}_i\cdot\nabla^2\mathbf{B}_j\,dV - \oint_{r=1}\mathbf{B}_i\times\left(\boldsymbol{\nabla}\times\mathbf{B}_j\right)\cdot\mathbf{dS} \\
&= -\int_V(\boldsymbol{\nabla}\times\mathbf{B}_i)\cdot(\boldsymbol{\nabla}\times\mathbf{B}_j)\,dV
\end{aligned} \qquad (4.23)$$

and similarly for equation (4.21). Note that B and D are defined over all space and are both symmetric; in particular the discretised diffusion term is identical to that which comes about from a Galerkin treatment of the Laplacian operator (see section 4.6.1).

4.4 Choice of basis

4.4.1 Properties of the basis functions

We detail in this section the choice of basis that we shall use in our study. Recall that each basis function must individually satisfy all boundary and other relevant conditions. In particular they

must be solenoidal so we expand as usual in poloidal-toroidal form; the insulating boundary conditions mean that they can be explicitly written as

$$\mathbf{S}_\alpha = \begin{cases} \boldsymbol{\nabla}\times\boldsymbol{\nabla}\times[S_\alpha Y_\alpha \hat{\mathbf{r}}] & \text{in } V \\ -l_\alpha S_\alpha(1)\boldsymbol{\nabla}\left(\frac{Y_\alpha}{r^{l_\alpha+1}}\right) & \text{in } \hat{V} \end{cases} \tag{4.24}$$

$$\mathbf{T}_\alpha = \begin{cases} \boldsymbol{\nabla}\times[T_\alpha Y_\alpha \hat{\mathbf{r}}] & \text{in } V \\ 0 & \text{in } \hat{V} \end{cases} \tag{4.25}$$

with

$$\frac{dS_\alpha}{dr}(1) + l_\alpha S_\alpha(1) = 0 \qquad T_\alpha(1) = 0 \tag{4.26}$$

Regularity imposes that as $r \to 0$

$$S_\alpha(r) = O(r^{(l+1)}) \qquad T_\alpha(r) = O(r^{(l+1)}) \tag{4.27}$$

In addition, since we include the origin in our domain, the system must be invariant under the mapping:

$$\theta \to \theta + \pi \qquad r \to -r \tag{4.28}$$

The harmonics satisfy $Y_l^m(\theta,\phi) \to (-1)^l Y_l^m(\theta,\phi)$ under $\theta \to \theta+\pi$ see section (B.1.3). This means that

$$S_l^m(r) = (-1)^{(l+1)} S_l^m(-r) \tag{4.29}$$

which comes from, for example $l(l+1)Y_l^m\, S_l^m(r)/r^2$, the radial component of a poloidal vector harmonic, and noting that the extra power of -1 stems from the reversal of $\hat{\mathbf{r}}$ under such a transformation.

A similar result holds for the toroidal vectors, following from the behaviour of the θ-component. Again we see that $T_l^m(r) = (-1)^{l+1} T_l^m(-r)$ since $\hat{\boldsymbol{\theta}}$ changes sign.

This means that if l is an odd number, then the corresponding radial functions S_l^m and T_l^m must be even; similarly if l is an even number, then the functions must be odd. This parity separation means that we can either exclude all even or all odd functions in r for any given l, which increases the degree of expansion that we can achieve by a factor of two for a fixed number of terms. The interested reader should consult *Kerswell and Davey* (1996) for more details on symmetry separation.

We shall index the basis functions by the superscript n so that, for example

$$\mathbf{S}_\alpha = \sum_{n=1}^{N_{max}} \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times [A_\alpha Y_\alpha S_\alpha^n \hat{\mathbf{r}}] \tag{4.30}$$

for some coefficients $A_1, A_2, \ldots, A_{N_{max}}$.

Following the behavioural constraint at the origin, we might seek radial basis functions of the form:

$$S_\alpha^n(r) = r^{l_\alpha+1} f_\alpha^n(r) \qquad T_\alpha^n(r) = r^{l_\alpha+1} g_\alpha^n(r) \tag{4.31}$$

for some non-singular functions f_α^n and g_α^n. However, this is perhaps not a good way to proceed since significant relative errors are introduced when evaluating terms in $r \ll 1$, due to the large pre-multiplying power of r (Dr. David Ivers, personal communication, 2002). We instead follow *Zhang and Busse* (1988) in using an expansion of the form:

$$S_\alpha^n(r) = r\, f_\alpha^n(r) \qquad T_\alpha^n(r) = r\, g_\alpha^n(r) \tag{4.32}$$

where f_α^n and g_α^n are both $O(r)$ at the origin, which satisfies the weaker form of S_α^n and T_α^n being $O(r^2)$ at $r = 0$. We truncate these functions as $n \leq N_{max}$ similarly to that in l, using $l \leq L_{max}$.

Although the regularity condition of $O(r^{l+1})$ is necessary, we might expect that the solution obtained automatically satisfies this constraint whatever basis functions we choose. We therefore use it only to hone the basis and do not worry too much about whether each basis function individually satisfies it. This approach is typical in a finite difference scheme where only $O(r)$ behaviour is enforced, since power law constraints are not in a tractable form for such a method.

4.4.2 Chebyshev Polynomials

We will make extensive use of Chebyshev polynomials in the basis that we choose, so a few details are given here. Chebyshev polynomials may be defined by

$$T_n(\cos\theta) = \cos(n\theta) \tag{4.33}$$

and satisfy the equation:

$$(1-x^2)T_n''(x) - xT_n'(x) + n^2 T_n(x) = 0 \tag{4.34}$$

They are efficiently determined by the three-term recurrence relation and starting values:

$$T_0(x) = 1 \tag{4.35a}$$

$$T_1(x) = x \tag{4.35b}$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{4.35c}$$

Other useful relations are:

$$2T_m(x)T_n(x) = T_{n+m}(x) + T_{n-m}(x) \qquad n \geq m \tag{4.36}$$

$$T_n(1) = 1 \tag{4.37}$$

$$T_n(-x) = (-1)^n T_n(x) \tag{4.38}$$

$$T_n(0) = \begin{cases} (-1)^m & n = 2m \\ 0 & n = 2m+1 \end{cases} \tag{4.39}$$

$$T_n'(1) = n^2 \tag{4.40}$$

$$T_n''(1) = \frac{1}{3}n^2(n^2-1) \tag{4.41}$$

$$\int_0^1 T_n(x)dx = \left[\frac{T_{n+1}(x)}{2(n+1)} + \frac{T_{n-1}(x)}{2(n-1)}\right]_{x=0}^{x=1} \qquad n \geq 2 \tag{4.42}$$

$$\int_{-1}^1 \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}dx = \begin{cases} 0 & m \neq n \\ \frac{\pi}{2}(1+\delta_{n0}) & n = m \end{cases} \tag{4.43}$$

$$(1-x^2)T_n'(x) = -nxT_n(x) + nT_{n-1}(x) \tag{4.44}$$

A recurrence relation links the derivatives:

$$T_k'(x) = \sum_{i=0}^{k} a_i T_i(x) \tag{4.45a}$$

$$a_i = \begin{cases} 0 & i \geq k \\ \frac{2k\delta_{i(k-1)} + a_{i+2}}{1+\delta_{i0}} & i = k-1, k-2, k-3, \ldots, 0 \end{cases} \tag{4.45b}$$

For example,

$$T_6'(x) = 12T_5(x) + 12T_3(x) + 12T_1(x) \tag{4.46}$$

$$T_7'(x) = 14T_6(x) + 14T_4(x) + 14T_2(x) + 7T_0(x) \tag{4.47}$$

A Chebyshev series is closely related to a cosine Fourier series (4.33) and consequently has many similar properties; in particular they converge exponentially fast (see *Boyd*, 2001,

section 2.11). It may also be shown that the Chebyshev polynomials are complete. Another point of note is that they have a definite parity (4.38), which we will exploit.

4.4.3 The Chebyshev basis

No combination of the Chebyshev polynomials T_0, T_1 and T_2 can be used to fulfil the regularity, boundary condition and parity requirements of equations (4.27), (4.26) and (4.29), and so they are used as building blocks for the radial basis. The following is a set of basis functions that satisfy all the required constraints, and additionally exhibit completeness and excellent convergence properties.

$$T_\alpha^n(r) = r \begin{cases} T_{2n+1}(r) - T_1(r) & l_\alpha \text{ odd} \\ T_{2n+2}(r) - T_0(r) & l_\alpha \text{ even, n odd} \\ T_{2n+2}(r) - T_2(r) & l_\alpha \text{ even, n even} \end{cases} \tag{4.48}$$

$$S_\alpha^n(r) = r \begin{cases} T_{2n+1}(r) - \frac{(2n+1)^2+l+1}{l+2} T_1(r) & l_\alpha \text{ odd} \\ T_{2n+2}(r) - o_1 T_2(r) - o_2 T_0(r) & l_\alpha \text{ even, n odd} \\ T_{2n+2}(r) - e_1 T_2(r) - e_2 T_0(r) & l_\alpha \text{ even, n even} \end{cases} \tag{4.49}$$

with $l_\alpha \geq 1$,$n \geq 1$ and

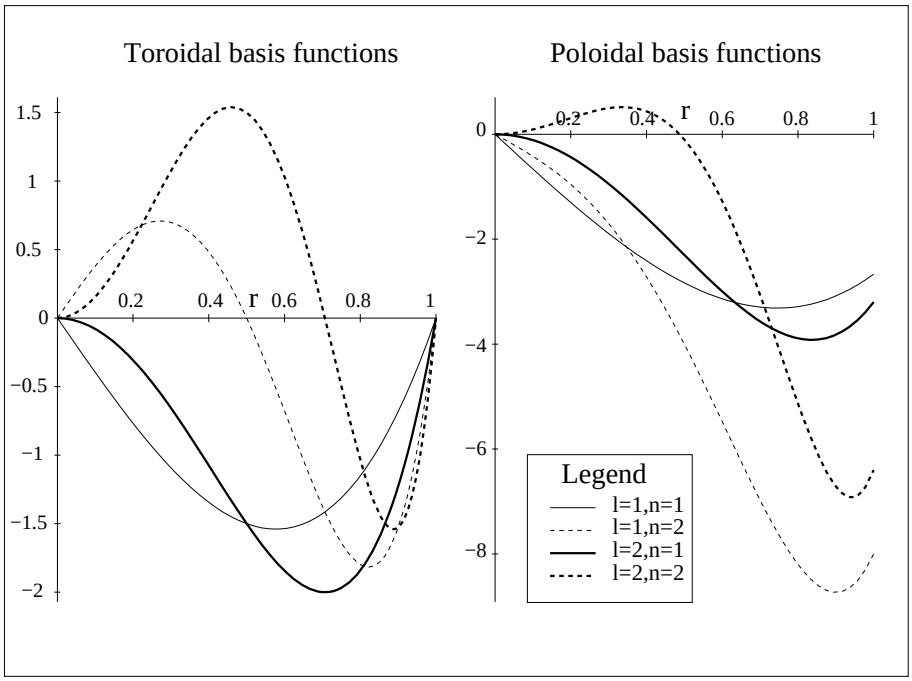
$$\begin{array}{ll} o_1 = \frac{(2n+2)^2}{2(l_\alpha+3)} & o_2 = \frac{(2n+2)^2+2l_\alpha+6}{2(l_\alpha+3)} \\ e_1 = \frac{(2n+2)^2+2l_\alpha+2}{2(l_\alpha+3)} & e_2 = \frac{(2n+2)^2-4}{2(l_\alpha+3)} \end{array} \tag{4.50}$$

Note the dependence on whether n is odd or even. This is due to the change in $T_{2n+2}(0) = (-1)^{n+1}$. The parity separation means that for a given radial truncation N_{max}, the maximal degree of polynomials used is $O(2N_{max})$ compared to $O(N_{max})$, as would be the case without such separation. Figure 4.1 shows the first four toroidal and poloidal basis functions on the domain $0 \leq r \leq 1$.

4.4.4 A sinusoidal basis

An alternative basis, motivated by *Zhang and Busse* (1988), is proposed below involving stretched Fourier series:

$$S_l^n(r) = r^{\Gamma(l)} sin(\nu_l^n r) \qquad T_l^n(r) = r^{\Gamma(l)} sin(n\pi r) \tag{4.51}$$



Explicit expressions			
l	n	T_l^n	S_l^n
1	1	$4r^2(r^2-1)$	$4r^2(r^2-\frac{5}{3})$
1	2	$4r^2(4r^4-5r^2+1)$	$4r^2(4r^4-5r^2-1)$
2	1	$8r^3(r^2-1)$	$8r^3(r^2-\frac{7}{5})$
2	2	$8r^3(4r^4-6r^2+2)$	$8r^3(4r^4-6r^2+\frac{6}{5})$

Figure 4.1: Toroidal and poloidal radial basis functions for $l = 1,2$; $n = 1,2$.

where $\Gamma(l) = 1$ if l is odd and 2 if even (to ensure the correct parity of the functions) and ν_l^n satisfies:

$$\tan(\nu_l^n) + \frac{\nu_l^n}{l+\Gamma(l)} = 0 \tag{4.52}$$

We let ν_l^n be the n^{th} positive root.

One might expect that Fourier series should converge exponentially fast; however, in the poloidal case above where each term is radially compressed by a factor ν_l^n rather than the usual $n\pi$, this is demonstrably false as shown later in section 4.7.1.

4.5 The matrix elements

4.5.1 Definition and choice of region

As has been outlined above, we use the matrix elements defined as

$$\mathsf{E}_{ij} = \int_V \mathbf{B}_i \cdot \mathsf{e}\mathbf{B}_j \, dV \qquad \mathsf{U}_{ij} = \int_V \mathbf{B}_i \cdot \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}_j) \, dV \tag{4.53}$$

$$[\mathsf{B}_{\mathscr{R}}]_{ij} = \int_{\mathscr{R}} \mathbf{B}_i \cdot \mathbf{B}_j \, dV \qquad [\mathsf{D}_{\mathscr{R}}]_{ij} = -\int_{\mathscr{R}} \mathbf{B}_i \cdot \boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}_j) \, dV \tag{4.54}$$

where in the cases of B and D the region $\mathscr{R}$ over which the elements are integrated is written explicitly. Since every such choice $\mathscr{R}$ must always include V, the only region in which $\mathbf{u}$ is non-zero, no such treatment is necessary for the matrices E and U. Unless stated we always use the Chebyshev basis since it has superior convergence properties.

In the case of the variational equations for the onset of magnetic energy instability, the region over which the integration is taken when computing the matrix elements is fixed: that of $V + \hat{V}$. There is some degree of flexibility however when formulating the Galerkin method for the induction equation. Any choice of domain $\mathscr{R} \supseteq V$ is a valid choice, since the magnetic field in $\hat{V}$ is completely determined by its behaviour on ∂V and the inclusion of the region V ensures that the velocity is fully represented. The calculation of the matrix elements of B and D are detailed in section 4.6.1.

Therefore every discretisation of the induction operator $\mathbf{L}$ is of the form

$$\mathsf{L} = \mathsf{B}_{\mathscr{R}}^{-1}(R_m \mathsf{U} + \mathsf{D}_{\mathscr{R}}) \tag{4.55}$$

There are infinitely many choices of L, all representing the same operator. This degeneracy might be regarded as that analogously obtained on changing the grid points in a finite difference or collocation method, clearly affecting the numerical results but nonetheless converging to the same solution. In general, there is an optimal grid that is chosen for such methods, and we might expect the same to be true in this case for an analogous choice of region. Table 4.2 shows the effect of the choice of region on the convergence to the $l = 1$ poloidal decay rate of $\lambda_P = -d_P^2 = -\pi^2$ using the Chebyshev basis. The value λ is the numerically computed eigenvalue for the choice of $\mathscr{R}$ for the decay problem. The error in the eigenvector is measured by

$$I(\mathscr{R}) = \int_V \left[\mathbf{B}_{\mathscr{R}} - \mathbf{B}_P\right]^2 dV \tag{4.56}$$

where $\mathbf{B}_P$ is the analytical decay mode and $\mathbf{B}_{\mathscr{R}}$ its numerical approximation using the region $\mathscr{R}$

in the Galerkin method. All fields are normalised such that the defining poloidal scalar attains 1 at $r = 1$.

	$\mathscr{R} = V$		$\mathscr{R} = V + \hat{V}$	
N_{max}	$\log_{10}\lvert\lambda - \lambda_P\rvert$	$\log_{10} I(V)$	$\log_{10}\lvert\lambda - \lambda_P\rvert$	$\log_{10} I(V+\hat{V})$
4	-3.4820	-8.3626	-7.9246	-8.6051
8	-10.9883	-23.6788	-12.0438	-24.3839
16	-11.2306	-24.9785	-11.4347	-24.1401

Table 4.2: Different choices of the region $\mathscr{R}$ lead to different convergence in the $l = 1$ poloidal decay rates of $\lambda_P = -d_P^2 = -\pi^2$. $I(\mathscr{R}) = \int_V \left[\mathbf{B}_{\mathscr{R}} - \mathbf{B}_P\right]^2 dV$ where $\mathbf{B}_{\mathscr{R}}$ is the numerical eigenvector using the Chebyshev basis, and $\mathbf{B}_P$ is the analytic decay mode. In all cases the fields are normalised so that the defining poloidal scalar attains a value of 1 at $r = 1$.

The error in the eigenvectors is not only very small, especially for $N_{max} = L_{max} \geq 8$, but remarkably similar for each choice of region. That this is so is not surprising since the true decay mode has only one projection onto the basis used, irrespective of which method is used to find it. This is only true when the projection is well defined; in the case of the sinusoidal basis this process is numerically ill-conditioned and many representations of a vector will almost agree leading to large errors.

The differences in the governing matrices are manifested in the differing convergence of the eigenvalues. Although both computations converge to a ceiling of around 10^{-11}, taking $\mathscr{R} = V + \hat{V}$ is seen to be vastly superior.

This is again seen in the non-quiescent flow case. We now compare the convergence using the different regions, in the most unstable eigenvalue of L as a function of truncation, for the case of the $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) flow, $\tau = 0.5$, $R_m = 55.0$ (defined in chapter five). The symmetry investigated is the critical one: $m = 1$, E^S; the eigenvalues are shown in table 4.3.

We also computed the integral $\int_V [\mathbf{B}_V - \mathbf{B}_{V+\hat{V}}]^2\, dV$, where $\mathbf{B}_V$ and $\mathbf{B}_{V+\hat{V}}$ are the eigenvectors from calculations with $\mathscr{R} = V$ and $\mathscr{R} = V + \hat{V}$ respectively, which measures the deviation in the eigenvectors for differing regions. For cases of $N_{max} = L_{max} \geq 16$ we found the integral to be at most 10^{-8}. This again means that the eigenmode has only one representation and any choice of region will give the same result. The eigenvectors were normalised so that $S_1^{1S}(1) = 1 + 0i$. The differing eigenvalues reflect the differing governing matrices.

In this case, as before, it is most expedient to choose the region $V + \hat{V}$ for swiftest convergence. We need only $N_{max} = L_{max} = 16$ to achieve convergence to five decimal places choosing $\mathscr{R} = V + \hat{V}$, whereas if we choose $\mathscr{R} = V$, we need more: $N_{max} = L_{max} = 32$ will give the same result.

We note in conclusion that any representation of $\mathbf{L}$ will work, although some better

$N_{max}=L_{max}$	λ_V	$\lambda_{V+\hat{V}}$
4	$0.4248+36.4035i$	$2.0087+33.1449i$
8	$-0.2482+35.0776i$	$0.04421+35.0348i$
16	$-0.002513+35.0741i$	$-0.001887+35.0749i$
32	$-0.001866+35.0750i$	$-0.001866+35.0750i$

Table 4.3: The most unstable eigenvalue of L as a function of truncation, for the different regions $\mathscr{R}=V$ and $\mathscr{R}=V+\hat{V}$, using the Chebyshev basis.

than others. Interestingly, to achieve the fastest convergence we choose $\mathscr{R}=V+\hat{V}$ and not the region available for finite difference and collocation methods of $\mathscr{R}=V$.

4.6 Computation of matrix elements

4.6.1 Calculation of Galerkin elements in $\hat{V}$

We have not touched upon how to compute elements of B and D in $\hat{V}$ but some clue has been given in section 4.1.2. It is a simple matter to extend **B** over $\hat{V}$ and so the elements of B follow easily. We may formulate the contribution to the diffusion matrix D_{ij} from $\hat{V}$ in one of two ways. Recall that the time dependence is governed in $\hat{V}$ by

$$\frac{\partial \mathbf{B}}{\partial t}=-\boldsymbol{\nabla}\times(\eta\boldsymbol{\nabla}\times\mathbf{B}) \tag{4.57}$$

where η is infinite in the well defined electrically insulating limit. The first method is to directly write the contribution as

$$-\int_{\hat{V}}\mathbf{B}_i\cdot\boldsymbol{\nabla}\times(\eta\boldsymbol{\nabla}\times\mathbf{B}_j)\,dV=-\oint_{r=1}\mathbf{B}_i\times(\boldsymbol{\nabla}\times\mathbf{B}_j)\cdot\mathbf{dS} \tag{4.58}$$

on taking the limit $\eta\to\infty$ which gives well defined matrix elements. The above equation follows directly from, for example, the analysis in section 3.9.7. Secondly, we may simply infer the behaviour of **B** in $\hat{V}$ from $r=1$. That is, for $r>1$,

$$-\boldsymbol{\nabla}\times(\eta\boldsymbol{\nabla}\times\mathbf{B})=\frac{\partial\mathbf{B}}{\partial t}(r)=\mathscr{E}\left(\frac{\partial\mathbf{B}}{\partial t}(1)\right)=\mathscr{E}\left(\nabla^2\mathbf{B}(1)\right) \tag{4.59}$$

where $\mathscr{E}$ is the potential field extension of a vector defined in V into the electrical insulator $\hat{V}$.

From sections 3.3.3 and 3.4.1 it follows that

$$\mathscr{E}\left(\nabla^2\mathbf{B}_\alpha(1)\right)=-l_\alpha\nabla_\alpha^2 S_\alpha(1)\boldsymbol{\nabla}\left(\frac{Y_\alpha}{r^{l_\alpha+1}}\right) \tag{4.60}$$

and so the contribution to $[\mathsf{D}]_{ij}$ from $\hat{V}$,where the index i refers to the dual index (α,n) and j to (β,m), is

$$\int_0^{2\pi}\int_0^{\pi}\int_1^{a}\mathbf{B}_\alpha^n\cdot\mathscr{E}\left(\nabla^2\mathbf{B}_\beta^m(1)\right)dr\sin\theta\,d\theta\,d\phi \tag{4.61}$$

$$=l_\alpha l_\beta\,\nabla_\beta^2 S_\beta^m(1)\,S_\alpha^n(1)\int_0^{2\pi}\int_0^{\pi}\int_1^{a}\boldsymbol{\nabla}\left(\frac{Y_\alpha}{r^{l_\alpha+1}}\right)\cdot\boldsymbol{\nabla}\left(\frac{Y_\beta}{r^{l_\beta+1}}\right)dr\sin\theta\,d\theta\,d\phi \tag{4.62}$$

$$=l_\alpha^2\,(l_\alpha+1)\,\nabla_\alpha^2 S_\alpha^m(1)\,S_\alpha^n(1)\,\delta_{\alpha\beta}\left(1-a^{-(2l_\alpha+1)}\right)N_\alpha^2 \tag{4.63}$$

$$\tag{4.64}$$

by a similar argument to that presented in B.3.3. In particular, if $a=\infty$ then the contribution from $\hat{V}$ to D is

$$l_\alpha^2\,(l_\alpha+1)\,\nabla_\alpha^2 S_\alpha^m(1)\,S_\alpha^n(1)\,N_\alpha^2\,\delta_{\alpha\beta} \tag{4.65}$$

4.6.2 Computation of elements in V

Unlike in $\hat{V}$, the functional form of the various matrix elements in V is complex. Although analytic formulae will exist, we do not attempt to compute the elements in this fashion; instead we use quadrature techniques that numerically integrate exactly if the elements behave in a certain way. In particular, to compute the integrals over solid angle, we use the vector harmonic transforms of *Lloyd and Gubbins* (1990) which is equivalent to using Gauss-Legendre quadrature in $\cos\theta$ and fast-Fourier transforms in ϕ. This method can be shown to compute the angular integrals exactly. If the velocity harmonic scalar functions are polynomials, then since the field scalars are also of this type, it follows that the radial behaviour is governed by a polynomial of some maximal degree (related to N_{max} and $\mathbf{u}$) which can be exactly treated by Gauss-Legendre quadrature. If the velocity scalars are sinusoidal for example, then we use increasing numbers of radial points until satisfactory convergence is achieved. In practice, we only need to compute the defining matrices once since they can then be saved; therefore computing on a fine grid is little hardship.

To handle the diffusive terms (noting that only the radial integration needs to be done), we use a Chebyshev transform although a radial quadrature method is equivalent. This is described in *Press et al.* (1992), *pg.149*.

It is interesting to note the relationship between the matrices U and E:

$$\mathsf{E}=\frac{1}{2}\left(\mathsf{U}+\mathsf{U}^T\right) \tag{4.66}$$

This follows easily from consideration of

$$\int_V \mathbf{B}_i \cdot \boldsymbol{\nabla} \times \left(\mathbf{u} \times \mathbf{B}_j \right) \, dV + \int_V \mathbf{B}_j \cdot \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}_i) \, dV \tag{4.67}$$

$$= 2 \int_V \mathbf{B}_i \cdot \mathsf{e} \mathbf{B}_j \, dV + \int_V \boldsymbol{\nabla} \cdot \left[(\mathbf{B}_i \cdot \mathbf{B}_j) \, \mathbf{u} \right] \, dV \tag{4.68}$$

similarly to section 3.5.2, and noting that the last integral vanishes since there is no normal flow at $r = 1$.

4.6.3 Checking the elements

To check the computations of the matrix elements, we independently evaluated some of them using the package *Maple* that uses computer algebra to compute them analytically. They were found to agree satisfactorily. In addition, the matrix E could be checked against $(\mathsf{U} + \mathsf{U}^T)/2$ and the matrices B and D used to solve the diffusion problem for which exact analytic solutions are known. We also benchmarked the Galerkin discretisation of the induction equation eigenvalue problem against existing finite difference codes and good agreement was found.

4.6.4 Solving the eigenvalue problem

Both the Galerkin form of the induction equation and the variational equations for maximal instantaneous magnetic energy growth are of generalised eigenvalue form, where for example

$$(R_m \mathsf{E} + \mathsf{D}_{V+\hat{V}}) \mathbf{v} = \lambda \mathsf{B}_{V+\hat{V}} \mathbf{v} \tag{4.69}$$

The matrices B and D are of block form due to orthogonality of the spherical harmonics:

$$\mathsf{B} = \begin{pmatrix} B_1 & & & & \\ & B_2 & & & \\ & & B_3 & & \\ & & & \ddots & \\ & & & & B_{NH} \end{pmatrix} \tag{4.70}$$

where NH is the number of harmonics and E (and similarly U) is in general dense. Instead of keeping the matrices of both sides of equation (4.69) in storage and finding the generalised eigenvalues, it is simpler to compute $\mathsf{C} = \mathsf{B}^{-1}_{V+\hat{V}} (R_m \mathsf{E} + \mathsf{D}_{V+\hat{V}})$ and find the usual eigenvalues instead. The matrix $\mathsf{B}^{-1}_{V+\hat{V}}$ is easy to compute since it is also in block form, with blocks being the inverses of those of $\mathsf{B}_{V+\hat{V}}$.

To find the eigenvalues, we use ARPACK routines[1] which use the implicitly restarted Arnoldi method to find a selected number of eigenvalues. These are very efficient at returning, for example, the n eigenmodes with the largest real part, where $n \ll N$, the dimension of C. Instead of solving the eigenproblem directly:

$$\mathsf{C}\mathbf{v} = \lambda\,\mathbf{v} \tag{4.71}$$

we instead choose to solve the transformed problem:

$$(\mathsf{C} - \sigma)^{-1}\mathbf{v} = \nu\,\mathbf{v} \tag{4.72}$$

where σ is a user specified real shift and $\nu = \frac{1}{\lambda-\sigma}$. The beauty of this approach is that the eigenvalues of largest magnitude in (4.72) correspond to those closest to σ in the untransformed problem (4.71). This is easiest seen in the relation $\lambda = \sigma + \frac{1}{\nu}$. Supposing then that we wanted to find the eigenvalue of C with the largest real part. We supply the Arnoldi routines with a large real shift, say $\sigma = 1000$, and find the eigenvalue with the largest magnitude of the transformed problem. This is simply related to the eigenvalue that we want.

Although the constituent matrices comprising C are symmetric, in general their product is not, so we cannot use a symmetric matrix eigenvalue solver in this case. Despite the problem being self-adjoint, symmetry of the underlying matrix does not necessary follow.

4.7 Convergence

4.7.1 The decay problem

The decay problem is a standard by which we can benchmark the code. Some results have already been presented, regarding the different choices of region over which we can choose to formulate the Galerkin method when applied to the induction equation. We now compute the convergence for three differing methods: two Galerkin formulations, using the Chebyshev and sinusoidal bases and a scheme using spherical harmonics in solid angle and finite differences in radius. Table 4.4 shows the convergence of the Chebyshev Galerkin scheme using $\mathscr{R} = V + \hat{V}$ to the $l = 1$ poloidal and toroidal decay rates as a function of N_{max} (we only include the S_1 harmonic in this case).

Both the poloidal and toroidal scalars converge exponentially to the analytic solution (i.e. increasing N_{max} from 2 to 4 at least halves the logarithmic error) until a numerical ceiling of

[1] See the website http://www.caam.rice.edu/software/ARPACK/

N_{max}	$log_{10}\|\lambda-\lambda_P\|$	$log_{10}\|\lambda-\lambda_T\|$
1	-0.1313	0.3635
2	-0.7304	-1.3049
3	-2.0015	-3.4629
4	-3.4820	-6.0471
5	-5.1513	-8.9645
6	-6.9721	-12.1272
7	-8.9212	-12.6786
8	-10.9820	-13.2733
16	-11.0705	-14.1484
32	-10.8967	-13.0023

Table 4.4: Absolute error of the $l=1$ poloidal and toroidal decay rates relative to their analytic values $\lambda_P=-d_P^2=-\pi^2$ and $\lambda_T=-d_T^2\approx-20.1907$ respectively, as a function of the radial truncation N_{max}. We use the Chebyshev Galerkin scheme with $\mathscr{R}=V+\hat{V}$.

10^{-11} and 10^{-13} respectively is reached. It is noteworthy that the poloidal scalar can never match the convergence of the toroidal scalar, why this is so is unclear. It most likely stems from the fact that the poloidal integrals are much larger in magnitude than those for the toroidal elements since there is an extra factor of $1/r^2$, allowing small numerical imprecision to creep in.

The diffusion problem, formulated using the sinusoidal Galerkin scheme over the regions V and $V+\hat{V}$ is shown in figure 4.2. The scheme exhibits exceptionally poor convergence

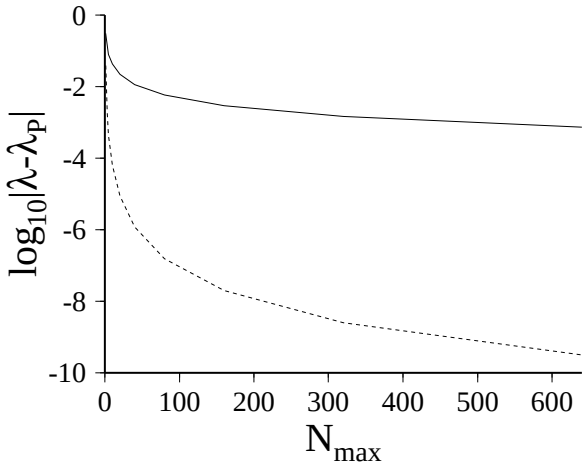


Figure 4.2: Comparison of convergence to $\lambda_P=-\pi^2$ for the $l=1$ poloidal decay problem for the Galerkin formulation with the sinusoidal basis. Solid is using $\mathscr{R}=V$, dashed is $\mathscr{R}=V+\hat{V}$.

when compared to the Chebyshev basis. Indeed, we need around 600 basis functions to get anywhere near the convergence using just 8 Chebyshev basis functions. The difference in the choice of region is also remarkable, and backs up a previous result that we must choose the region $V+\hat{V}$ (each of $\mathsf{B}_{V+\hat{V}}$ and $\mathsf{D}_{V+\hat{V}}$ being symmetric) to get the best convergence.

It is also noteworthy that not only do the eigenvalues differ in accuracy on changing the choice of region, but the eigenvectors do as well (in contrast with the Chebyshev result which shows little variation, section 4.5.1). This means that the projection of the decay mode onto this basis is very poorly conditioned: there exist many different representations of almost the same vector.

Lastly we present results of convergence of a finite difference calculation in radius, solving the diffusion problem in V. Again, we only solve for S_1 harmonic. Figure 4.3 shows the absolute error in the decay rate as a function of NR, the number of (equally spaced) grid points with three different orders of differencing scheme: solid is second order, short dashed is fourth order and long dashed is sixth order (*Gibbons*, 1998).

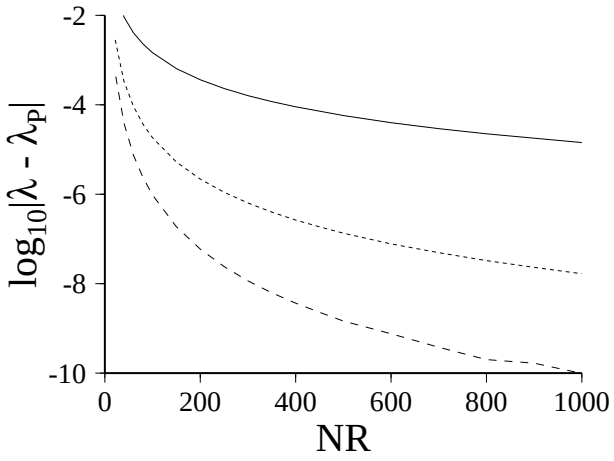


Figure 4.3: Comparison of convergence to $\lambda_P = -\pi^2$ for the $l = 1$ poloidal decay problem using a finite difference formulation. NR is the number of radial points on the uniform grid. The three lines show the order of the scheme: solid is second order, short-dashed is fourth order and long dashed is sixth order.

In order to get good convergence, not only do we need a sixth order scheme but also nearly 1000 grid points. This is in stark contrast to the Chebyshev Galerkin method.

4.7.2 Convergence in magnetic energy growth onset when $R_m \neq 0$

We now detail convergence in the case of the eigenvalue problem for λ_E, i.e. that of maximised instantaneous magnetic energy growth, using the Chebyshev basis. We use the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow with $\tau = 0.5$ (defined in chapter five). Table 4.5 shows λ_E for the $m = 0$ E^A field symmetry, for a range of different R_m. The convergence is exponential, at least for $R_m \leq 10^3$, since progressive increases in R_m by factors of 10 must be matched by additions of about 8 to $N_{max} = L_{max}$. At large truncations numerical errors begin to creep in, as demonstrated in the last significant fig-

ures for the 56 and 64 truncation levels. We must also check convergence of the eigenvectors

Truncation	λ_E for different values of R_m		
$N_{max} = L_{max}$	$R_m = 10$	$R_m = 10^2$	$R_m = 10^3$
4	-1.978674992	174.4142401	2120.712160
8	-1.920071338	198.7562174	2892.527679
16	**-1.920069544**	199.6333954	3277.361675
24	-1.920069544	**199.6333961**	3290.259497
32	-1.920069544	199.6333961	**3290.271507**
40	-1.920069544	199.6333961	3290.271507
48	-1.920069544	199.6333961	3290.271507
56	-1.920069544	199.6333961	3290.271509
64	-1.920069573	199.6333966	3290.271519

Table 4.5: Values of λ_E for the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$ with field of $m = 0$ E^A symmetry, $R_m = 10, 10^2, 10^3$. We use the Chebyshev basis; bold indicates convergence has been achieved.

however; this is shown in table 4.6 which shows the equivalent convergence in rms over the unit sphere, with normalisation chosen such that the harmonic component S_1 attains the value 1 on $r = 1$. The convergence in the eigenvectors matches that for the eigenvalues, perhaps not surprisingly because the problem is described by symmetric matrices and so is self-adjoint. We obtain converged solutions even up to $R_m = 1000$ with only $N_{max} = L_{max} = 32$.

Truncation	$\mathbf{B}_{rms}$ for different values of R_m		
$N_{max} = L_{max}$	$R_m = 10$	$R_m = 10^2$	$R_m = 10^3$
4	3.716996525	7.446646010	9.814727629
8	3.730705373	11.406394715	46.58724052
16	**3.730705866**	11.692283662	159.9496714
24	3.730705866	**11.692283920**	172.7132574
32	3.730705866	11.692283919	**172.7295699**
40	3.730705866	11.692283919	172.7295700
48	3.730705866	11.692283924	172.7295720
56	3.730705865	11.692283791	172.7296738
64	3.730705873	11.692284666	172.7324848

Table 4.6: Spatial rms measure of the size of the eigenvector corresponding to λ_E, $\mathbf{B}_{rms} = \sqrt{\frac{3}{4\pi}\int_V |\mathbf{B}|^2\, dV}$, normalised so that the field component $S_1(1) = 1$, for the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$ with field of $m = 0$ E^A symmetry. We use the Chebyshev basis; bold indicates convergence has been achieved.

Another way of visualising this convergence is to plot a power spectrum. Figure 4.4 shows $M_l = \frac{1}{2}\int_{V+\hat{V}} \mathbf{B}_l^2\, dV$ where $\mathbf{B}_l$ is the field containing only spherical harmonics of degree l, as a function of l for the poloidal (solid) and toroidal (dashed) components. The field is computed with $N_{max} = L_{max} = 64$, $R_m = 10^3$ and normalised such that M_l sum to 1. Both curves

bottom out at $l = 32$ indicating that this is the minimum truncation necessary for convergence, in agreement with tables 4.5 and 4.6.

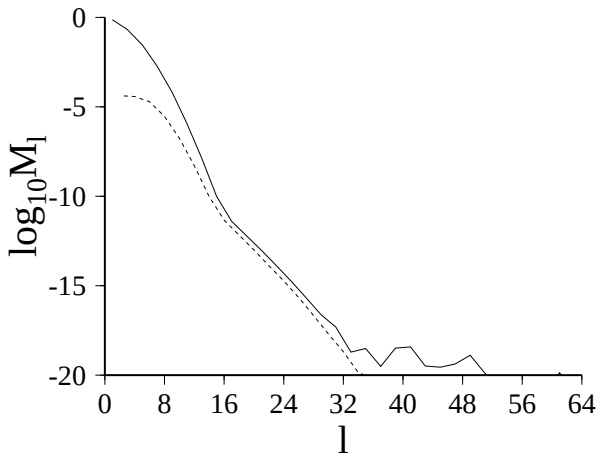


Figure 4.4: Power spectrum of the field associated with λ_E for $R_m = 10^3$ and $N_{max} = L_{max} = 64$, for the $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) flow, $\tau = 0.5$. The curves bottom out at $l = 32$ indicating that this is the minimum truncation necessary for convergence, in agreement with tables 4.5 and 4.6.

4.7.3 Convergence in the induction equation eigenvalues when $R_m \neq 0$

We now detail convergence in the case of the induction equation eigenvalue problem for λ. We use the same $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow with $\tau = 0.5$ for comparison with the previous section. The region of integration for the Chebyshev Galerkin scheme is V, since although this does not show the best convergence, we may directly compare the results with those of other schemes in the literature. Table 4.7 shows the eigenvalue λ for the $m = 0$ E^A field symmetry, for a range of different R_m. The eigenvalues are all real unless explicitly written otherwise.

Again we also check convergence of the eigenvectors; this is shown in table 4.8 which shows the associated rms values over the unit sphere, with normalisation chosen such that the harmonic component S_1 attains the value $1 + 0i$ on $r = 1$.

Convergence in the eigenvalues and eigenvectors is much poorer than that of the previous section. We only obtain converged solutions for $R_m = 10$; the cutoff between unconverged and converged solutions is somewhere in $10 < R_m \leq 100$. It is noteworthy that both the eigenvalues and eigenvectors converge at the same rate: we find the solution at $R_m = 10$ is fully converged with $N_{max} = L_{max} = 24$. This is at odds with previous work, for example, *Gibson and Roberts* (1967), who report that the eigenvalues tend to converge faster than the eigenvectors. This could be an artifact of either truncation or their numerical method used. We find no

Truncation	λ for different values of R_m		
$N_{max} = L_{max}$	$R_m = 10$	$R_m = 10^2$	$R_m = 10^3$
4	-10.64748350	-9.917964773 + 251.6136792i	-33.65136539
8	-8.182616767	-30.96805134	-33.71633508
16	-8.016049667	-19.59753494	-32.43089239
24	**-8.016002465**	-8.1361069667	-31.93038594
32	-8.016002462	-6.9701786935	-31.78237276
40	-8.016002463	-6.9294930685	-31.72673360
48	-8.016002462	-6.9288540530	-31.73916495
56	-8.016002469	-6.9168865200	-31.72331091
64	-8.016002694	-6.9872496777	-31.72243873

Table 4.7: Values of λ for the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$ with field of $m = 0$ E^A symmetry, $R_m = 10, 10^2, 10^3$. We use the Chebyshev basis; bold indicates convergence has been achieved. Computed with the discretisation of $\mathbf{L}$ over $\mathscr{R} = V$.

Truncation	$\mathbf{B}_{rms}$ for different values of R_m		
$N_{max} = L_{max}$	$R_m = 10$	$R_m = 10^2$	$R_m = 10^3$
4	5.902958827	26.84085496	1.555194495
8	5.492042537	1.604959010	1.560766045
16	5.467587247	18.70823767	1.599028128
24	**5.467581088**	17.08907330	1.618530117
32	5.467581088	17.04782017	1.624227421
40	5.467581088	17.04690463	1.625880546
48	5.467581088	17.04689382	1.626355927
56	5.467581091	17.04063620	1.626486797
64	5.467581167	17.05841207	1.626520113

Table 4.8: Spatial rms measure of the size of the eigenvector corresponding to λ, $\mathbf{B}_{rms} = \sqrt{\frac{3}{4\pi}\int_V |\mathbf{B}|^2\, dV}$, normalised so that the field component $S_1(1) = 1 + 0i$, for the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$ with field of $m = 0$ E^A symmetry. Bold indicates convergence has been achieved. Computed with the discretisation of $\mathbf{L}$ over $\mathscr{R} = V$.

evidence to substantiate such a claim here using a fully spectral Galerkin method.

Tables 4.9 and 4.10 show the equivalent data for the induction operator $\mathbf{L}$ discretised over $V + \hat{V}$, in all cases showing much improved convergence. Well converged solutions are found for $R_m = 10$ at $N_{max} = L_{max} = 16$, compared to the truncation level of 24 required for $\mathscr{R} = V$. Most of the previous literature in kinematic dynamo theory uses finite difference schemes in radius, for example *Gubbins et al.* (2000a), who employed truncation levels of up to $L_{max} = 18$ with 50 radial points. Our calculations suggest that, depending in the convergence required, that $N_{max} = L_{max} = 24$ is a reasonable truncation level and highlights the difficulty in finding properly converged eigenvalues, an issue that has plagued the subject since its conception. Many authors, for example, *Bullard and Gellman* (1954) and *Lilley* (1970) found numerical results that were

later found to be spurious, owing to insufficient numerical resolution.

Truncation	λ for different values of R_m		
$N_{max} = L_{max}$	$R_m = 10$	$R_m = 10^2$	$R_m = 10^3$
4	-8.011955933	-6.164763605	-33.65136539
8	-8.015997179	-6.826803514	-33.71633508
16	**-8.016002462**	-6.928777793	-32.43089239
24	-8.016002462	-6.928848707	-31.93038594
32	-8.016002461	-6.928848717	-31.78237276
40	-8.016002459	-6.928848321	-31.73916494
48	-8.016002464	-6.928849576	-31.72673360
56	-8.016002428	-6.921928612	-31.72331091
64	-8.016002755	-6.955875275	-31.72244082

Table 4.9: Values of λ for the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$ with field of $m = 0\ E^A$ symmetry, $R_m = 10, 10^2, 10^3$. We use the Chebyshev basis; bold indicates convergence has been achieved. Computed with the discretisation of $\mathbf{L}$ over $\mathscr{R} = V + \hat{V}$.

Truncation	$\mathbf{B}_{rms}$ for different values of R_m		
$N_{max} = L_{max}$	$R_m = 10$	$R_m = 10^2$	$R_m = 10^3$
4	5.459553693	16.81626987	1.626486796
8	5.467576745	16.09726939	1.560766045
16	**5.467581088**	17.03662230	1.599028128
24	5.467581088	17.04687639	1.618530117
32	5.467581088	17.04689369	1.624227421
40	5.467581087	17.04689364	1.625880546
48	5.467581088	17.04689386	1.626355927
56	5.467581072	17.04293268	1.626486797
64	5.467581189	17.05447682	1.626520070

Table 4.10: Spatial rms measure of the size of the eigenvector corresponding to λ, $\mathbf{B}_{rms} = \sqrt{\frac{3}{4\pi}\int_V |\mathbf{B}|^2\, dV}$, normalised so that the field component $S_1(1) = 1 + 0i$, for the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$ with field of $m = 0\ E^A$ symmetry. Bold indicates convergence has been achieved. Computed with the discretisation of $\mathbf{L}$ over $\mathscr{R} = V + \hat{V}$.

4.8 Maximal finite time growth

In section 3.9.5 we saw theoretically how we might go about computing maximal relative finite time magnetic energy growth. We present the numerical details here. Suppose then that we would like to evaluate, for some time t

$$\max_{\mathbf{B}(0)} \frac{M(t)}{M(0)} \tag{4.73}$$

where the maximisation is taken over all initial fields $\mathbf{B}(0)$. We discretise the induction operator $\mathbf{L}$ over all space, forming L, and so we may write the exact solution of

$$\frac{\partial \mathbf{B}}{\partial t} = \mathbf{L}\mathbf{B} \tag{4.74}$$

as

$$\mathbf{v}(t) = e^{\mathsf{L}t}\mathbf{v}(0) \tag{4.75}$$

where $\mathbf{v}$ represents the field $\mathbf{B}$ relative to the Chebyshev basis used. We may then rewrite the maximisation problem as

$$\max_{\mathbf{B}(0)} \frac{\int_{V+\hat{V}} \mathbf{B}(t)^2\, dV}{\int_{V+\hat{V}} \mathbf{B}(0)^2\, dV} = \max_{\mathbf{v}(0)} \frac{\mathbf{v}(t)^T \mathsf{B}_{V+\hat{V}} \mathbf{v}(t)}{\mathbf{v}(0)^T \mathsf{B}_{V+\hat{V}} \mathbf{v}(0)} = \frac{\left[e^{\mathsf{L}t}\mathbf{v}(0)\right]^T B_{V+\hat{V}} e^{\mathsf{L}t}\mathbf{v}(0)}{\mathbf{v}(0)^T \mathsf{B}_{V+\hat{V}} \mathbf{v}(0)} \tag{4.76}$$

recalling that the matrix $\mathsf{B}_{V+\hat{V}}$ is the basis energy inner product matrix. Since this is symmetric and positive definite, we may write it as $\mathsf{B}_{V+\hat{V}} = \mathsf{F}^T\mathsf{F}$ using the Cholesky decomposition (see *Press et al.*, 1992). Equation (4.76) then becomes

$$\max_{\mathbf{v}(0)} \frac{|\mathsf{F}\, e^{\mathsf{L}t}\mathbf{v}(0)|^2}{|\mathsf{F}\,\mathbf{v}(0)|^2} \tag{4.77}$$

where $|\quad|^2$ is the usual Eulerian 2-norm. Writing $\mathbf{u} = \mathsf{F}\mathbf{v}(0)$ this becomes

$$\max_{\mathbf{u}} \frac{|\mathsf{F}\, e^{\mathsf{L}t}\,\mathsf{F}^{-1}\mathbf{u}|^2}{|\,\mathbf{u}|^2} = \|\mathsf{F}\, e^{\mathsf{L}t}\,\mathsf{F}^{-1}\|_2^2 \tag{4.78}$$

where $\|\quad\|_2^2$ is the Eulerian 2-norm of a matrix, which may be effectively calculated using the singular value decomposition.

In contrast with computations for λ_E, we do not include convergence results for maximal finite time growth calculations here since they depend on the particular flow studied and no generic behaviour was found; the reader is referred instead to chapter six where all the relevant results are presented.

4.8.1 Finding the norm using Singular value decomposition

We may write any $n \times n$ matrix A as

$$\mathsf{A} = \mathsf{U}\,\mathsf{W}\,\mathsf{V}^T \tag{4.79}$$

where U and V are orthogonal $n \times n$ matrices (that is, their inverses are equal to their transposes) and $\mathsf{W} = diag(w_1, \ldots, w_n)$, the singular values of A. This may be rewritten

$$\mathsf{A}\,\mathsf{V} = \mathsf{U}\,\mathsf{W} \tag{4.80}$$

so that the i^{th} column vectors of V and U satisfy

$$\mathsf{A}\,\mathbf{v}_i = w_i\,\mathbf{u}_i \tag{4.81}$$

We may expand a vector $\mathbf{v}$ in terms of the columns of V, being an orthogonal spanning set:

$$\mathbf{v} = \sum_{i=1}^{N} \alpha_i \mathbf{v}_i \tag{4.82}$$

so that

$$\mathsf{A}\,\mathbf{v} = \sum_{i=1}^{N} \alpha_i\, w_i\, \mathbf{u}_i \tag{4.83}$$

Then by the orthogonality of $\mathbf{u}_i$ and $\mathbf{v}_i$:

$$\begin{aligned} \|A\|_2^2 &= \max_{\alpha_i} \frac{\sum_{i=1}^{N} (\alpha_i\, w_i)^2}{\sum_{i=1}^{N} \alpha_i^2} \\ &= (\max_i \{w_i\})^2 \end{aligned} \tag{4.84}$$

attained when $\mathbf{v} = \mathbf{v}_i$, or any scalar multiple of it. If A has orthogonal eigenvectors, then the singular values match the eigenvalues. A particular case of this arises if we compute $\mathsf{A}^T\mathsf{A}$, and apply equation (4.79); this may be written

$$\mathsf{A}^T\mathsf{A} = \mathsf{V}\mathsf{W}\mathsf{U}^T\mathsf{U}\mathsf{W}\mathsf{V}^T = \mathsf{V}\mathsf{W}^2\mathsf{V}^T \tag{4.85}$$

and so has singular values which are the squares of those of A. Therefore the maximum eigenvalue of $\mathsf{A}^T\mathsf{A}$ and the square of the largest singular value of A give the sought maximum, and offer two independent means of computation.

4.8.2 The matrix exponential

In order to compute the maximum relative magnetic energy growth over some time t, we need to be able to compute the matrix exponential of L. Although we may make sense of this using a

familiar series expansion:

$$e^{\mathsf{L}t} = \sum_{i=0} (\mathsf{L}t)^i / i! \qquad , \tag{4.86}$$

actually computing it this way is highly inefficient and convergence is slow.

An alternative method to compute $e^{\mathsf{L}t}$, popular in the literature discussed in section 5.1, is to project L onto its N most unstable eigenvectors and to use a truncated version of the identity:

$$e^{\mathsf{L}t} = \sum_{n=0} \left(\mathsf{V}\Lambda\mathsf{V}^{-1}\right)^n / n! = \sum_{n=0} \mathsf{V}\Lambda^n\mathsf{V}^{-1}/n! = \mathsf{V}e^{\Lambda t}\mathsf{V}^{-1} \tag{4.87}$$

where $\mathsf{L} = \mathsf{V}\Lambda\mathsf{V}^{-1}$, the matrix V comprising the eigenvectors of L as columns, and Λ being the diagonal matrix of the eigenvalues of L. The beauty of this methodology is that $e^{\Lambda t}$ is trivial to compute, being itself diagonal with entries $[e^{\Lambda t}]_{ii} = e^{\Lambda_{ii} t}$. The idea is to increase the number of eigenvectors involved in the truncation and hope that the computations converge. Unfortunately in our case this does not occur, seemingly because more eigenvectors are needed in the expansion than for which eigenvector convergence is possible. Using very high spectral resolution does not seem to help, increasing only the numerical error and creating spurious eigenvalues. Recall that historically it has been hard enough to prove the convergence of a single eigenvalue in kinematic dynamo computations; in this case we would typically require 20.

Instead we use the Padé approximation of order 6 with scaling and squaring (see *Molar and Van Loan*, 1978), identical to that used by the package *Matlab*.

In order to test the method, we used various different time-stepping schemes to find the time-propagator matrix, which acts upon a field and evolves it to time t later on. This is equivalent to the matrix $e^{\mathsf{L}t}$, and excellent agreement in equation (4.78) was found as the number of time steps was increased. We do not use such methods since they present yet another parameter on which the convergence of the solution would depend. Although the order of the Padé approximation is adjustable, we find that the usual value of 6 is sufficient for all of our calculations.

4.9 Using finite time growth for the onset of magnetic energy growth

We may use the initial gradient of the maximised finite time magnetic energy growth calculations of equation 4.73 (which in chapter six will be referred to as 'envelopes') to compute the maximum instantaneous magnetic energy growth rate λ_E. Recall that

$$\lambda_E = \max_{\mathbf{B}} \frac{1}{2M}\frac{dM}{dt} \tag{4.88}$$

so that $2\lambda_E$ is the initial envelope gradient. Such calculations are a useful check on our numerical procedures since the techniques involved are very different from the direct eigenvalue methods associated with the variational approach. We may also use such a method to find R_m^E, by finding the value of R_m for which the envelope initially has zero slope. Table 4.11 shows values of R_m^E evaluated using the two methods, for the $m = 0\ E^A$ field symmetry with $N_{max} = L_{max} = 16$. The values of R_m^E agree well enough for us to be sure that they converge to the same solution.

Flow	R_m^E (Variational)	R_m^E (Envelope)
$\mathbf{s}_2$(MDJ), $\tau = 1/2$	6.1240425	6.1240446
$\mathbf{t}_1\,\mathbf{s}_2$(MDJ), $\tau = 1/2$	11.3957947	11.3957882
$\mathbf{t}_1\,\mathbf{s}_2$(KR), $\tau = 1/2$	12.8193697	12.8193652

Table 4.11: A comparison in the value of R_m^E using the two different methods. Computed for the $m = 0\ E^A$ field symmetries, $N_{max} = L_{max} = 16$. The envelope gradient calculations were carried out using the differences between $t = 0$ and $t = 10^{-10}$.

However, the envelope method is numerically more inaccurate than the direct variational method, for when $t \ll 1$ we must compare the value of $1+\varepsilon$, where $(0 < \varepsilon \ll 1)$, with its value of 1 at $t = 0$ to see if it initially increases. The problem is that as $t \to 0$, $\varepsilon \to 0$ also, possibly more rapidly than t, and finite machine precision becomes important when comparing two very similar numbers.

4.10 The choice of domain for computations of λ_E

If we were to study the onset problem of flow energy growth in a fluid dynamical case, the choice of the region over which to define the energy is natural, since the fluid only exists in a finite region. Choosing to define the energy over a subset of this is also possible but not physically very intuitive. In the magnetic case, although the field in the external insulator depends completely on the field at the boundary, we may choose the energy to be defined over the region $0 < r < a$ where $a \geq 1$. There are infinitely many such choices of a, which one do we choose?

Any reasonable choice of region would reflect what is going on; for example, if $R_m = 0$ one would expect that $\lambda_E < 0$. This diffusion problem should also converge easily, for if it does not, then we will expect extreme difficulty seeking solutions for which $R_m > 0$.

For instantaneous instability, we are forced into choosing the region over which the diffusion operator is self-adjoint, since any other measure does not capture this physical process. We illustrate this idea with the study of a simple 1D model.

4.10.1 Simple 1D model

Consider the simple 1D equation for $u = u(x,t)$ already introduced in section 4.1:

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \tag{4.89}$$

subject to the boundary conditions

$$\frac{\partial u}{\partial x} + \frac{u}{x} = 0 \qquad x \geq 1, t \geq 0 \tag{4.90}$$

$$u(0,t) = 0 \tag{4.91}$$

with $u(x,t)$ and its first order spatial derivatives continuous everywhere.

4.10.2 Energetic instability

We define the energy of the solution to be

$$M = \frac{1}{2}\int_0^a u^2\, dx \tag{4.92}$$

and

$$\lambda_E = \max_{u(x,0)} \frac{1}{2M}\frac{dM}{dt} \tag{4.93}$$

We use a variational principle to locate the maximum by multiplying equation (4.89) by u, integrating and taking the first variation:

$$2\lambda_E \int_0^a \delta u\, u\, dx = \int_0^a \left\{ \frac{\partial u}{\partial t}\,\delta u + \frac{\partial \delta u}{\partial t}\, u \right\} dx \tag{4.94}$$

which, on employing a Galerkin method with the basis U_n (section 4.1.1), is of generalised eigenvalue form.

4.10.3 Self-adjoint case

When $a = 1$ the equation becomes self-adjoint. This is because

$$\int_0^1 \frac{\partial^2 u}{\partial x^2}\,\delta u\, dx = \left[\frac{\partial u}{\partial x}\,\delta u(x)\right]_0^1 - \int_0^1 \frac{\partial u}{\partial x}\frac{\partial \delta u}{\partial x}\, dx = -u(1)\,\delta u(1) - \int_0^1 \frac{\partial u}{\partial x}\frac{\partial \delta u}{\partial x}\, dx \tag{4.95}$$

which is symmetric. In this case, the normal modes are orthogonal, and the energy stability becomes identical to that of the usual linear eigenvalue case. That the normal modes are orthogonal

is demonstrably true:

$$\int_0^1 \sin(\nu_n x)\ \sin(\nu_m x)\ dx = \frac{-\nu_m \tan\nu_n + \nu_n \tan\nu_m}{(\nu_n^2 - \nu_m^2)\cos\nu_n \cos\nu_m} \tag{4.96}$$

which vanishes by definition of ν if ν_n and ν_m are different. It is non zero if they are the same.

4.10.4 Other choices of a

Table 4.12 shows values of λ_E for various values of a and N, the number of odd basis functions used. The value k is the matrix condition number (for $N = 8$) which is the product of the smallest and largest singular values. It gives a measure of the numerical difficulty in computing λ_E. We used *Maple*, able to compute the matrix elements exactly using computer algebra to an accuracy of 50 decimal places, which in the cases where k was extremely high was necessary to get a converged solution. The only converged results are for $a = 1$ for which the decay rate

	N				
a	2	4	6	8	k
0.2	272.72	4883.83	24628.86	76957.10	10^{24}
0.4	57.14	1141.57	5896.08	18624.64	10^{21}
0.6	16.00	433.33	2372.03	7686.86	10^{16}
0.8	0.49	153.52	1012.53	3537.94	10^{12}
1	**-0.41**	**-0.41**	**-0.41**	**-0.41**	10^{6}
2	0.02	103.98	805.04	3267.90	10^{7}
6	2.92	167.53	1181.18	4618.38	10^{7}
10	3.47	178.93	1247.44	4854.93	10^{7}
100	4.19	193.77	1333.34	5161.23	10^{7}

Table 4.12: Values of λ_E for different choices of a as a function of N, the number of basis functions used, restricted to those which are of odd parity since the sought solution is odd. k is the condition number (for $N = 8$). Bold represents converged solutions.

$\nu_1 = -0.41$ is recovered; all other choices give λ_E as a strictly increasing function of truncation. So what is going on, and why aren't we getting converged solutions for $a \neq 1$? To answer this, figure 4.5 shows the solution maximising the instantaneous growth for $a = 0.4$, $N = 8$. Note the scale disparity for the inset graph. The discontinuous solution it is trying to converge to appears to be one which vanishes in the region $0 \leq x \leq a$ and is non zero elsewhere. These results may be easily understood by considering diffusion. A solution that is zero in a region and large nearby will experience large instantaneous growth in the region as the solution diffuses into it. Note that overall the solution may be reducing in magnitude over time, but the region considered will experience a large initial growth. As a tends to 1 from below, the boundary conditions become increasingly important. If u is zero in $[0, a]$ where $1 - a \ll 1$, the solution

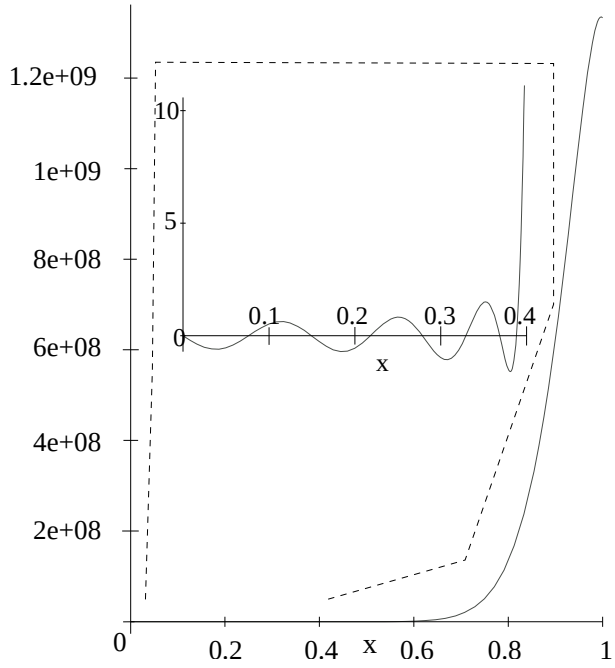


Figure 4.5: The solution maximising $\lambda_E = 18624.64$ for $N = 8$ using the energy measure for $a = 0.4$. The solution has energy $M = 1/2$ in the region $0 \leq x \leq a$.

will be almost zero everywhere since the internal solution is communicated linearly through the boundary. This is clearly not the optimal solution since almost no diffusion will occur; the most unstable configuration when $a = 1$ is u non-zero in $x > 1$, decaying in line with the normal modes.

The discontinuous solutions for $a \neq 1$ live outside the space of functions realisable in our Chebyshev basis which is continuously differentiable. Perhaps numerically a better method would be a finite difference scheme, which would allow this kind of solution to be well approximated. Unfortunately in this Galerkin paradigm, we are not at liberty to use such a tool.

4.10.5 Relevance

When defining a region over which we measure energetic instability, we must choose it such that the diffusion operator is self-adjoint, otherwise computations do not converge to a physically meaningful value. In the case here, we must choose the region $[0,1]$. In the case of the 3D kinematic dynamo, the required region is $V + \hat{V}$. This is demonstrated in table 4.13 where we demonstrate the behaviour of λ_E as a function of region chosen, using only the S_1 harmonic. Only the choice $a = \infty$, corresponding to the region $V + \hat{V}$, converges. Indeed, it converges to the physically sensible (negative) normal mode value of $-\pi^2$. The methodology for producing these results is as follows. Referring back to section 4.6.1 we see that the contribution to the

	N_{max}		
a	16	24	32
1	6.4023	28.1099	60.1305
1.25	-6.8425	-3.1094	2.3106
∞	-9.8696	-9.8696	-9.8696

Table 4.13: Convergence in λ_E for the S_1 harmonic in N_{max}, for different choices of region.

matrix elements of $\mathsf{B}_{V+\hat{V}}$ in $\hat{V}$ are multiplied by a factor $(1-a^{-3})$ when $l=1$. Thus we may compute the matrix B_a of the energy inner products of the basis in $0 \leq r \leq a$ where $a \geq 1$, by simply forming $\mathsf{B}_V + (1-a^{-3})\left[\mathsf{B}_{V+\hat{V}} - \mathsf{B}_V\right]$.

The value of λ_E associated with any region is the maximum eigenvalue of the matrix $[\mathsf{L}+\mathsf{L}^\dagger]/2$ (as follows from section 3.9.4) where we choose (arbitrarily) to form L by discretising over $V+\hat{V}$. The adjoint matrix $\mathsf{L}^\dagger$ may be formed by consideration of its definition:

$$\int_{0\leq r\leq a} \mathbf{B}_1 \cdot \mathbf{L}\,\mathbf{B}_2\, dV = \int_{0\leq r\leq a} \mathbf{L}^\dagger \mathbf{B}_1 \cdot \mathbf{B}_2\, dV \tag{4.97}$$

Expressing $\mathbf{B}_1$ relative to the Chebyshev basis as $\mathbf{v}_1$ and similarly for $\mathbf{B}_2$, this can be written:

$$\mathbf{v}_1^T\, \mathsf{B}_a\, \mathsf{L}\, \mathbf{v}_2 = \left[\mathsf{L}^\dagger \mathbf{v}_1\right]^T\, \mathsf{B}_a\, \mathbf{v}_2 \tag{4.98}$$

This this must hold for all vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, it follows that

$$\mathsf{L}^\dagger = \mathsf{B}_a^{-1}\, \mathsf{L}^T\, \mathsf{B}_a \tag{4.99}$$

Chapter 5

The onset of magnetic energy growth

5.1 Motivation for non-eigenmode analysis

The motivation for the study of so called non-modal (non-eigenmode) analysis of linear systems came about in the effort to explain the hydrodynamical phenomenon of transition to turbulence. In this non-magnetic case, if a flow is driven hard enough in a steady basic state, growing perturbations emerge which push the system into a non-linear regime. The numerical approach relating to this was to linearise the Navier-Stokes equations about some steady state and to look for growing instabilities. If one is found that grows sufficiently large in magnitude, then the initial linearisation assumption breaks down and non-linear effects become important. Historically workers have solved this stability problem using an eigenmode formulation, that is, $\mathbf{u} \propto e^{\lambda t}$ where $\mathbf{u}$ is the velocity perturbation. In certain cases this was very successful, and in others, disastrous as table 5.1 shows. The three flows listed here are: Rayleigh-Bénard convection (cellular convection driven from bottom heating in an infinite plane layer), plane Couette flow (with a linear velocity profile between two infinite plates moving parallel to one another) and plane Poiseuille flow (with a parabolic flow profile in a channel with parallel sides). In the case of convection with a quiescent base state, the important parameter is the non-dimensional vigour of heating, measured by the Rayleigh number R_a. If R_a is large enough, the fluid will begin to move in a convection roll formation, a more efficient mechanism of transferring heat than static conduction. For the planar flows which are linearised about a steady but non-zero base state, the controlling parameter is the non-dimensional flow speed: the kinetic Reynolds number, R_e. The threshold R_c is the critical value of R (either R_e or R_a depending on the situation), above which a growing eigenmode perturbation to the flow will grow. R^E is the energetic critical value, below which all disturbances decay in energy, but transient growth may occur if $R^E < R < R^c$. Lastly,

R^T denotes that found in experiments and signifies transition from the base state (see *Schmid and Henningson*, 2001). It is striking that the above three separate analyses agree so well in

	Rayleigh-Bénard convection (R_a)	Plane Couette flow (R_e)	Plane Poiseuille flow (R_e)
R^c	1708	∞	5772
R^E	1708	20.7	49.6
R^T	1710 ± 10	350	1000

Table 5.1: The critical linear R^c, energetic R^E and experimental R^T thresholds on R, denoting the Rayleigh number R_a or the kinetic Reynolds number R_e depending on the flow. From *Schmid and Henningson* (2001).

Rayleigh-Bénard convection, and differ so much in the other two cases. Plane Couette flow is eigenmode or linearly stable for all values of R_e, in huge disagreement with experimental results. In fact it seems that the energetic analysis, which signals the onset of growth (at least instantaneously before possible decay thereafter) is a better indicator of the physical behaviour, despite being an order of magnitude too low. In both planar flow cases, even if all the eigenmodes decay (i.e. the system is linearly subcritical) it is still possible to get significant transiently growing perturbations (that eventually decay).

The disagreement between eigenvalue and energetic analysis can be traced to the underlying operator's non self-adjointness. In the literature this is also known as 'non-normality', and signifies non-orthogonal eigenvectors. In a typical system, an initial condition may be represented as the superposition of nearly cancelling eigenvectors. The period of scientific interest may be then described by how the pattern of cancellation evolves and not of the behaviour of the individual eigenmodes. Indeed, in strongly non-normal cases it may be argued that there is no good scientific reason to study eigenmode growth (*Trefethen*, 1997). In this fashion, even if each eigenmode individually decays the energy may transiently grow. In the geophysical case, we are not limited so much by the timescale but on whether or not the magnetic field grows large enough so that the Lorentz force becomes significant; if such non-eigenmode growth can occur then it is clearly of interest to quantify it.

Non-normal theory has enjoyed great success when applied to the transition to turbulence in the Navier-Stokes equations, although alternative mechanisms have been proposed (*Waleffe*, 1995). In particular, it has correctly predicted stream-wise streaks and vortices (*Chapman*, 2002) seen experimentally in both the planar flows described above. In MHD, using the idea of the pseudospectrum (see below) the resistive Alfvén paradox has been resolved, where in the limit of vanishing magnetic diffusivity, the Alfvén velocity spectrum appeared to be discontinuous (*Borba et al.*, 1994). In a study of the Hartmann layer, *Gerard-Varet* (2002) showed that

significant subcritical perturbation growth can occur, leading to dynamics not predicted by eigenmode analysis. Energetic instability of magnetic fields has already been studied by *Farrell and Ioannou* (1999a) in a cylindrical geometry. They find possibilities for large sub-critical transient behaviour, of many orders of magnitude. Of particular relevance for the Earth is their discovery of asymptotically unbounded growth (as $R_m \to \infty$) in finite time of axisymmetric fields. *Cowling* (1933) proved that no such field may be maintained indefinitely by dynamo action. However, energetic instability may be important in explaining its large dominance in the Earth's field.

As has already been discussed in chapter two, the kinematic eigenvalue problem is extremely sensitive to the choice of flow. In contrast, non-normal theory is in general robust; in the geophysical context, this means that many flows that are approximately the same will allow similar magnetic fields to grow. In fact, if we replace the steady flow with a suitably statistically steady state, we can sustain magnetic fields by transient growth mechanisms even in, for example, the axisymmetric case where the field would have ordinarily decayed (*Farrell and Ioannou*, 1999b).

5.1.1 Pseudospectra and sensitivity

Having motivated the subject of non-normal (non-eigenmode) growth, we now detail some of the relevant theory. There are many important papers in the field, some of which are: *Trefethen et al.* (1993), *Reddy and Henningson* (1993), *Schmid and Henningson* (1994), *Trefethen* (1997), *Farrell and Ioannou* (1996).

A matrix L is normal if it has a complete set of orthogonal eigenvectors. Equivalently, it satisfies $\mathsf{L}\mathsf{L}^\dagger = \mathsf{L}^\dagger\mathsf{L}$ where its adjoint is $\mathsf{L}^\dagger$. Let us denote the set of eigenvalues of L as $\Lambda(\mathsf{L})$ and $z \in \Lambda(\mathsf{L})$ occurs if and only if $(z\mathsf{I} - \mathsf{L})$ is singular, that is, it has at least one singular value of 0 (I is the identity matrix). We define the resolvent to be the matrix $(z\mathsf{I} - \mathsf{L})^{-1}$ which has reciprocal singular values to $(z\mathsf{I} - \mathsf{L})$. This follows simply from section 4.8.1 where we write a matrix A as $\mathsf{U}\mathsf{W}\mathsf{V}^T$. If A is invertible then

$$\mathsf{A}^{-1} = \mathsf{V}\mathsf{W}^{-1}\mathsf{U}^T \tag{5.1}$$

and so its singular values are the reciprocals of those of A.

If z is 'near' but not equal to an eigenvalue of L, then at least one singular value is very

large. We define the ε-pseudospectrum of L, $\Lambda_\varepsilon(\mathsf{L})$, by the following equivalent statements:

$$\Lambda_\varepsilon(\mathsf{L}) = \left\{ z \in \mathbb{C} : \|(z\mathsf{I} - \mathsf{L})^{-1}\| \geq \varepsilon^{-1} \right\} \tag{5.2}$$

$$\Lambda_\varepsilon(\mathsf{L}) = \{ z \in \mathbb{C} : z \in \Lambda(\mathsf{L} + \Delta\mathsf{L}) \ \text{ for some } \Delta\mathsf{L} \text{ with } \|\Delta\mathsf{L}\| \leq \varepsilon \} \tag{5.3}$$

where $\|\mathsf{M}\|^2$ denotes the square of the matrix norm of M, in our case, the square of the largest singular value of the matrix $\mathsf{F}\,\mathsf{M}\,\mathsf{F}^{-1}$ (see the explanation below). $\mathbb{C}$ is the set of complex numbers.

If L is normal, the pseudospectra are circular regions centred on the eigenvalues of radius ε (see figure 5.1(a)). If L is non-normal then significant deviations can occur and in general they are much larger for a given ε (figure 5.1(b)). We may consider the pseudospectra as regions representing 'almost eigenvalues'. If inaccuracies creep in to our eigenvalue calculations, then another value in Λ_ε might be selected over the true value in Λ. This is rigourised in definition (5.3) which relates errors or changes in the matrix L to the effect on Λ. In the case of kinematic dynamo eigenvalue calculations, a small change in the flow or errors introduced by truncation or numerical effects may significantly affect the eigenvalue spectrum if the problem is sufficiently non-normal. Not only does this mean that as the non self-adjointness (proportional to R_m) is increased, it becomes harder to find numerically accurate solutions, but that it is increasingly difficult to find robust physical interpretation of the computations. This is important, since a small change in the flow should not alter our physical understanding of the processes involved, but it might switch off the dynamo by shifting the eigenvalues into a stable region.

The norms given in equations (5.2) and (5.3) in our case are taken using the energy inner product, where **u** and **v** represent magnetic fields relative to the Chebyshev basis:

$$(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{v}_1^T \, \mathsf{B}_{V+\hat{V}} \, \mathbf{v}_2 \qquad |\mathbf{v}|^2 = \mathbf{v}^T \, \mathsf{B}_{V+\hat{V}} \, \mathbf{v} \tag{5.4}$$

and $\|\mathsf{M}\|^2$ denotes the matrix norm: $\max \frac{|\mathsf{M}\mathbf{u}|}{|\mathbf{u}|}$, the largest singular value of the matrix $\mathsf{F}\,\mathsf{M}\,\mathsf{F}^{-1}$ (see section 4.8). We can write $\mathsf{F}^T\,\mathsf{F} = \mathsf{B}_{V+\hat{V}}$ since $\mathsf{B}_{V+\hat{V}}$, the basis energy inner product matrix, is positive definite (see *Press et al.*, 1992). If the time-dependence is described by the equation

$$\frac{\partial \mathbf{v}}{\partial t} = \mathsf{L}\,\mathbf{v} \tag{5.5}$$

then we may transform this into:

$$\frac{\partial (\mathsf{F}\,\mathbf{v})}{\partial t} = (\mathsf{F}\,\mathsf{L}\,\mathsf{F}^{-1})\,(\mathsf{F}\mathbf{v}) \tag{5.6}$$

Writing $\mathbf{u} = \mathsf{F}\mathbf{v}$, we see that the system can be re-written as

$$\frac{\partial \mathbf{u}}{\partial t} = (\mathsf{F}\mathsf{L}\mathsf{F}^{-1})\,\mathbf{u} \tag{5.7}$$

$$(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{u}_1^T \mathbf{u}_2 \qquad\qquad |\mathbf{u}|^2 = \mathbf{u}^T \mathbf{u} \tag{5.8}$$

In the transformed system, $\|\mathsf{M}\|$ denotes the largest singular value of the matrix M. This facilitates analysis of the pseudospectra equations and allows us to use the package *EigTool*[1], to plot contours of Λ_ε using the usual Euclidean norm.

Figure 5.1 shows pseudospectra in the complex plane, for the cases of (a) the diffusion problem and (b) the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$, $R_m = 50$ with an $m = 0$ E^A field symmetry. In (a) the operator is self-adjoint and so Λ_ε are circles of radius ε centred on the eigenvalues. In (b) the non-normality significantly effects the structure; note also that the scale is around four orders of magnitude smaller. The contour line $\varepsilon = 10^{-4.5}$ cuts the imaginary axis so that a change in the operator of a norm of this magnitude will give a growing eigenmode, as follows from property (5.3). To put this in perspective, a change in a flow of 1% by rms can incur a norm of $\Delta\mathsf{L}$ of $O(100)$. Thus a change in the flow of figure 5.1(b) of $O(10^{-6})$ by rms could potentially push an eigenvalue into the unstable real half-plane. Of course, perturbations to the induction operator of the required magnitude to create linear instability may not be associated with realisable velocity perturbations (i.e. they could merely be a numerical artifact), invalidating the above sensitivity analysis. However, as we follow higher contours of ε this argument becomes increasing more difficult to believe as they protrude greater distances into the real half-plane. Thus the induction operator may not be as sensitive as the above analysis suggests, but fractional flow perturbations of $O(10^{-4}$–$10^{-5})$ could be significantly alter the eigen-spectrum.

The above results are based on the $\mathbf{t}_1$ and KR flows (both are defined in section 5.3), KR comprising 99% of $\mathbf{t}_1$ by rms. Although these flows differ by just 1% , the norm of the vector difference of the two flows was 370 (for $N_{max} = L_{max} = 8$, and field symmetry E^A, m even and $R_m = 600$). A similar figure of 4000 was derived from the same truncation, field symmetry and R_m but from the difference between the STW and KR flows. The importance of these flows is that KR supports growing eigenmode solutions (if R_m is sufficiently large) but $\mathbf{t}_1$ does not. We already discussed in chapter two the difficulties of attributing a robust physical process to the kinematic dynamo in such a case, a small change in the flow having disastrous consequences. This issue is quantified in pseudospectra computations.

The pseudospectra are not only linked to the sensitivity of the operator, but also to the

[1] Eigtool is available at http://www.comlab.ox.ac.uk/pseudospectra/eigtool/

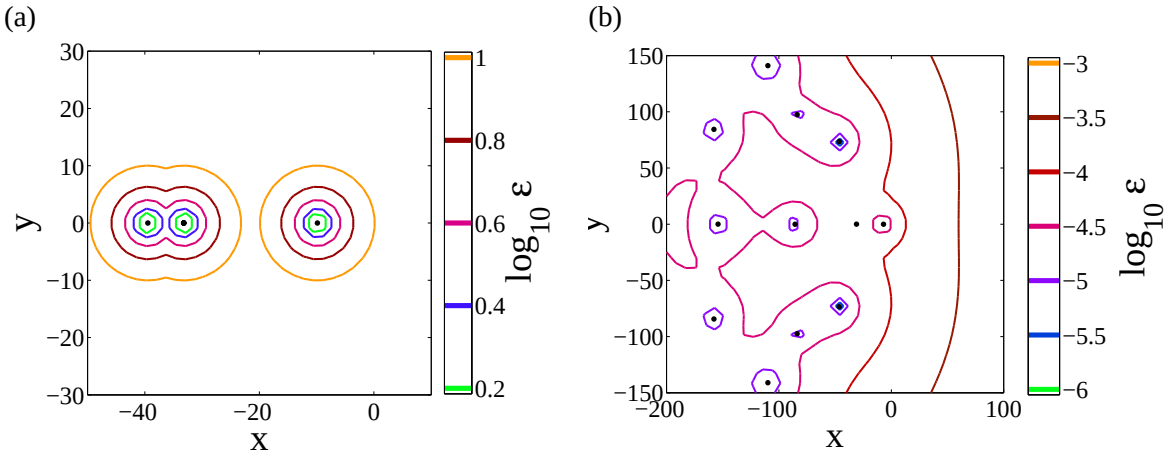


Figure 5.1: Pseudospectra for (a) the diffusion problem and (b) the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$, $R_m = 50$ with an $m = 0$ E^A field symmetry. The contours are of $log_{10}\varepsilon$ and are indicated to the right of each picture.

problems of maximal instantaneous energy growth and transient behaviour. Roughly speaking, if the contours protrude far into the real half-plane, then large transient growth is possible. We do not pursue this topic further here, but refer the interested reader to the references given at the start of this section.

5.2 Properties of flows

This section is adapted from *Acheson* (1990), chapter six, and is included here for completeness.

5.2.1 Fundamental concepts

Strain

The relative change in shape, size or volume due to externally applied forces is called the *strain*. The precise definition will depend on the geometry and the body concerned. Note that this is a dimensionless number.

Stress

The internal force per unit area associated with a strain is called the *stress*.

In a solid body such as a spring, *Hooke's Law* states that the strain is proportional to stress. If it is loaded at one end with a weight, it will stretch until the resistive elastic forces are equal and opposite to the externally applied force. Crucially this arises since the spring is solid and the different parts may not move relative to one another (apart from being stretched of course). In a fluid, if a stress is applied, fluid particles may slip past one another and no such analogue to the restoring force exists. It is more reasonable to assert that the stress is proportional

to the rate of strain (i.e. the rate of deformation). The constant of proportionality is the viscosity.

5.2.2 The stress tensor

The stress **t** exerted on an infinitesimal surface with unit normal **n** in the fluid, is given by

$$t_i = \mathsf{T}_{ij} n_j \tag{5.9}$$

where T is the *stress tensor*, and is a function only of position. Consider the simple case of a quiescent flow ($\mathbf{u} = \mathbf{0}$) where the only stresses are those caused by the pressure, and so

$$\mathsf{T}_{ij} = -p\,\delta_{ij} \tag{5.10}$$

The scalar function, p (the pressure) is the force per unit area exerted on the fluid into which **n** points by the fluid on the other side of the surface. Hence the force per unit area exerted on the surface is $-p\,\mathbf{n}$ (i.e. large pressures mean large compressional forces).

In general though, a viscous fluid will have an extra term

$$\mathsf{T}_{ij} = -p\,\delta_{ij} + \tau_{ij} \tag{5.11}$$

where τ_{ij} is the stress induced by the relative motions of the fluid, communicated by the action of viscosity. It may be shown that under the assumptions that

(i) τ_{ij} is linear in the velocity gradients (this defines a Newtonian fluid).

(ii) τ_{ij} should vanish if there is no deformation of the fluid elements.

(iii) The fluid is isotropic, that is, there is no preferred direction for its physical properties.

(iv) The fluid is incompressible.

then

$$\tau_{ij} = 2\mu\,\mathsf{e}_{ij} \qquad \mathsf{e}_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{5.12}$$

where μ is the viscosity (not to be confused with the maximum strain rate used later on).

The stress tensor T is therefore given by

$$\mathsf{T}_{ij} = -p\,\delta_{ij} + 2\mu\,\mathsf{e}_{ij} \tag{5.13}$$

where e is called the *rate of strain tensor*. The strain itself is an instantaneous measure of total fluid deformation since time zero, giving no indication of the stress at that instant. The strain rates give a measure of the local deformation of the fluid at that point. However, since the fluid is incompressible they must necessarily sum to zero (e has zero trace since $\boldsymbol{\nabla} \cdot \mathbf{u} = 0$) and so on 'average' there is no deformation. This is precisely what we mean physically by an incompressible fluid.

5.2.3 Strain rates and flow reversal

When a flow reverses its sign, its physical properties can change remarkably. At first sight this may seem rather strange since it merely flows in the other direction — surely its properties are preserved? In the 3-D case of the $\mathbf{s}_2$(MDJ) flow for example, fluid flows inwards along the equator and outwards towards the poles on the z axis. Flows of this type can remain incompressible as long as the outward flow is sufficiently strong, relative to the equatorial inflow. Near the z-axis there is a region of large vertical stress as fluid particles are accelerated along the z-axis, away from the origin. If we reverse the flow, these regions of large stress transform to regions of large vertical squashing (but equatorial spreading). Applying the idea *frozen flux*, if a magnetic field was 'frozen in' and aligned with the flow at a location of strong vertical stretching, magnetic energy would be locally created. In an equivalent setup but with the flow reversed, the magnetic energy would locally decrease. Thus in 3-D flow reversal has a significant effect.

In a 2-D incompressible flow however, the strain rate tensor e is of dimension 2×2 and the eigenvalues are $\pm\sqrt{\mathsf{e}_{12}^2 - \mathsf{e}_{11}\,\mathsf{e}_{22}}$ (using the facts that e is symmetric and has zero trace). These are of the form $\pm\sigma$, and we note explicitly that

$$\mathsf{e}\, v_\pm = \pm\sigma v_\pm \tag{5.14}$$

where $v_\pm$ are the eigenvectors associated with eigenvalues $\sigma v_\pm$. If the flow is reversed, so that $\mathsf{e} \to -\mathsf{e}$, we must have that

$$(-\mathsf{e})\, v_\pm = \mp\sigma v_\pm \tag{5.15}$$

Thus the maximum strain rate is unaffected, but the corresponding direction is now what used to be that of minimal strain rate (and vice versa). This is intuitive: on reversing the direction of flow, it merely causes the fluid to stress in the opposite direction, so that the direction of maximal squashing becomes that of maximal stretching.

Before going any further, we first clarify some nomenclature. In general, we will denote the pointwise maximum strain rate (the pointwise maximum eigenvalue of e) by μ, and

its global maximum by $\mu_{max} = \max \mu$.

5.3 Flows studied

In this section we outline the flows that will be studied in this thesis. We measure the typical velocity (on which R_m is defined) by their rms values. For a flow in poloidal-toroidal form, this is

$$\frac{4\pi}{3}\mathbf{u}_{rms}^2 = \sum_\alpha l_\alpha(l_\alpha+1)N_\alpha^2 \int_0^1 \left[\frac{l_\alpha(l_\alpha+1)}{r^2}s_\alpha^2 + \left(\frac{ds_\alpha}{dr}\right)^2 + t_\alpha^2\right] dr \tag{5.16}$$

where the radial scalar functions s and t are similar to those used for the magnetic field; we will always use lower case for the flow to distinguish between them.

We normalise each flow by its rms value and then multiply accordingly by the value of R_m required. To do this, each set of flow scalars is defined up to a constant K^{-1} which is chosen to be such that the flow has unit rms. Depending on the definition of flow, in certain simple cases this can be analytically evaluated; more often we use quadrature (in radius) to compute it.

5.3.1 An axisymmetric toroidal flow

No toroidal-only flow can sustain a magnetic field indefinitely. However, these are an important ingredient in many flows that can support dynamo action and are therefore worthy of a study in their own right. We define an axisymmetric toroidal flow by the scalar

$$t_1(r) = K^{-1} r^2 (1 - r^2) \tag{5.17}$$

This defines a differentially rotating flow which satisfies the non-slip condition on $r = 1$ and also the regularity condition of section 3.3.4 that ensures infinite differentiability. The normalisation constant is $K = \sqrt{\frac{16}{315}}$ in this case. Differential rotation is thought to be a crucial process in the geodynamo, generating toroidal field from poloidal field, and there is evidence that it is an important aspect of rotating convection (*Busse*, 2002).

This flow can be visualised by contours of its sole non-zero component, u_ϕ, in a meridian plane as shown in figure 5.2(a). Figure 5.2(b) shows contours of μ, the pointwise maximum strain rate.

5.3.2 Axisymmetric poloidal flows

We study three axisymmetric poloidal flows denoted $\mathbf{s}_2$(MDJ), $\mathbf{s}_2$(KR) and $\mathbf{s}_2$(IC). These are based on the flows of *Dudley and James* (1989) and *Kumar and Roberts* (1975), the former flow

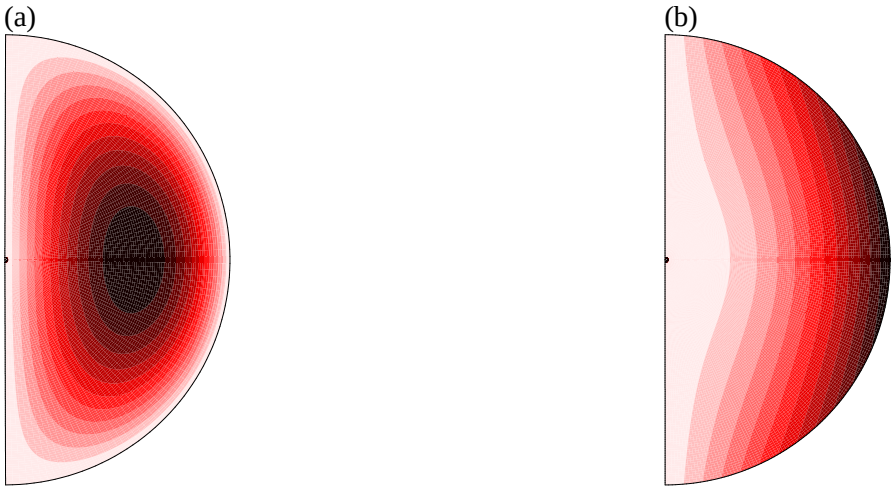


Figure 5.2: Contours in a meridian plane of (a) $0 \leq u_\phi \leq 1.708$ and (b) $0 \leq \mu \leq 4.437$, the pointwise maximum strain rate. The flow is strongest midstream and vanishes on the boundary where μ is maximal, corresponding to large shear.

being modified into a polynomial form (hence *MDJ*). These are defined as

$$s_2(r) = K^{-1} r^3 (1-r^2)^2 \qquad \text{for MDJ} \tag{5.18}$$

$$s_2(r) = K^{-1} r^6 (1-r^2)^3 \qquad \text{for KR} \tag{5.19}$$

and for $\mathbf{s}_2$(IC):

$$s_2(r) = \begin{cases} 0 & r \leq 0.35 \\ (r-0.35)^3 (1-r^2)^2 & 0.35 \leq r \leq 1 \end{cases} \tag{5.20}$$

These flows have no azimuthal component, so we can view streamlines in meridian planes as shown in figure 5.3. All flows are non-slip at $r = 1$ (requiring both $s_2(1) = 0$ and $s_2'(1) = 0$) but the $\mathbf{s}_2$(KR) flow does not satisfy the regularity conditions at the origin, required for infinite differentiability. However, due to the factor of r^6, in fact the first two derivatives of $s_2(r)$ vanish at the origin; thus the first derivatives of $\mathbf{u}$ are continuous which is all that is required for the kinematic dynamo problem.

The streamlines of the $\mathbf{s}_2$(MDJ) flow represent a convection pattern that has a stagnation point at the origin. In the $\mathbf{s}_2$(KR) flow, the streamlines are 'bent' away from the origin which is not only a stagnate point but one at which the strain rates also vanish. We also modify the scalar of $\mathbf{s}_2$(MDJ) to incorporate a stagnate inner core of radius $r = 0.35$, forming the $\mathbf{s}_2$(IC) flow, modelling a solid (iron) inner core of the same conductivity as the liquid outer core.

These flows model the large scale poloidal overturn in the outer core driven by some

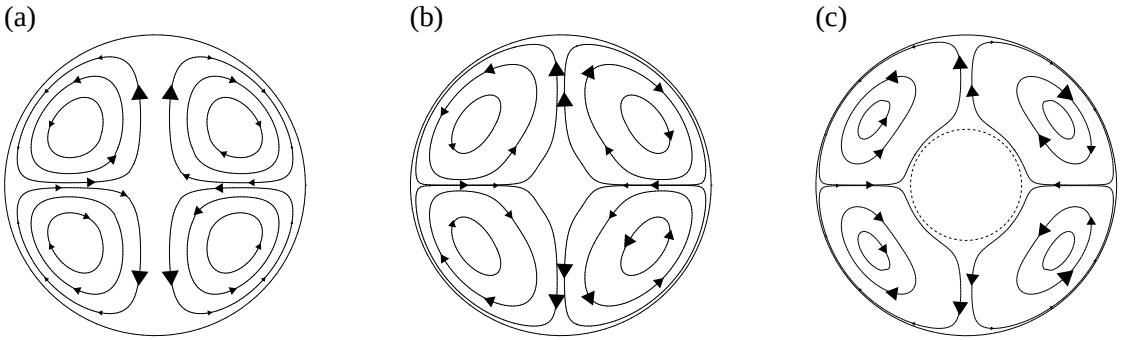


Figure 5.3: Streamlines in a meridian plane of (a) $\mathbf{s}_2$ (MDJ), (b) $\mathbf{s}_2$ (KR) and (c) $\mathbf{s}_2$ (IC). Arrows show the direction of flow. In (c), a solid inner core of radius 0.35 is represented by a quiescent flow as shown by the dashed circle.

kind of convective process. They may be further motivated by results of axisymmetric convection which take on an $\mathbf{s}_2$ form (see *Weir*, 1976; *Chandrasekhar*, 1961).

5.3.3 $\mathbf{t}_1\,\mathbf{s}_2$ flows

The above $\mathbf{t}_1$ and $\mathbf{s}_2$ flows can be combined to form $\mathbf{t}_1\,\mathbf{s}_2$ velocities. We define the ratio of the rms values of the two constituent components by τ, where

$$\tau = \frac{\sqrt{\frac{3}{4\pi}\int_V |\mathbf{s}_2|^2\,dV}}{\sqrt{\frac{3}{4\pi}\int_V |\mathbf{t}_1|^2\,dV}} \tag{5.21}$$

The relative signs of the two components should be obvious from the context. We shall particularly make use of the flows: $\mathbf{t}_1\,\mathbf{s}_2$(KR) and $\mathbf{t}_1\,\mathbf{s}_2$(MDJ), since although the components behave approximately the same, only the latter flow exhibits growing eigenmode solutions.

5.3.4 Other flows

We shall also study the full *Kumar and Roberts* (1975) flow (KR) and the *Sarson* (2003) flow (STW — Sarson thermal wind). These are defined respectively as

$$\begin{aligned} t_1(r) &= K^{-1}r^2(1-r^2) & s_2(r) &= \varepsilon_1 K^{-1}r^6(1-r^2)^3 \\ s_2^{2s}(r) &= \varepsilon_2 K^{-1}r^4(1-r^2)^2\cos pr & s_2^{2c}(r) &= \varepsilon_3 K^{-1}r^4(1-r^2)^2\sin pr \end{aligned} \tag{5.22}$$

and

$$
\begin{aligned}
t_1(r) &= K^{-1} r^2 (1-r^2) \\
s_2^{2s}(r) &= \varepsilon_2 K^{-1} r^4 (1-r^2)^2 \cos pr \qquad s_2^{2c}(r) = \varepsilon_3 K^{-1} r^4 (1-r^2)^2 \sin pr \\
t_3^{2s}(r) &= -\varepsilon_3 K^{-1} \frac{4}{5}\sqrt{5}\left(\frac{d}{dr} - \frac{3}{r}\right) r^4 (1-r^2)^2 \sin pr \\
t_3^{2c}(r) &= +\varepsilon_2 K^{-1} \frac{4}{5}\sqrt{5}\left(\frac{d}{dr} - \frac{3}{r}\right) r^4 (1-r^2)^2 \cos pr
\end{aligned} \tag{5.23}
$$

The value K is chosen to give the flows an rms value of 1. Kumar and Roberts found well converged eigenvalue instability for the values $\varepsilon_1 = 0.03$, $\varepsilon_2 = \varepsilon_3 = 0.04$, $p = 3\pi$ giving $R_m^c = 890$. Sarson found similar instability with the values $\varepsilon_2 = 0.04$, $\varepsilon_3 = -0.04$, $p = 3\pi$, with $R_m^c = 479$. The STW flow is thermally driven and dynamically self-consistent in a geostrophic regime, where a force balance is achieved between pressure, buoyancy and rotation. Although the thermal source is not geophysically motivated, it is clearly a step towards more Earth-like flows.

The components $\mathbf{s}_2^{2s/c}$ represent a non-axisymmetric spiralling convective flow; the degree of tilt (variation of azimuthal phase with radius) is described by the parameter p (*Sarson and Busse*, 1998). We lastly note that the KR and STW flows comprise 99% and 96% $\mathbf{t}_1$ by rms respectively. Thus a small change in rms of the axisymmetric toroidal component converts these flows that operate as kinematic dynamos to $\mathbf{t}_1$, which does not.

5.4 Some eigenvalue results

We do not dwell on presenting many eigenvalue calculation results; instead we refer the reader to one of the many works cited in section 2.6. However, a few calculations are worth detailing since they aid us in our understanding of the difficulties therein.

5.4.1 A working $\mathbf{t}_1\,\mathbf{s}_2$ kinematic dynamo

Figure 5.4 shows $\lambda_R = \Re(\lambda)$ as a function of R_m, for the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow with (a) $\tau = 0.5$ and (b) $\tau = 1$ (flow energy equipartition). The short-dashed solid line shows the $m = 0$ symmetry, solid the $m = 1$ and long-dashed the $m = 2$. The truncation used is $N_{max} = L_{max} = 40$ for the $m = 0$ symmetry and $N_{max} = L_{max} = 32$ for $m = 1, 2$; the solutions are all satisfactorily converged. In (a) the flow supports a growing field (of $m = 1\ E^S$ symmetry) at a critical magnetic Reynolds number of $R_m^c = 55.0$. Here the solution is oscillatory, the eigenvalue being $\lambda = 0 \pm 35.075i$.

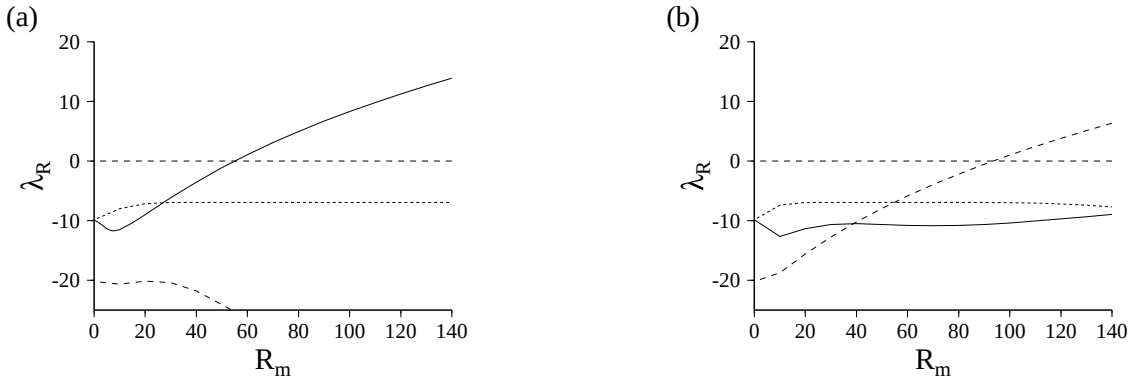


Figure 5.4: The induction equation eigenvalue problem for the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow with (a) $\tau = 0.5$, (b) $\tau = 1$. Plotted is $\Re(\lambda)$ against R_m. Short-dashed is the $m = 0$ field symmetry, solid $m = 1$ and long-dashed $m = 2$. The critical value $\Re(\lambda) = 0$ is shown as the horizontal dashed line.

The $m = 0$ field solution, precluded by Cowling's theorem initially increases in its value of λ_R, and remains at around $\lambda_R = -7$ for most of the range shown; the $m = 2$ symmetry decays. In (b) the kinematic dynamo has a growing $m = 2\ E^S$ field solution, indicating a different mechanism of field generation from (a) although with the same equatorial symmetry. The $m = 1$ solutions now decay in this range though λ_R shows some signs of increasing at larger R_m. No higher calculations in R_m converged sufficiently, so whether or not a growing $m = 1$ eigenmode solution exists is unknown. These two kinematic dynamos both support exponentially growing field solutions, but the mechanism by which this occurs is not robust. In contrast, the $m = 0$ solution seems to behave similarly: in this case changing τ seems to have little effect.

In general if the kinematic dynamo fails, then the eigenvalues decrease with increasing R_m. Even if there exists a growing eigenmode for a value of R_m^c, it does not necessarily follow that the kinematic dynamo will still work for all $R_m > R_m^c$, the subject of so called fast dynamo theory. Intuitively one might have thought that increasing R_m leads robustly to more efficient dynamo action. In view of the above however, this is false if we only consider the eigenvalue problem. In fact, a far more physical picture is obtained using the methods of non-normal analysis (see section 5.5 onwards).

That the behaviour of the $m = 0$ symmetry is robust and generally tends towards the region in which $\Re(\lambda) > 0$ as R_m is increased, is significant if we consider the pseudospectra. Recall that contours of the pseudospectra surround the eigenvalues, and that the more positive (or less negative) the eigenvalue, the smaller the operator perturbation needed to obtain an unstable eigenvalue. In this sense then, as R_m is increased the inclination of the system to undergo large robust $m = 0$ field growth will increase; this is confirmed by direct calculations in chapter six.

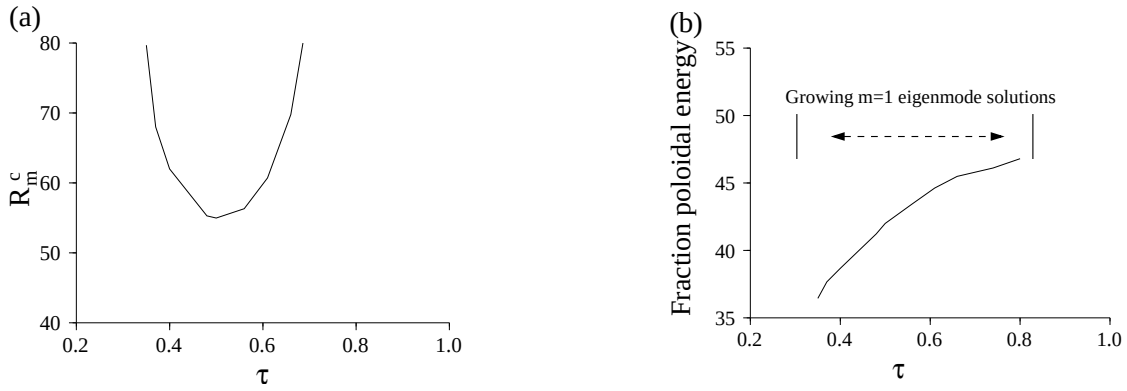


Figure 5.5: (a) A plot of R_m^c for $m = 1 E^S$ fields against τ for the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow. (b) The fraction of poloidal energy at R_m^c against τ.

5.4.2 Sensitivity and interpretation

Figure 5.5(a) shows R_m^c for the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow against τ for the $m = 1$ field symmetry. For fields of this type, kinematic dynamo action is only possible within $0.3 \leq \tau \leq 0.8$, a fairly narrow range of τ. Some hint as to the generating mechanism can be found in figure 5.5(b), showing the fraction of poloidal field (by energy) at R_m^c associated with each value of τ. As the poloidal flow strength (τ) increases, a higher fraction of poloidal field is generated. This mechanism is fairly intuitive, if we associate the poloidal flow strength with the efficiency of converting toroidal field to poloidal field. It is somewhat surprising, given that the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow supports growing eigenmode solutions, that on changing the $\mathbf{s}_2$ defining scalar to that of $\mathbf{s}_2$(KR), kinematic dynamo action ceases altogether, at least in this range of R_m for $m = 1, 2$. Both $\mathbf{s}_2$ flows describe the same physical process of convection, but only one promotes dynamo action. This again flags the sensitivity issue of these calculations. The velocity term of the induction equation is a function of the scalar quantities: s_2, s_2', s_2'', t_1 and t_1' so perhaps the sensitivity stems from one or more of these?

Figure 5.6 shows a comparison of s_2, s_2',s_2'' of the $\mathbf{s}_2$(MDJ) and $\mathbf{s}_2$(KR) flows, each normalised to have unit rms value. The scalars s and their first derivatives agree quite well; however their second derivatives differ considerably over the domain. This may explain why some ostensibly similar flows show different kinematic dynamo properties: not only must we take into account the large scale behaviour but also the strain rates (also found by *Holme*, 2003). An additional point is that the $\mathbf{s}_2$ (KR) flow has not only vanishing velocity at the origin but also vanishing strain rate. This may be important in explaining why no growing eigenmode solutions of the form $\hat{\mathbf{B}}(\mathbf{r})\,e^{\lambda t}$ are supported, for these require the rate of generation (λ) to be everywhere the same. If this process is inhibited near the origin then no kinematic dynamo can operate.

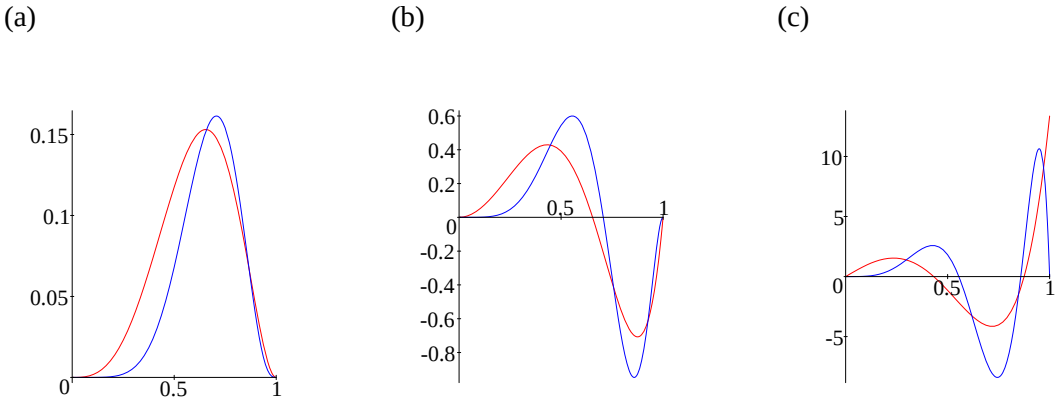


Figure 5.6: Comparison of scalar functions and their derivatives for the $\mathbf{s}_2$ (MDJ) (shown in red) and $\mathbf{s}_2$(KR) (shown in blue) flows, $0 \leq r \leq 1$. (a) shows $s_2(r)$ (b) $s_2{}'$(r) and (c) $s_2{}''(r)$. The second derivatives disagree considerably over the domain.

5.5 Instantaneous magnetic energy growth

5.5.1 Recap of important ideas

Recall that the equation governing the time-dependence of magnetic energy was

$$\frac{dM}{dt} = R_m \int_V \mathbf{B} \cdot \mathsf{e}\,\mathbf{B}\, dV - \int_V |\boldsymbol{\nabla} \times \mathbf{B}|^2\, dV \tag{5.24}$$

so that energy is increased by the action of e, physically representing the stretching of field lines by the flow, and decreased through Ohmic dissipation (stemming from the diffusion of magnetic field). The problem of the onset of magnetic energy growth i.e. the maximised instantaneous magnetic energy growth rate is described by

$$\lambda_E = \max_{\mathbf{B}} \frac{1}{\sqrt{M}} \frac{d}{dt} \sqrt{M} \tag{5.25}$$

and that this was numerically formulated by the generalised eigenvalue problem:

$$\lambda_E \, \mathsf{B}_{V+\hat{V}}\, \mathbf{v} = \left(R_m \mathsf{E} + D_{V+\hat{V}} \right) \mathbf{v} \tag{5.26}$$

where the matrix entries are given in section 4.5.1.

The idea is to see how fast a field can grow in energy initially, as measured by λ_E. The lowest value of R_m giving $\lambda_E = 0$ is termed R_m^E, describing a critical state below which all field structures decay, and above which at least one grows. In the following analysis, we essentially

elaborate on the results of *Livermore and Jackson* (2004).

5.5.2 Magnetic field line stretching

The stretching effect of a fluid can locally cause properly aligned magnetic fields to grow in energy. The rate of strain on a field line pointing in the direction of the unit vector $\hat{\mathbf{B}}$ is given by $\mathsf{e}\hat{\mathbf{B}}$. The rate at which the field line is actually stretched is the back projection of this quantity onto the field itself i.e. $\hat{\mathbf{B}}\cdot\mathsf{e}\hat{\mathbf{B}}$. This is a pointwise quantity and will in general vary throughout the flow. Plotting contours of this gives an indication of local alignment of the field to the direction of maximal strain rate.

At large values of R_m it might be supposed that the optimal magnetic field configuration would be some sort of local solution around the location of maximal rate of strain in the flow, in order to maximise its growth through stretching for a given amount of field energy. However, this is not so, and in such cases dissipation plays a large role due to the small length scales involved.

We illustrate the key ideas in a simple example.

5.5.3 Example

We consider a 2-D flow in Cartesian coordinates:

$$\mathbf{u} = x(1-x)\mathbf{j} \tag{5.27}$$

$0 \leq x \leq 1$, where the unit vector $\mathbf{j}$ points in the y-direction. The flow profile looks much like figure 5.7(a), with the maximum gradients of the profile are highlighted in the heavy dashed lines.

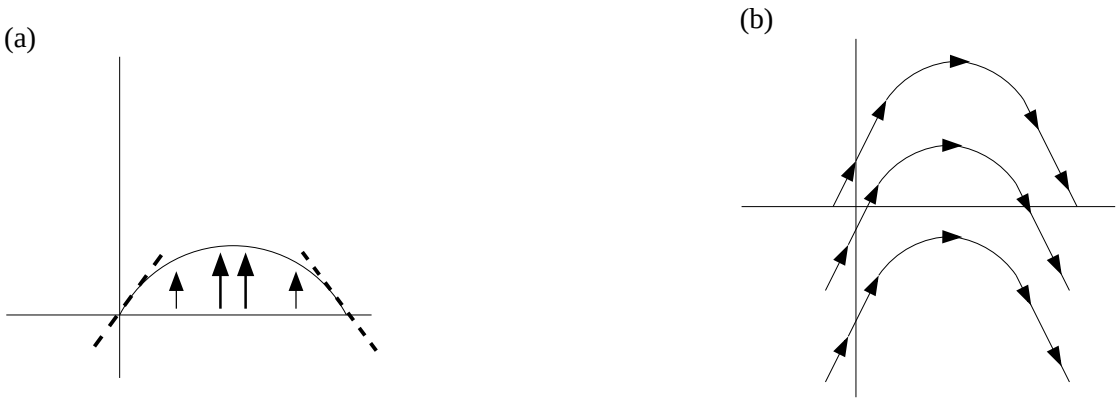


Figure 5.7: (a) Cartoon of the flow profile with the maximum gradients (maximum stretching) highlighted in heavy dashed line. (b) To maximise the instantaneous energy growth, the field might align itself in this kind of configuration.

To find the direction of maximal strain rate, we must compute e:

$$\mathsf{e} = \frac{1}{2}\begin{pmatrix} 0 & 1-2x \\ 1-2x & 0 \end{pmatrix} \tag{5.28}$$

The eigenvalue decomposition of e is explicitly $\frac{1}{2}(1-2x)$ with eigenvector $(1,1)$ and $-\frac{1}{2}(1-2x)$ with eigenvector $(1,-1)$. In $0 \leq x \leq 1$, the value of $\mu_{max} = 1/2$, attained at $x = 0$ and $x = 1$.

Therefore if we were to find a magnetic field that maximised $\mathbf{B} \cdot \mathsf{e}\mathbf{B}$ in this geometry, with $\mathbf{B}$ lying in the plane, we would expect it to look similar to that of figure 5.7(b). The field lines are aligned along the maximal stretching directions at $x = 0$ and $x = 1$.

Note that this is just a cartoon; there are other effects that may be important in the general case, for example, diffusion, the solenoidal condition and boundary conditions. Nonetheless, it is worth keeping this picture in mind to help with visualising the field structure in 3-D flows.

We note that energetic instability in the simple infinite shear flow of $\mathbf{u} = x\mathbf{j}$ has been solved analytically by *Farrell and Ioannou* (1999a), corresponding to the local first order flow at the origin in this example. They find $R_m^E = \frac{16\sqrt{3}}{9}$ and that λ_E scales linearly with R_m as $R_m \to \infty$, where the most unstable field is aligned with the direction $(1,1)$, just as we postulated above.

5.6 An axisymmetric toroidal flow

The rate of strain tensor for the axisymmetric toroidal flow of section 5.3.1 is given in spherical polar coordinates (e.g. see *Acheson*, 1990, for the components of e in this geometry) by

$$\mathsf{e} = -r^2 \sin\theta K^{-1} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{5.29}$$

The value of $K = \sqrt{(16/315)}$ ensures the flow is of unit rms. The eigenvalue decomposition of e is: $K^{-1}r^2\sin\theta$ with eigenvector $(1,0,-1)$, $-K^{-1}r^2\sin\theta$ with eigenvector $(1,0,1)$ and 0 with eigenvector $(0,1,0)$ (a null space).

The global maximum rate of strain (μ_{max}) is $K^{-1} = 4.437$, attained at $r = 1$, $\theta = \frac{\pi}{2}$ (see figure 5.2(b)). The associated eigenvector for the maximal growth at any point is $(1,0,-1)$, since $r^2\sin\theta > 0$ everywhere. To maximise the stretching, the field must try to align itself pointwise with $(1,0,-1)$ as best it can, especially near regions where the strain rate is maximal, whilst also minimising the effects of diffusion.

Also note that since e has only two eigenvalues, they necessarily differ only by sign. This means that a reversal in flow direction makes no difference to the analysis, other than the selection of the other eigenvector, amounting to nothing more than a reversal in the direction of $\hat{\phi}$.

5.6.1 Symmetry separation

Figure 5.8 shows λ_E as a function of R_m for the $m = 0$ (blue), $m = 1$ (red), E^A (solid) and E^S (dashed) symmetries. The $m = 1$ E^S is the critical symmetry which becomes energetically unstable at $R_m^E = 16.8$, despite being eigenvalue stable for all R_m. Also note that the symmetries do not differ greatly in their instability: the $m = 0$ E^A symmetry has $R_m^E = 19.3$.

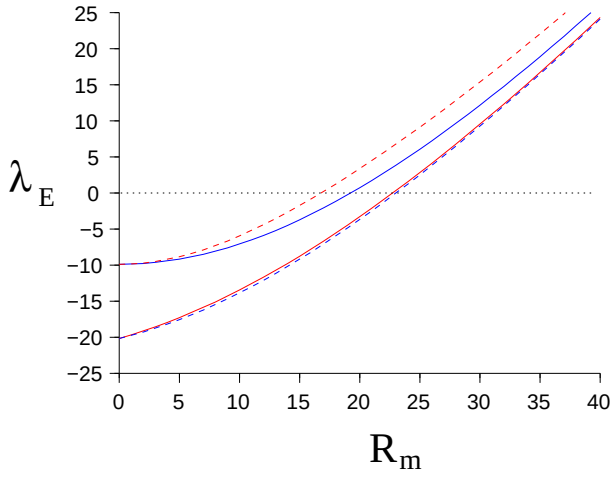


Figure 5.8: Field symmetry separation in energetic instability. λ_E is plotted as a function of R_m. Red shows $m = 1$, blue $m = 0$, solid E^A and dashed E^S. Computed for $N_{max} = L_{max} = 16$.

5.6.2 The structure of the critical field

Figure 5.9(a) shows field lines of the critical $m = 1$ E^S symmetry in an equatorial plane. Crucially the field aligns itself with the direction of maximal strain rate, as shown by contours of $\hat{\mathbf{B}}\mathsf{e}\hat{\mathbf{B}}$ in figure 5.9(b) (where $\hat{\mathbf{B}} = \mathbf{B}/|\mathbf{B}|$). Its structure is such that it has large intensity near such regions, shown by contours of $R_m^E \mathbf{B}\mathsf{e}\mathbf{B}$ in figure 5.9(c). The locations where the field is misaligned, that is compressed by the flow, are very low in field intensity so that they contribute little to the energy instability. Figure 5.9(d) shows the sum of pointwise stretching and dissipative contributions to the energy integral: contours of $R_m^E \mathbf{B}\mathsf{e}\mathbf{B} - |\boldsymbol{\nabla} \times \mathbf{B}|^2$.

In trying to predict such behaviour, we would need to take into account alignment of the field with the maximal stretching directions along with the negative effect of dissipation

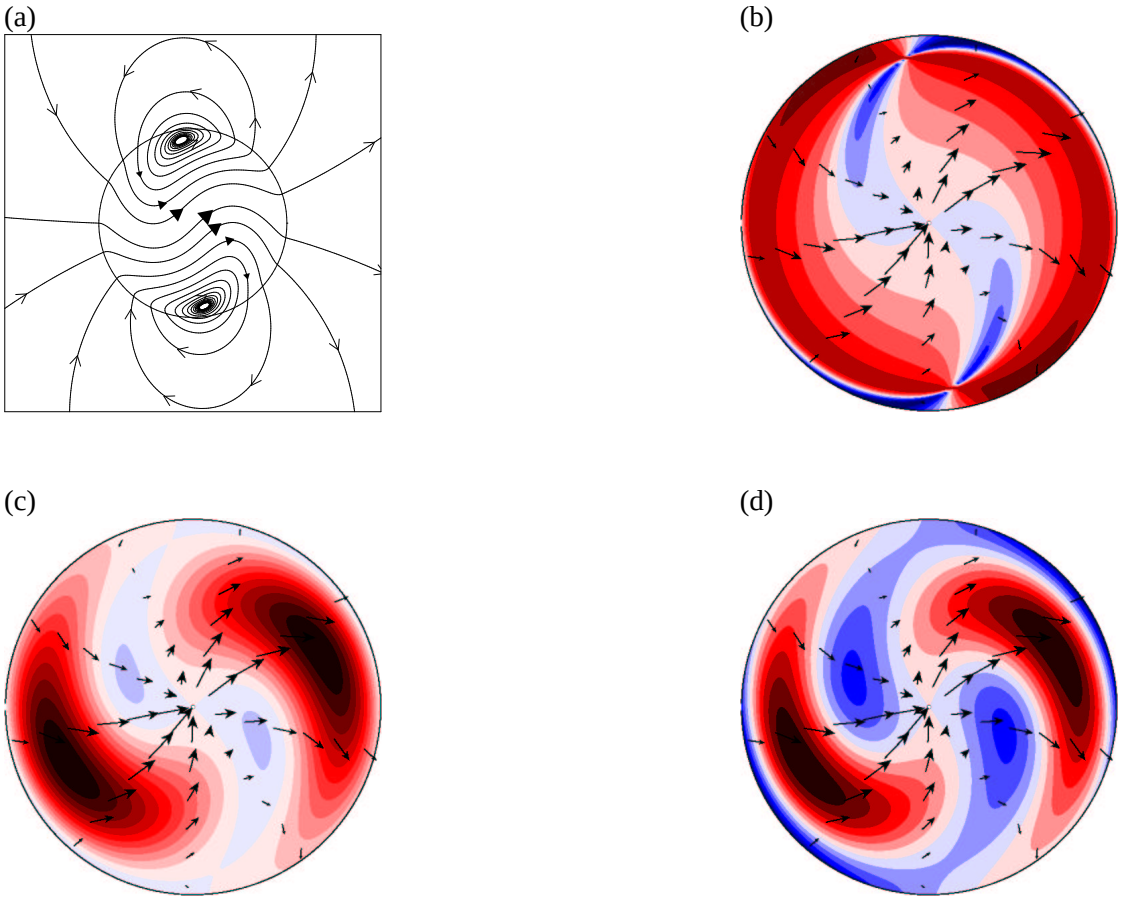


Figure 5.9: The energetically critical field of $m = 1$ E^S symmetry for $\mathbf{t}_1$ at $R_m^E = 16.8$, in an equatorial section ($\theta = \frac{\pi}{2}$) showing: (a) fieldlines, (b) contours of $-4.435 \leq \hat{\mathbf{B}}\mathsf{e}\hat{\mathbf{B}} \leq 4.437$, (c) contours of $-5.503 \leq R_m^E \mathbf{B}\mathsf{e}\mathbf{B} \leq 45.895$, (d) contours of $-35.983 \leq R_m^E \mathbf{B}\mathsf{e}\mathbf{B} - |\boldsymbol{\nabla} \times \mathbf{B}|^2 \leq 45.880$. The field is normalised to have unit energy. Red is positive, blue is negative; arrows show the field projected onto the plane. $\phi = 0$ is located at the 3 o'clock position.

on small length scales. The location of μ_{max} at the equator and its associated direction, means that necessarily E^S field symmetries are favoured. This is because the E^A symmetry forces $\mathbf{B} = (0, B_\theta, 0)$ there, and so lies in the null space of e — i.e. it is not stretched there at all. The $m = 1$ subspace of E^S symmetries contains spatially simpler harmonics than those of $m = 0$, so dissipation has a lesser effect. Consequently the critical symmetry calculated numerically agrees with this intuitive reasoning.

5.7 Axisymmetric poloidal flows

We consider now the instantaneous magnetic energy stability of the $\mathbf{s}_2$ (MDJ), $\mathbf{s}_2$(KR) and $\mathbf{s}_2$(IC) flows, modelling poloidal overturn convection patterns in the Earth's core. None can support growing eigenmode solutions on their own but they can, as in the toroidal case, support magnetic

energy instability. Each flow is non-slip and has a stagnation point at the origin; in the case of the $\mathbf{s}_2$(KR) flow the strain rates also vanish there and in $\mathbf{s}_2$(IC) the whole region $r \leq 0.35$ is quiescent.

5.7.1 Symmetry separation

Figure 5.10 shows (left) plots of λ_E as a function of R_m for $m = 1$ (red), $m = 0$ (blue), E^A (solid) and E^S (dashed) symmetries along with (right) fieldlines of the critical (for $R_m^E > 0$) $m = 0$ E^A (poloidal only) field in each case. Each row of the figure corresponds to a different flow: pictures (a) and (b) to $\mathbf{s}_2$(MDJ), (c) and (d) to $\mathbf{s}_2$(KR) and (e) and (f) to $\mathbf{s}_2$(IC). Note the striking similarity between all of these flows; in particular the similar dominating behaviour of E^A fields is apparent for $R_m > 0$, with $m = 0$ E^A being the critical symmetry. The reversed flows ($R_m < 0$) show no such clear separation.

5.7.2 Structure of the critical fields

In the case of the $\mathbf{s}_2$(MDJ) flow, since it is infinitely differentiable everywhere, its strain rate tensor is straightforward to compute. However, due to the increased flow complexity (relative to $\mathbf{t}_1$ above), e is spatially dependent. The flow attains μ_{max} at the origin, which turns out to be a point of fundamental importance. Here, e is

$$\mathsf{e} = \frac{3}{K}\begin{pmatrix} 2\cos^2\theta - \sin^2\theta & -3\cos\theta\sin\theta & 0 \\ -3\cos\theta\sin\theta & 2\sin^2\theta - \cos^2\theta & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{5.30}$$

Alternatively we may transform this back into Cartesian coordinates in which it takes on a simple form:

$$\mathsf{e} = \frac{1}{K}\begin{pmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 6 \end{pmatrix} \tag{5.31}$$

It is clear that the associated direction of maximal stretching is vertically up (and down), with a strain rate of $6K^{-1}$. This is physically intuitive, once we consider the 3-D geometry of the flow. In the Cartesian system, the $\mathbf{s}_2$ (MDJ) flow is

$$\mathbf{u} = \frac{1}{K}\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\left[\left(\frac{3}{2}z^2 - \frac{1}{2}(x^2+y^2+z^2)\right)\left(1-(x^2+y^2+z^2)^2\right)\right](x,y,z) \tag{5.32}$$

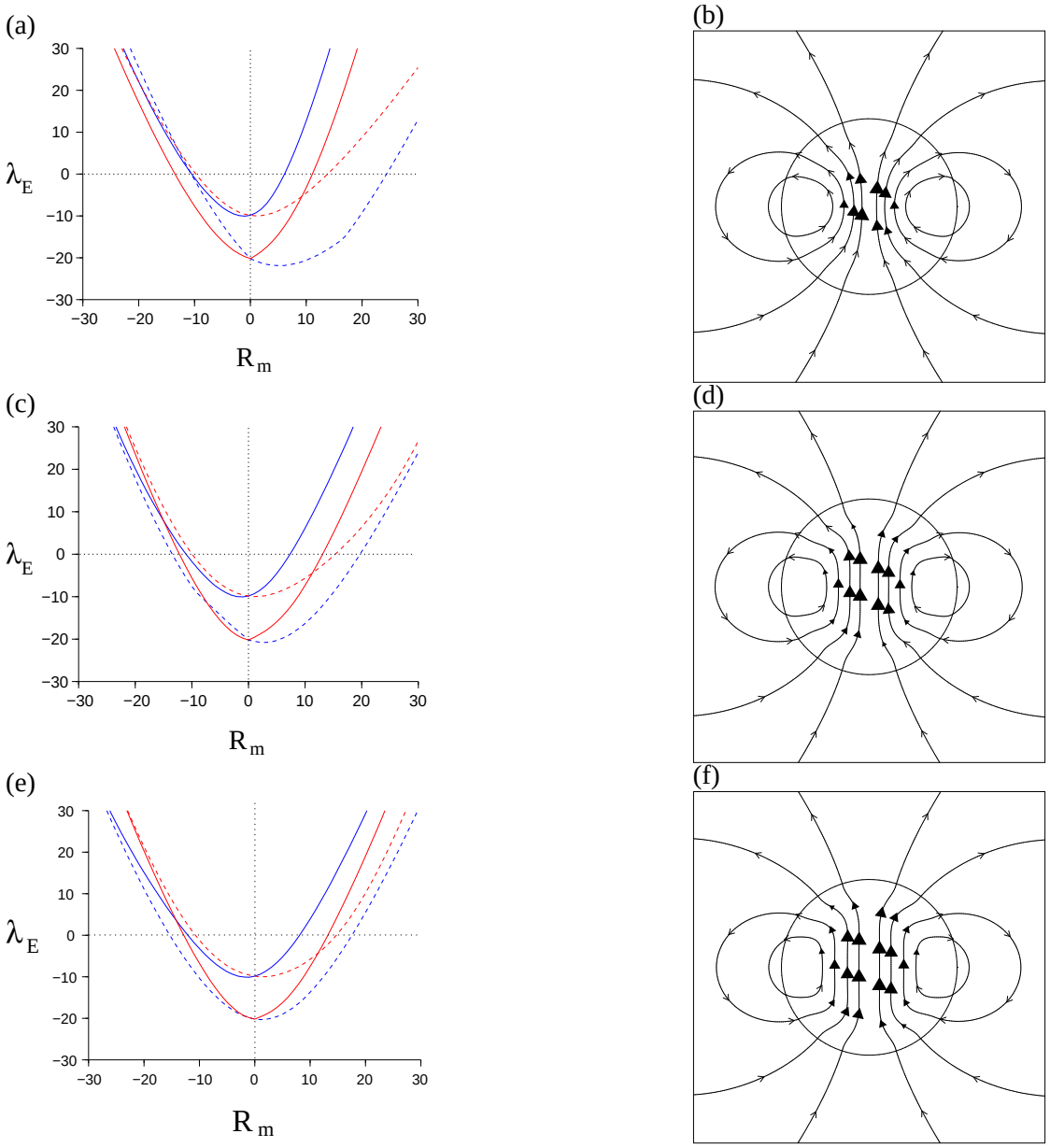


Figure 5.10: Field symmetry separation in energy instability (left) and fieldlines of the critical field in a meridian plane (right). Pictures (a) and (b) refer to the $\mathbf{s}_2$(MDJ) flow; (c) and (d) to $\mathbf{s}_2$(KR) and (e) and (f) to $\mathbf{s}_2$(IC). In (a), (c) and (e) λ_E is plotted as a function of R_m. Red indicates $m = 1$, blue $m = 0$, solid E^A, and dashed E^S. The critical symmetry in each case is $m = 0$ E^A and of poloidal form only, which means that fieldlines are confined to meridian planes. They are shown in (b), (d) and (f). Arrows show the direction of the field; solid signifies high intensity. The values of R_m^E in each case are 6.1, 7.4 and 8.1 respectively. The truncation level used is $N_{max} = L_{max} = 16$.

which to first order in (x,y,z) at the origin is $\frac{1}{K}(-3x,-3y,6z)$. It is clear how the fluid behaves here: there is convergent horizontal flow and vertical divergent flow of twice the strength.

Hence we get strong straining (of rate $\frac{6}{K}$ = 10.027) in $\hat{\mathbf{k}}$, and weaker compression in $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ (of rate $\frac{3}{K} = 5.017$). On flow reversal, this transforms to strong compression along the z-axis, and weak stretching in horizontal directions. The origin is the location of maximal strain rate, and has an associated stretching direction that is easily accessible to E^A magnetic fields.

Figure 5.11 shows contour plots of the local maximum strain rate μ for the (a) forward and (b) reversed $\mathbf{s}_2$(MDJ) flows and that for the $\mathbf{s}_2$ (KR) flow in (c) and the $\mathbf{s}_2$ (IC) flow in (d). Note that changing the flow direction completely changes the behaviour of μ, for example the maximum strain rate above becomes $3K^{-1}$. It is no surprise therefore that this has a dramatic effect on the energetics.

The vanishing strain rates at the origin in the $\mathbf{s}_2$(KR) and $\mathbf{s}_2$(IC) flows mean that the location of μ_{max} has moved to near the boundary. However, immediately above and below the origin there is strong vertical stretching (figures 5.11(c) and 5.11(d)), and although not maximal these locations dominate the field solution. The critical fields are shown in figure 5.12 with their relation to the flows. Again each row represents each $\mathbf{s}_2$ type flow studied: figures 5.12(a)-(c) the $\mathbf{s}_2$(MDJ) flow, 5.12(d)-(f) the $\mathbf{s}_2$(KR) flow and 5.12(g)-(i) the $\mathbf{s}_2$(IC) flow . Contours of $|\mathbf{B}|$ are showed in Column 1, $R_m^E \mathbf{B}\cdot \mathtt{e}\mathbf{B}$ in column 2 and $R_m^E \mathbf{B}\cdot \mathtt{e}\mathbf{B} - |\boldsymbol{\nabla}\times\mathbf{B}|^2$ in column 3. Since the fields are axisymmetric, the latter quantity must integrate to zero over the region shown, since the field is energetically critical.

All the critical fields align themselves with strong vertical stretching at or near the origin. The $\mathbf{s}_2$(MDJ) flow stretches the field particularly strongly (note that the contours of figure (b) are roughly 4 times those of (e) and (h)) at the origin, the location of its maximal strain rate. The field is correctly aligned at this point and in the nearby region in order to feel the effect of this maximally. The $\mathbf{s}_2$(KR) and $\mathbf{s}_2$(IC) flows have no rate of strain at the origin. However, vertically aligned fields above and below the origin get stretched in a similar way. The contour values show that the positive contributions of $R_m^E \mathbf{B}\cdot \mathtt{e}\mathbf{B}$ are not affected by Ohmic dissipation, but it instead decreases the lowest contour level. The averaged effect of dissipation must match that of the generation by stretching, the field being energetically critical.

It is an important result that all the critical fields show the same characteristics; indeed, it might have been thought that the introduction of an inner core, removing the origin as a location for strong stretching, might fundamentally change the selected field symmetry. This is demonstrably false however, and we may conclude that the mechanism for the onset of magnetic energy growth by axisymmetric poloidal convection is robust, with or without an inner core.

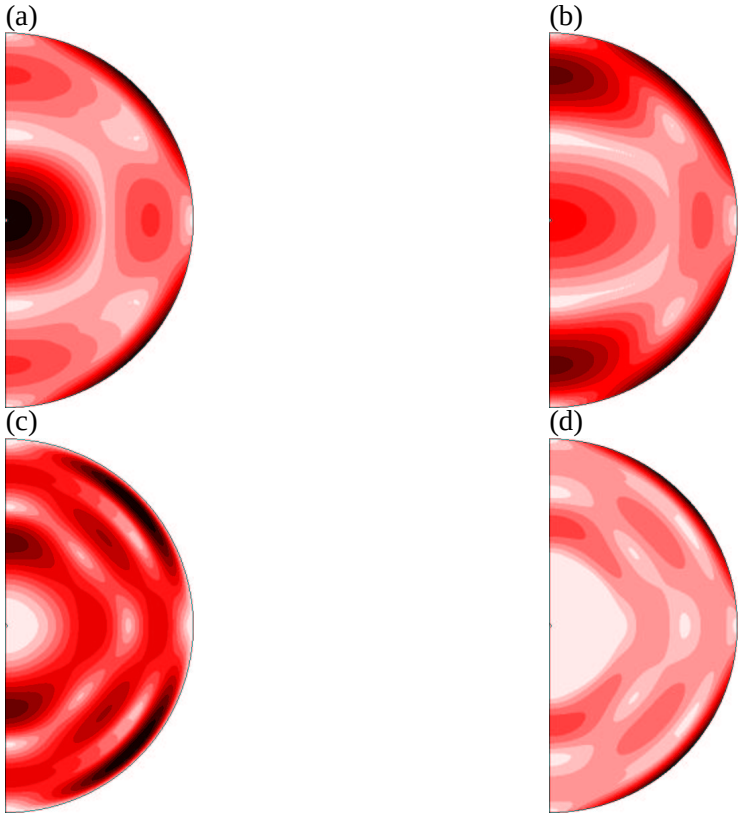


Figure 5.11: In a meridian plane, (a) and (b) show contours of μ for the forward and reversed $\mathbf{s}_2$(MDJ) flows respectively. (c) and (d) show contours for the $\mathbf{s}_2$(KR) and $\mathbf{s}_2$(IC) (forward) flows respectively. The flows are normalised to have a unit rms value. Contour levels are (a) $0.063 \leq \mu \leq 10.027$ (b) $0.050 \leq \mu \leq 10.027$ (c) $0 \leq \mu \leq 8.792$. (d) $0 \leq \mu \leq 21.193$. Red shows positive values.

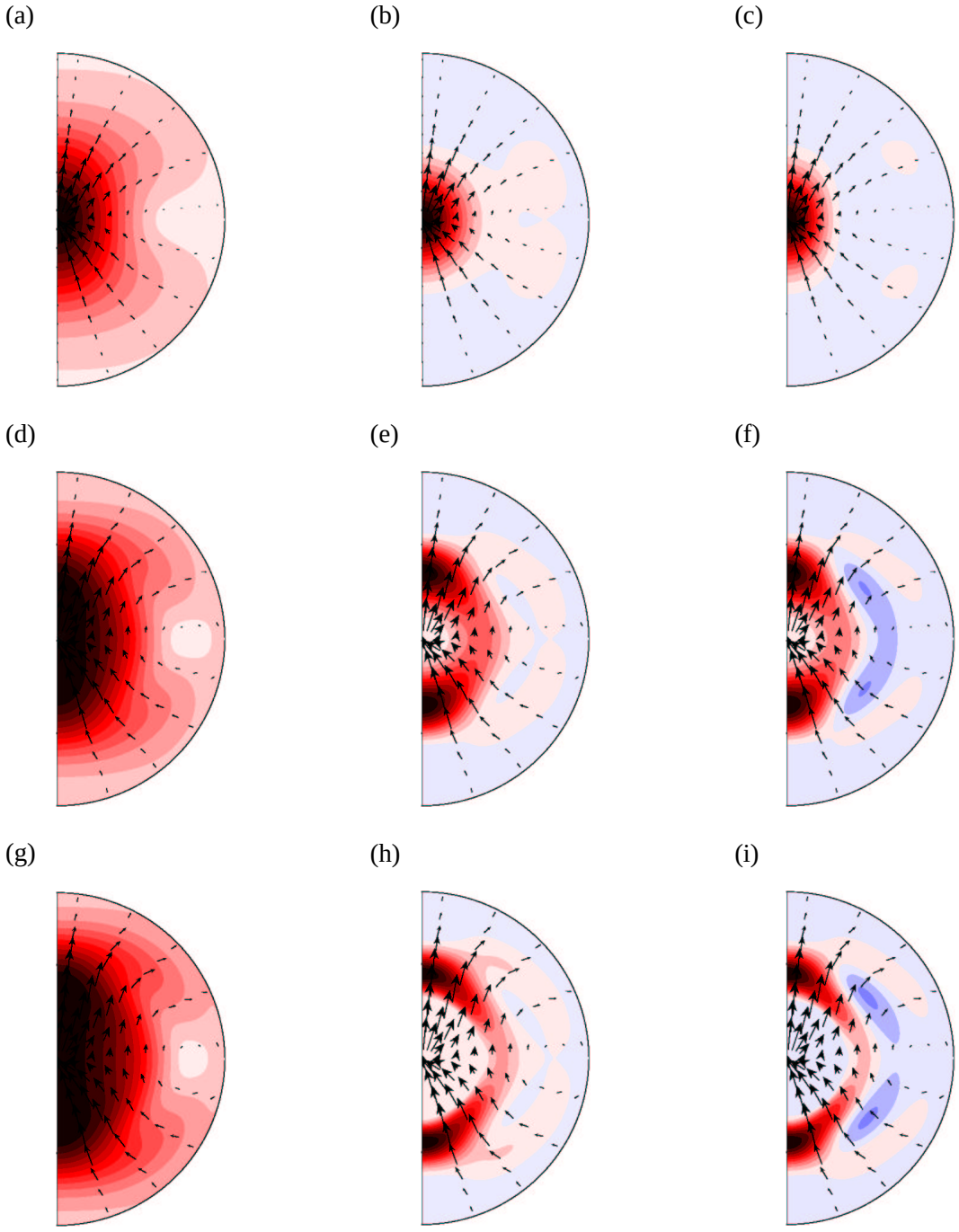


Figure 5.12: The field at R_m^E in a meridian plane for the three $\mathbf{s}_2$ flows: MDJ (row 1), KR (row 2) and IC (row 3). Column 1 gives contours of $|\mathbf{B}|$, column 2 contours of $R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B}$ and column 3 contours of $R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B} - |\boldsymbol{\nabla}\times\mathbf{B}|^2$. Red is positive, blue negative. Contour levels are as follows:
(a) $0 \le |\mathbf{B}| \le 3.021$, (b) $-14.011 \le R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B} \le 393.151$, (c) $-29.762 \le R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B} - |\boldsymbol{\nabla}\times\mathbf{B}|^2 \le 394.482$
(d) $0 \le |\mathbf{B}| \le 1.919$, (e) $-10.023 \le R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B} \le 144.687$, (f) $-23.056 \le R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B} - |\boldsymbol{\nabla}\times\mathbf{B}|^2 \le 110.087$
(g) $0 \le |\mathbf{B}| \le 1.609$, (h) $-8.528 \le R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B} \le 150.933$, (i) $-23.937 \le R_m^E \mathbf{B}\cdot\mathtt{e}\mathbf{B} - |\boldsymbol{\nabla}\times\mathbf{B}|^2 \le 106.726$.
Arrows denote the direction of the field which is normalised to have unit energy.

always have $R_m > 0$. Such lack of symmetry separation then does not concern us.

5.8 $\mathbf{t}_1\,\mathbf{s}_2$ flows

Flows of $\mathbf{t}_1\mathbf{s}_2$ type are combinations of the flows in the previous two sections. We find linear instability (the existence of R_m^c) only for a narrow range of τ: $0.3 \leq \tau \leq 0.9$ in MDJ.

Figures 5.13(a) and 5.13(b) compare λ_E for the various symmetries with $\tau = 0.5$ for the (a) $\mathbf{t}_1\mathbf{s}_2$(MDJ) and (b) $\mathbf{t}_1\mathbf{s}_2$(KR) flows. Notice that they are almost identical for $R_m > 0$ and fairly similar for $R_m < 0$. The effect of the vertical straining motion in the forward flow, at or near the origin, still has a large effect on the solution giving rise to the dominant $m = 0$ E^A symmetry. The value $\tau = 0.5$ was chosen since R_m necessary for the $\mathbf{t}_1\mathbf{s}_2$(MDJ) flow to exhibit linear instability ($\Re(\lambda) = 0$) was minimal ($R_m^c = 55.0$), see figure 5.5(a). The dominant

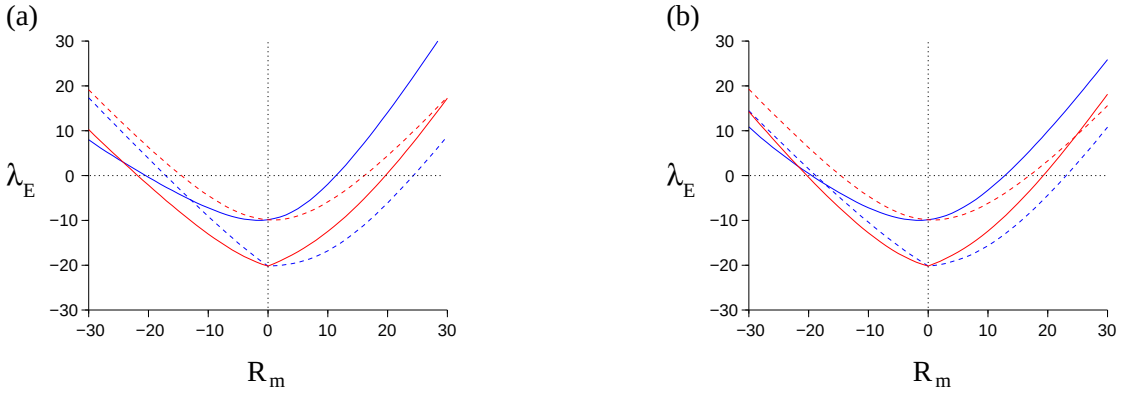


Figure 5.13: Field symmetry separation in energy instability of the (a) $\mathbf{t}_1\mathbf{s}_2$(MDJ), $\tau = 0.5$ and (b) $\mathbf{t}_1\mathbf{s}_2$(KR), $\tau = 0.5$ flows. λ_E is plotted as a function of R_m. Red is $m = 1$, blue $m = 0$, solid E^A, dashed E^S.

symmetry of λ_E is preserved at higher values of R_m. In particular at $R_m = 55.0 (= R_m^c)$ the energetically most unstable symmetry remains $m = 0$ E^A. The structure of the most unstable induction equation eigenmode ($m = 1$ E^S) therefore differs fundamentally from that of energetic analysis ($m = 0$ E^A) even at the same R_m.

The critical symmetry at $\tau = 0$ (where $\mathbf{u} = \mathbf{t}_1$) and $\tau = \infty$ (where $\mathbf{u} = \mathbf{s}_2$(MDJ)) differ, being $m = 1$ E^S and $m = 0$ E^A respectively. We now investigate at what value of τ this crossover occurs. Geophysically we observe that the Earth's field is dominated by the latter symmetry, so it is of interest to quantify how large τ must be in order for this to become the most energetically favourable.

We would like to study not only the $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) flow where each component has a positive sign, but those in which the relative signs differ. In fact, if we write $-\tau$ to represent an

energy ratio of τ but where the two components differ by sign then

$$\mathbf{u} = -R_m(\mathbf{t}_1 + \tau \mathbf{s}_2) = R_m(\mathbf{t}_1 - \tau \mathbf{s}_2) \tag{5.33}$$

since a change of sign in $\mathbf{t}_1$ only affects the direction of $\hat{\phi}$ of which $\mathbf{s}_2$ is independent. Therefore negative values of τ merely correspond to a reversed flow direction. Figure 5.14 shows a plot of R_m^E against τ for the $m = 0$ field symmetry (solid) and the $m = 1$ symmetry (dashed). For $\tau > 0$

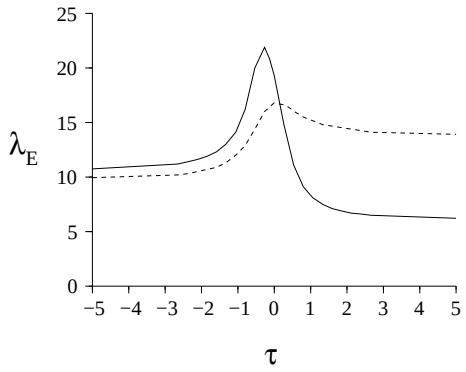


Figure 5.14: R_m^E plotted as a function of τ for the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow, solid is $m = 0$ and dashed is $m = 1$. Values of $\tau < 0$ represent reversed flows with a velocity component ratio of $|\tau|$. $\tau \geq 0.2$ gives the critical energetic symmetry as E^A $m = 0$.

the critical symmetry (the minimum of the two curves) switches at around $\tau = 0.2$. For $\tau < 0$ the $m = 1$ symmetry dominates. Geophysically, this means that in order to obtain an energetically unstable field of $m = 0$ E^A symmetry we need $\tau > 0.2$, which is easily achievable if convection dominates the flow pattern in the outer core.

5.9 The KR and STW flows

We now investigate more complex flows that do not give as great a symmetry separation as those preceding. Figure 5.15 shows λ_E as a function of R_m for the (a) KR and (b) STW flows. The coloured curves represent not just one value of m but a whole set: red curves contain odd values of m and blue even values of m. Solid denotes E^A symmetry and dashed E^S. In both cases, the critical symmetry is that of m odd and E^S, with $R_m^E = 15.8$ and $R_m^E = 13.6$ for the KR and STW flows respectively. Note the close similarity of both plots to that of the $\mathbf{t}_1$ flow in section 5.6 that has an identical critical symmetry and $\lambda_E = 16.8$. The fact that this $\mathbf{t}_1$ component makes up 99% and 96% by rms of the KR and STW flows respectively, demonstrates that the energetic instability method is again robust, in that small changes to the flow leave the energetic stability almost unaltered.

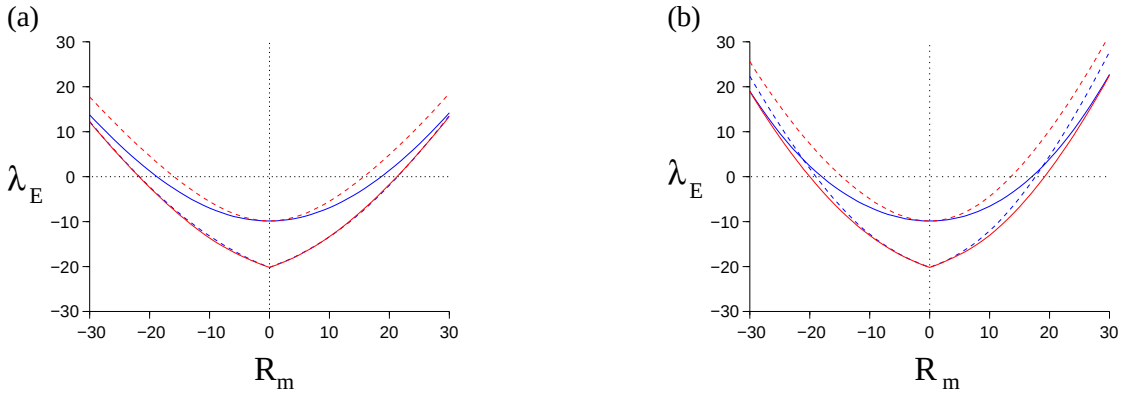


Figure 5.15: Field symmetry separation in energetic instability, for the (a) KR and (b) STW flows. λ_E is plotted as a function of R_m. Red contains odd values of m, blue contains even values of m; solid is E^A and dashed is E^S.

5.10 Improvements to Proctor's bound

Energetic instability is a necessary condition for magnetic field growth, and provides a lower bound on R_m for long-term dynamo action. Bounds of this nature play an important role, especially in physically realisable dynamo experiments where it is useful to have minimum size and flow velocity specifications. In table 5.2, we compare Proctor's (analytic) bound and the (numerical) values found in the previous sections to improve lower bounds on R_m for energetic instability. The value of μ_{max} is calculated using a grid search method.

Velocity	R_m^c	μ_{max}	$12.29/\mu_{max}$	R_m^E	Critical symmetry	R_m^{E*}
$\mathbf{t}_1$	-	4.44	2.77	16.8	$m=1$ E^S	74.6
$\mathbf{s}_2$(MDJ)	-	10.03	1.23	6.1	$m=0$ E^A	61.2
$\mathbf{s}_2$(KR)	-	8.83	1.39	7.4	$m=0$ E^A	65.3
$\mathbf{s}_2$(IC)	-	21.19	0.58	8.1	$m=0$ E^A	171.64
$\mathbf{t}_1\mathbf{s}_2$ (MDJ), $\tau=0.5$	55.0	5.38	2.28	11.4	$m=0$ E^A	61.3
$\mathbf{t}_1\mathbf{s}_2$ (KR), $\tau=0.5$	-	4.81	2.56	12.8	$m=0$ E^A	61.6
KR	890	5.62	2.18	15.8	$m=1$ E^S	88.8
STW	479	7.81	1.57	13.6	$m=1$ E^S	106.2

Table 5.2: Comparison of lower bounds on R_m for energetic stability. Second column shows R_m^c if it exists, the third shows μ_{max} of the flow, normalised to have unit rms value. The fourth shows the bound of *Proctor* (1977a) and the fifth shows the greatest lower bound R_m^E, with the associated critical symmetry in the sixth (in general this differs from that of linear analysis). Column seven shows R_m^E based on a flow with unit μ_{max} instead of unit rms.

The value of R_m^{E*} is the magnetic Reynolds number based on a flow of unit μ_{max} rather than of unit rms value, being a measure of how well the field can align itself with the stretching directions of the flow. It also depends on the distribution of μ in the flow: if the flow is everywhere static apart from one small region of vigour, $R_m^{E*} \gg R_m^E$ since the rms value of the flow

will be much smaller than the maximum strain rate. This is manifested in the relatively large value of R_m^{E*} for the $\mathbf{s}_2$(IC) flow, which has a reduced flow region in which to generate magnetic energy. Indeed, for a given maximum flow velocity, its rms value will be substantially less than that of any of the two other $\mathbf{s}_2$ flows, because of the presence of the inner core. This makes the process less efficient, a fact that is hidden using R_m based on the rms value of the flow.

In comparing Proctor's bound of $12.29/\mu_{max}$ with R_m^E, we achieve a gain of between 5 to 14 times, depending on the flow. Note the increase from *Livermore and Jackson* (2004) who showed an improvement of between 5 and 10 times, by the inclusion of the $\mathbf{s}_2$(IC) flow. Both choices of flow normalisation predict that the $\mathbf{s}_2$(MDJ) flow is the most energetically unstable. Generically, energetic instability is reached when R_m is $O(10)$, based on flows of unit rms.

5.11 Behaviour as $R_m \to \infty$

The subject of fast dynamos considers the ability of a flow to support magnetic instabilities in the limit $R_m \to \infty$. In eigenmode analysis, since a flow of a certain strength (and structure) is required in order to achieve instability, naively one might expect that higher values of R_m will serve only to increase it. However, this is not the case, and many flows support eigenmode instabilities only in a finite range of R_m, if at all. This is due to the pointwise field growth necessary for such a mode, being very sensitive to the exact form of flow, its strength, and the effect of diffusion. Numerical difficulties hamper studies at large R_m since not only are smaller and smaller length scales favoured leading to resolution problems, but the problem becomes increasing non-normal which makes it numerically much more difficult to find the eigenvalues. It is therefore not always easy to establish whether a flow is a fast dynamo or not. Much of fast dynamo theory uses a different scale for t in the induction equation: that of advection rather than diffusion time. This leads to a trivial rescaling of the results.

The behaviour of energetic instability however follows a more intuitive path: the stronger the flow, the greater the instability, apparently even in the asymptotic limit. In the range $-30 \leq R_m \leq 30$, many of the previous figures suggest that the dependence of λ_E on R_m is linear, and is independent of the flow choice and particular symmetry chosen. Figure 5.16 shows a plot of λ_E as a function of R_m on an extended scale. Dotted is shown a linear trend $\lambda_E \propto R_m$, and dashed the upper bound of $R_m \mu_{max}$ from section 3.7.6. The curve reaches its apparent (linear) asymptotic dependence at around $R_m = 1500$, indicating that the bounds in 3.7.6 are sharp, in the context of the scaling of magnetic instability as a linear power law in R_m. Well converged solutions are obtained in this range (see chapter four). There is however a ceiling to how hard

we may push the analysis, which with a convergence tolerance of 9 s.f. is around $R_m = 3000$, achievable with a truncation level $N_{max} = L_{max} = 40$.

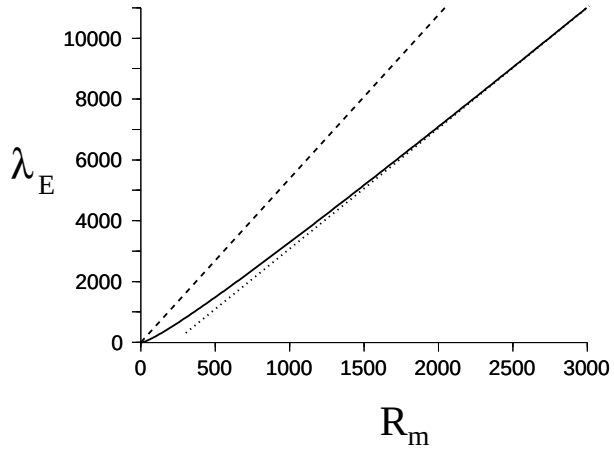


Figure 5.16: The value of λ_E is shown as a function of R_m for the $\mathbf{t}_1\mathbf{s}_2$(MDJ) flow with $\tau = 0.5$ (solid line). Dotted shows the asymptote that is reached at around $R_m = 1500$. Dashed is the upper bound on growth: $\lambda_E \leq R_m \mu_{max}$, from section 3.7.6.

5.12 Discussion

Overwhelmingly the most remarkable result found in this chapter is the robustness of the analysis regarding the onset of magnetic energy growth. Not only is λ_E insensitive to small flow perturbations, but the symmetry of the critical magnetic field remains unaffected. We may therefore explain the results by the same physical process, that of magnetic field line stretching by the flow.

We found that the $\mathbf{t}_1$, KR and STW flows had almost identical magnetic energy instability, all being roughly of the same flow structure, KR and STW differing only by 1% and 4% respectively from $\mathbf{t}_1$, by rms. The critical field symmetry was that of m:odd E^S. By contrast the linear stability characteristics are very different: $\mathbf{t}_1$ can never support growing eigenmodes but KR and STW can. This means that no robust physical interpretation for such an analysis can exist.

The three $\mathbf{s}_2$ flows, representing convection in the Earth's outer core, all exhibited very similar energy growth mechanisms, having the same critical symmetry of $m = 0$ E^A. In fact, $\mathbf{S}_1$ was the largest component of these fields, dominating also in the Earth. Eigenmode analysis cannot offer an explanation of this, due to Cowling's theorem, barring the infinite sustenance of all axisymmetric fields. However, we have shown here that these are the most energetically unstable in plausible geophysical flows, and could therefore go some way in explaining their presence in

the Earth's field. When adding in a $\mathbf{t}_1$ component into these $\mathbf{s}_2$ flows, provided that the sense of convection was geophysical ($R_m > 0$) and that the poloidal component was sufficiently large ($\tau > 0.2$) then again the Earth-like $m = 0$ E^A magnetic field symmetry dominates. This again flags the robustness of this analysis.

It is interesting to compare the sensitivity of this methodology with that of eigenmode growth: this scheme is robust and really only takes into account the large scale flow. In contrast, the eigenvalue spectrum depends critically on the choice of flow. The flow driving the geodynamo in the outer core plausibly has a significant large scale $\mathbf{s}_2$ convective component. It is extremely unlikely that the flow is everywhere the same as a kinematic dynamo model that we may study, for which growing eigenmode solutions might be found; only robust characteristics can be relied upon. That the geophysically dominant $m = 0$ E^A field symmetry is the most unstable at magnetic energy growth onset is significant, especially when this is insensitive to the particular choice of convective flow used. Thus such a mechanism may partially explain how the Earth's field came to be of its current structure: if the flow is turbulent (so that traditional large-scale eigenmode theory does not apply) and constantly feeds energy into the field, it may be that this field symmetry is continuously generated. In any case, in view of the success of non-normal theory in explaining transition to turbulence in the hydrodynamic case, this process may explain how magnetic fields grow insensitively to the choice of flow, and lead into the non-linear regime where the Lorentz force is important.

As R_m is increased, we find that λ_E asymptotically increases linearly. Physically this means that the faster the flow, the easier it is to generate magnetic field energy. In contrast, if unstable eigenmodes do not exist, the eigenvalues tend to become increasingly negative with R_m, implying that the generation mechanism is less efficient. We found that the asymptotic linear scaling for λ_E is reached at $R_m = 1500$ which was well within our convergence range.

Converged eigenvalue calculations become increasingly harder to obtain as R_m increases, due not only to the higher truncation level required but the increased non-normality of the system. This means that progressively larger subsets of the complex plane become numerically close to the true eigenvalues (i.e. $|\mathsf{A}\mathbf{v} - \sigma\mathbf{v}| \approx 0$ for the controlling matrix A and some $\sigma \in \mathbb{C}$, where σ may not be within the convergence tolerance of the true eigenvalue λ). On increasing the truncation levels, small numerical errors will start to creep in and thus imprecise eigenvalues will be reported.

We have dwelt in this chapter on the onset of magnetic energy instability, ignoring questions such as how large the magnetic energy might become and over what timescale it grows. These are addressed in chapter six.

Chapter 6

Transient magnetic energy growth

6.1 Transient growth

The idea that transient effects could be of importance in the geodynamo is not new, indeed both *Moffatt* (1978) and *Childress and Gilbert* (1995) discuss non-normal growth. In fact, the notion that the eigenmodes may not fully represent the capability of a flow to support growing magnetic fields was first proposed by *Backus* (1958). He considered a cyclic time sequence of various stationary flows, each linearly stable to magnetic fields. A period of stasis was imposed at each flow change-over, killing off high degree field harmonics. An initial field grew transiently in energy, and before it entered its decaying phase the flow changed and transient growth began again. At not all parts of the cycle did magnetic field grow: in some the field was merely transformed from one structure to another, for example, by a solid body rotation. Nonetheless, despite at no time this flow being able to support growing eigenmodes, this kinematic dynamo mechanism shows the importance of non-normal effects.

In general, even if a flow is excluded from indefinitely sustaining a magnetic field by the antidynamo theorems of section 2.5 (i.e. all eigenvalues are stable), the field may decay over such a long period of time that we may not be able to exclude them as being geophysically or astrophysically irrelevant (*Ivers and James*, 1984). Such a situation was considered by *Backus* (1957) who studied axisymmetric poloidal fields, in an axisymmetric flow, necessarily decaying by the application of Cowling's theorem. He found that the decay time was up to four times longer than that of static diffusion, corresponding to around $80,000$ years. Plausibly, by this time the flow pattern could have changed and the field may have begun to grow again. Therefore, ruling out such a situation on linear stability grounds could give misleading or even incorrect results.

In the Earth where the flow structure in the outer core is undoubtedly turbulent and time dependent, it is difficult to rule out any dynamo mechanism on eigenmode stability grounds, since not only does such an analysis depend on the flow being stationary, but transient effects are ignored. Instead, two approaches may be taken. Firstly, we can try to decompose such a flow into a time sequence of stationary flows, as in *Backus* (1958). We may then investigate magnetic energy growth under each fixed flow; if the growth is robust under small flow perturbations and occurs over a period of time where the stationary approximation is valid, then this might be at least part of the mechanism which sustains the Earth's field. In chapter one, we saw that after magnetic reversals, the field intensity along with its axisymmetric dipole component recovers rapidly over a timescale of around 10,000 years. No linear eigenvalue mechanism can supply a mechanism since axisymmetric fields are barred from growing exponentially; perhaps transient growth could offer an explanation? From a geomagnetic observational point of view, non-normal growth may explain other phenomena. Observations at Steens Mountain, Oregon (*Mankinen et al.*, 1985), indicated very fast (on a timescale of days or weeks) field intensity and directional changes during a field transition. Perhaps some kind of ephemeral linear magnetic field growth might account for this.

Alternatively, if the flow is statistically steady, it may be argued that transient growth could be continually fed by the flow, giving rise to a dynamo mechanism. It is not the linearly unstable magnetic fields which are manifested, but so called pseudo-modes, those field structures which grow transiently the most. This has been investigated by *Farrell and Ioannou* (1999b) in a cylindrical geometry, where they find that axisymmetric magnetic fields can be indefinitely sustained by stochastic forcing, despite the application of Cowling's theorem indicating the contrary.

In order the the concept of non-normal growth is clear, we first introduce a simple example.

6.1.1 A simple example

In a simple case, we show how the superposition of two decaying eigenmodes can produce transient growth. We consider first two such modes which are spatially orthogonal in the usual Euclidean inner product, for example $\mathbf{v}_1 = (0,1)\,e^{-2t}$ and $\mathbf{v}_2 = (1,0)\,e^{-t}$. Let us form some vector quantity from a combination of these two as $\mathbf{u} = \alpha\,\mathbf{v}_1 + \beta\,\mathbf{v}_2$; its energy E would then be

$$E = \mathbf{u}\cdot\mathbf{u} = \left[\alpha\,e^{-2t}\right]^2 + \left[\beta\,e^{-t}\right]^2 \tag{6.1}$$

which is shown in figure 6.1 as a dashed line, in the case when $|\alpha| = |\beta|$. The curve is normalised to have a maximum value of one. Secondly we consider two non-orthogonal eigenmodes, $\mathbf{v}_1 = (0,1)\,e^{-2t}$ and $\mathbf{v}_2 = (0,1)\,e^{-t}$. If $\mathbf{u} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2$ as before then the energy is

$$E = \mathbf{u}\cdot\mathbf{u} = \left[\alpha e^{-2t} + \beta e^{-t}\right]^2 \tag{6.2}$$

This is plotted as a solid line for the case $\alpha = -\beta = -1$ under the same normalisation, so that $\mathbf{u} = (0,1)\left(e^{-t} - e^{-2t}\right)$. The transient growth is caused by the initial destructive superposition

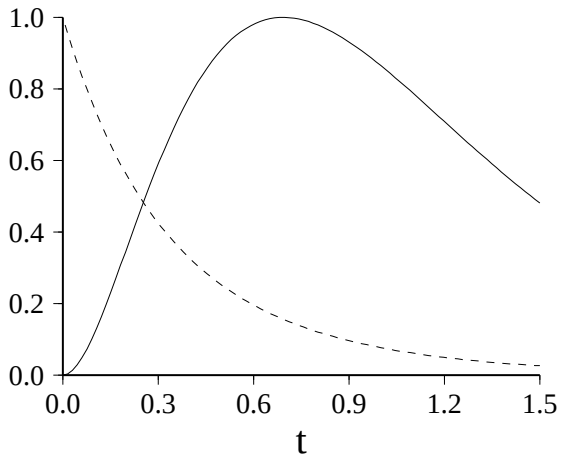


Figure 6.1: Energies in the cases of (a) orthogonal decaying eigenmodes (dashed) and (b) nonorthogonal decaying eigenmodes (solid). Each curve is normalised to have a maximum value of 1.

of non-orthogonal eigenmodes, combining to give zero initial energy. In this case, non-zero energy at any time $t > 0$ represents an infinite relative amplification from its initial value. This is a somewhat contrived case however, and in general only finite energy growth is possible since the eigenvectors are almost never exactly linearly dependent. Such transient behaviour is completely dependent on the initial conditions: transient decay is also possible by choosing $\alpha = \beta$ in the latter case. However, such effects will never be physically manifested and so we do not consider them here.

6.1.2 Envelopes

We now consider the maximal relative magnetic energy growth at some time t, a measure of the possible transient behaviour:

$$\sigma^2(t) = \max \frac{M(t)}{M(0)} = \max \frac{\frac{1}{2}\int_{V+\hat{V}} \mathbf{B}(t)^2\, dV}{\frac{1}{2}\int_{V+\hat{V}} \mathbf{B}(0)^2\, dV} \tag{6.3}$$

As already detailed in chapter four, this can be numerically computed by finding σ^2, where σ is the largest singular value of the matrix $\mathsf{F}\, e^{\mathsf{L}t}\, \mathsf{F}^{-1}$, L being the discretised induction operator and $\mathsf{F}^T\mathsf{F} = B_{V+\hat{V}}$, the Chebyshev basis energy inner product matrix. The optimal initial condition is given by $\mathsf{F}^{-1}\mathbf{v}$, where $\mathbf{v}$ is the corresponding right singular vector. If we plot σ^2 as a function of t, this defines an envelope of the possible relative energy growth. Every point on this curve is associated with an optimal initial condition which evolves to give this maximal growth; individual growth curves of a particular initial field with unit energy may touch the envelope tangentially at one or more points, but they may not cross it.

This is illustrated in figure 6.2 where we show the envelope of magnetic energy growth under the action of the $\mathbf{t}_1$ flow with $R_m = 1000$ (dashed); the evolution of two particular initial conditions is shown as solid lines, chosen to optimise the energy growth at $t = 0.001$ and $t = 0.1$. They touch the envelope at these points but in general it bounds them strictly from above. The relevant field symmetry is $m = 0\ E^A$ with $N_{max} = L_{max} = 16$. The initial condition chosen to

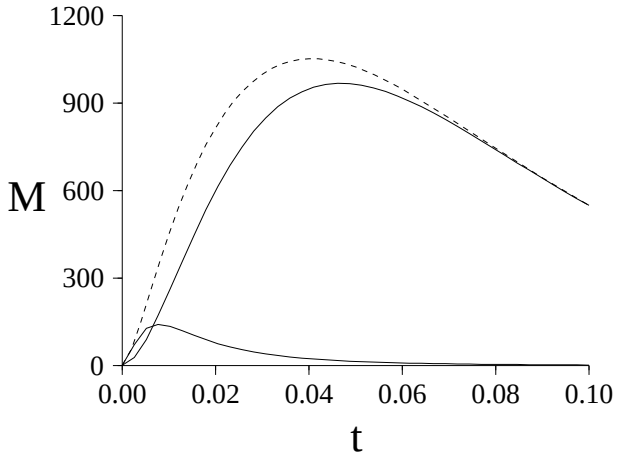


Figure 6.2: The magnetic energy envelope of the $\mathbf{t}_1$ flow with $R_m = 1000$ (dashed), and the evolution of two particular initial conditions (solid), chosen to optimise the energy growth at $t = 0.001$ and $t = 0.1$. The field symmetry is $m = 0\ E^A$ with truncation $N_{max} = L_{max} = 16$.

optimise energy growth for $t = 0.1$ apparently approximates the envelope for most times fairly well. In contrast, the optimal curve for $t = 0.001$ does not, and decays for $t > 0.01$.

We note explicitly that the initial gradient of a magnetic energy envelope is equal to the maximised instantaneous rate of magnetic energy growth subject to fields having unit energy. This is none other than twice the value of λ_E of chapter five. If $R_m < R_m^E$ so that $\lambda_E < 0$, the maximum gradient of the envelope will always be negative, i.e. it will decrease monotonically in time. If $R_m > R_m^E$ on the other hand, the envelope will initially increase, although may eventually decay. If $R_m > R_m^c$, should this linear critical magnetic Reynolds number exist, then the envelope will asymptotically increase exponentially.

6.1.3 Scaling laws

In relating such maximal growth calculations to the Earth, we are interested in how large the magnetic energy amplification can be and not necessarily at what time it is attained, so long as the timescales involved are geophysically relevant. It therefore makes no sense to seek $\sigma^2(t)$ for some fixed value of t; instead, we measure the scope of transient growth by the height of the envelope. It is in general not *a priori* possible to predict this height and the time at which it occurs, these only being known after the computation.

In general, computations at high R_m are numerically difficult since progressively smaller length scales become important and increased truncation levels result in very large matrix systems. In these cases, we try to fit quantities such as the height of the envelope peak and the time it is attained as power laws in R_m. If such a procedure can be carried out, this gives us more confidence of the behaviour at high R_m where convergence is difficult.

Scaling law analysis has been applied (see e.g. *Schmid and Henningson*, 2001) to the plane Poiseuille and plane Couette flows described in chapter five. Numerically it is found that the height of the energy envelope scales as R_e^2, that is, an initial disturbance of energy $O(R_e^{-2})$ grows to a perturbation of unit energy. Whether or not this final state leads to a transition to turbulence or not is another matter; in fact the energy of the initial field leading to the turbulent state is $O(R_e^{-2})$ and $O(R_e^{-3})$ respectively for these two simple flows, taking into account nonlinear interactions (*Chapman*, 2002). The height of the envelope peak is then a good benchmark for determining the effect of growing perturbations which lead to transition. In their study of transient magnetic energy growth in a cylindrical geometry, *Farrell and Ioannou* (1999a) found that the magnetic energy envelope height scales as $R_m^{2/3}$ and is attained at a time $O(R_m^{-2/3})$. It is expected that a spherical geometry has a significant effect on these scalings, although these indicate transient energy growth of $O(100)$ can occur when $R_m = 10^3$.

6.1.4 The rescaling of magnetic fields to the Earth

Since the induction equation is linear in **B**, our original non-dimensionalisation left the magnitude of the field unspecified. We now detail how we scale our numerically computed solutions to Earth-like values.

Ideally, since all of our calculations are normalised to have unit magnetic energy, we would simply scale the computed magnetic energy by that of the Earth; unfortunately this is unknown. Instead, we normalise the initial field to take on an external dipole coefficient for example, g_1^0, matching that of the observational inverse model detailed in Appendix A. There is in general always a choice between the three such dipole coefficients; we choose to scale by which ever is the largest in the numerical solution. For example, in many cases the $m = 0\, E^A$ symmetry dominates and so we match the associated value of g_1^0 to the Earth. At the Earth's surface ($r = 1.83$ in our units), the numerical solution gives

$$-29616.13\,\mathrm{nT} = g_1^0 = 1.83^{-3}\, S_1(1) \tag{6.4}$$

using equation 4.24 and Appendix A. This means that we must scale to $S_1(1) = 181502\,\mathrm{nT}$.

The timescale that we use is that of diffusion time $\mathscr{T}_d = \mathscr{L}^2/\eta_0$, one unit of which corresponds to 192,000 years. This figure is computed using the length scale $\mathscr{L}$ = 3485km, the CMB radius, and the core magnetic diffusivity of $2\,\mathrm{m^2\,s^{-1}}$. The e-folding time of the slowest decaying magnetic field (of $l = 1$ poloidal form) is π^{-2}, approximately 20,000 years in dimensional terms. We contrast this with the turnover time for the core of around 500 years, based on the 'frozen flux' flow speed estimate of $2 \times 10^{-4}\,\mathrm{m\,s^{-1}}$ and length scale $\mathscr{L}$.

6.2 The transient magnetic energy amplification of the $\mathbf{t}_1$ flow

6.2.1 Symmetry separation

As before, the different field symmetries which separate under this flow may be studied independently. We show a comparison at fixed R_m in figure 6.3 of the magnetic energy envelopes. Blue shows the $m = 0$ symmetry, red $m = 1$, solid E^A and dashed E^S. It is apparent that the blue curves ($m = 0$ field symmetries) dominate over the timescale shown. A non-dimensional time of 0.2 corresponds to 2 dipole diffusion times (about 40,000 years in dimensional terms). In fact it is the axisymmetric dipole symmetry that attains the highest envelope peak and it is this which we shall study further. At small times a different field symmetry becomes important: instantaneous energy growth calculations indicate that the critical field symmetry is $m = 1\, E^S$ (section 5.6). It

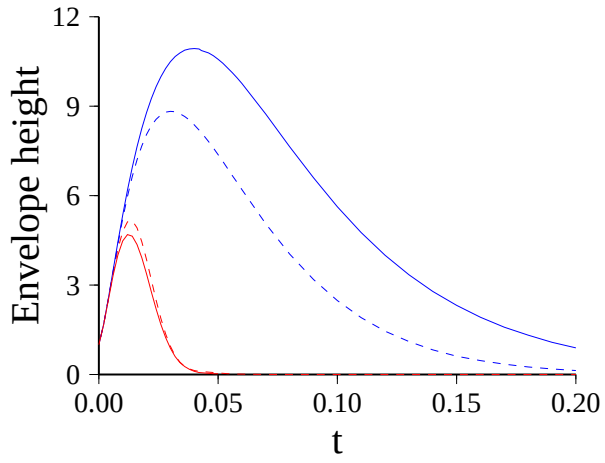


Figure 6.3: Energy envelope calculations for different symmetries at $R_m = 100$, truncation level $N_{max} = L_{max} = 24$. Blue is the $m = 0$ symmetry, red is $m = 1$, solid is E^A, dashed is E^S.

is therefore of note that the critical field symmetry depends on the timescale we are interested in.

6.2.2 Convergence

The convergence in truncation of the envelope height is shown in table 6.1, with the time at which it is attained in parentheses. We use a simple bisection algorithm in time to determine the height.

Interestingly the necessary values of $N_{max} = L_{max} = 8$ (shown in bold) do not change with increasing R_m, indicating that some large scale field is the optimum initial condition that maximises the envelope height for all values of R_m. It would also appear that the time at which this envelope maximum occurs is also independent of R_m, at around $t = 0.0407$ (7,800 years).

We note the very different truncation levels required to calculate λ_E in section 5.6, requiring $N_{max} = L_{max} = 32$ for $R_m = 10^3$. This indicates that the spatial structure of the initial condition giving rise to optimal magnetic energy growth is a function of time. Vanishing small t (corresponding to λ_E) and $t > 0$ (corresponding to $\sigma^2(t)$) have very different associated optimal initial field structures.

6.2.3 Scaling of the envelope height with R_m

We now investigate how the height of the envelope for the dominant $m = 0$ E^A symmetry and its associated optimising time vary, as functions of R_m. Figures 6.4(a) and (b) show respectively the peak of the energy growth envelope as functions of R_m and the associated optimising time (the time at which the maximum is attained). On the log-log scale, the graph of figure 6.4(a) is

Truncation	Envelope height and time attained		
$N_{max} = L_{max}$	$R_m = 10^2$	$R_m = 10^3$	$R_m = 10^4$
2	8.747978952	834.7320734	83433.80125
	(4.91986E-2)	(5.08927E-2)	(5.09095E-3)
4	10.92474901	1051.700336	105129.8025
	(3.98367E-2)	(4.08643E-2)	(4.08746E-2)
8	**10.93463857**	**1052.636658**	**105223.3837**
	(3.96589E-2)	(4.06905E-2)	(4.07004E-2)
16	10.93463857	1052.636657	105223.3837
	(3.96587E-2)	(4.06901E-2)	(4.06993E-2)
24	10.93463840	1052.636650	105223.3837
	(3.96654E-2)	(4.06950E-2)	(4.06990E-2)

Table 6.1: Convergence in truncation of the envelope height; the time at which it is attained is shown in parentheses. Bold indicates convergence has been obtained.

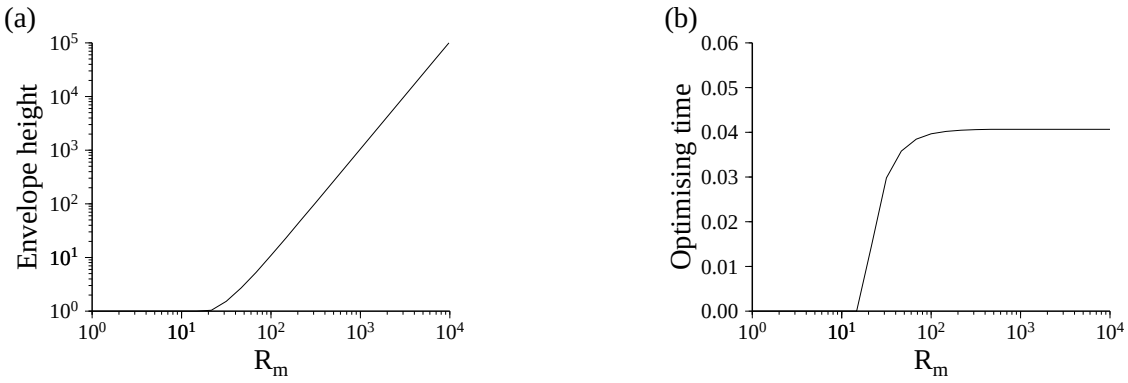


Figure 6.4: For the $\mathbf{t}_1$ flow, $m = 0$ E^A field symmetry, truncation level $N_{max} = L_{max} = 24$. (a) shows the height of the energy envelope as a function of R_m, (b) shows the time at which it is attained.

almost perfectly linear for $R_m \geq 100$. This indicates a power-law scaling which we compute to be $1.052 \times 10^{-3}\, R_m^2$. The optimising time tends to a constant value of 0.0407 as R_m is increased; the initial field configuration also converges and becomes independent of R_m for $R_m \geq 100$.

6.2.4 The nature of the transient growth

We now investigate the form of this transient behaviour. Figure 6.5(a) shows the toroidal (dashed) and poloidal (solid) energies respectively, for the time evolution of the initial magnetic field with unit energy chosen to maximise the energy at $t = 0.04$, $R_m = 100$, for the $m = 0$ E^A field symmetry. A snapshot of $\mathbf{B}_\phi$ is shown in figure 6.5(b); the field is dominated by its $\mathbf{T}_2$ component.

The poloidal energy decays since there is no process to sustain it; it is the toroidal energy which grows by the shearing action of the flow on poloidal field lines. Notice that its

initial value is 0 so that the optimal initial field configuration is poloidal only (see figure 6.6(a)).

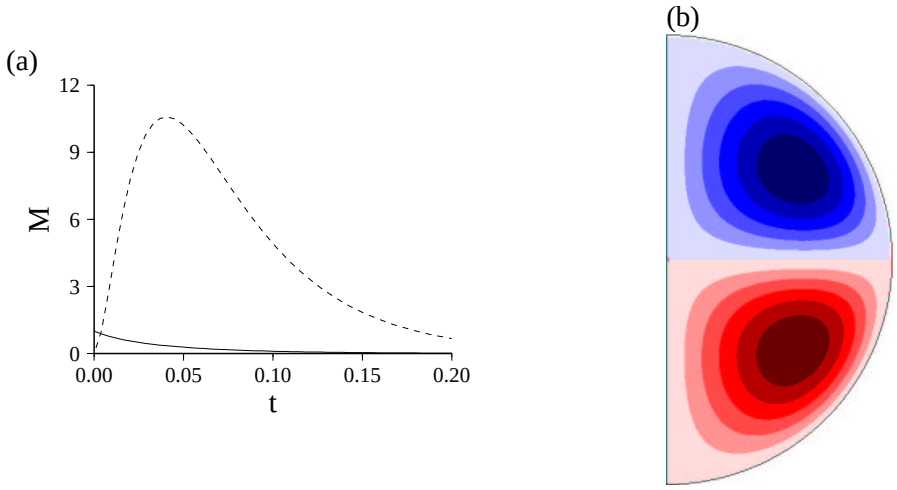
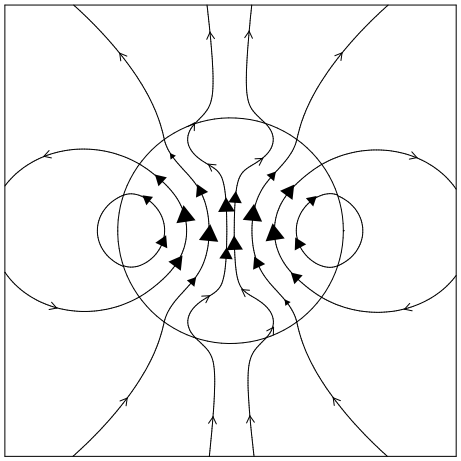


Figure 6.5: (a) Time evolution of the toroidal (dashed) and poloidal (solid) energy of the initial field with unit energy chosen to maximise the total energy at $t = 0.04$. Calculated with $N_{max} = L_{max} = 16$ for the $m = 0\ E^A$ field symmetry with $R_m = 100$.
(b) A snapshot of $-0.900 \leq \mathbf{B}_\phi \leq 0.900$ at $t = 0.04$. The field is dominated by $\mathbf{T}_2$, and normalised to have unit energy. Red shows positive contours, blue negative.

6.3 Initial fields

A comparison of the initial fields which maximise the envelope height for $R_m \geq 100$ shows that they are (i) independent of R_m (ii) of $m = 0\ E^A$ symmetry and (iii) of poloidal-only structure so that field lines are constrained to lie in meridian planes, shown in 6.6(a). Figure 6.6(b) shows the fraction of poloidal energy comprising the initial fields as a function of optimising time for $R_m = 50$. The solid line represents the $m = 0\ E^A$ symmetry and dashed the $m = 1\ E^S$ symmetry. To interpret figure 6.6(b) we first isolate the solid curve. At small $t < 10^{-3}$, the energy splits roughly into two, between its toroidal and poloidal components. As the optimising time is increased, the initial field becomes increasingly more poloidal until it reaches total saturation at around $t = 0.04$. The dashed curve representing the $m = 1\ E^S$ field symmetry behaves similarly. We therefore have two regimes: for small t, the optimal initial field is of similar configuration to that of instantaneous energy growth (i.e. that as $t \to 0$); the energy growth mechanism is then one of field line stretching. For large t, the energy growth process is one of finite time shearing, being far more efficient at generating magnetic energy over these time periods than merely field line stretching. The initial fields are poloidal-only fields which are configured such that (i) they can be optimally sheared and (ii) they decay the slowest. It is no surprise therefore that they are of a



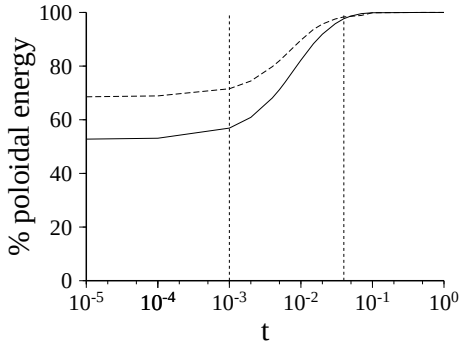


Figure 6.6: (a) Field lines of the optimising initial field, maximising the envelope height for $R_m \geq 100$ for the $m = 0\ E^A$ field symmetry. The truncation is $N_{max} = L_{max} = 24$.
(b) Percentage poloidal field energy of the initial field versus optimising time, for $R_m = 50$, $N_{max} = L_{max} = 16$. Solid is the $m = 0\ E^A$ field symmetry; dashed $m = 1\ E^S$. Vertical dotted lines mark $t = 0.001$ and $t = 0.04$.

similar structure to the $l = 1$ poloidal decay mode, shown in figure 6.6(a).

It is interesting to note that the length scales involved in this problem are all reasonably large, requiring much lower truncations than typical eigenmode computations for the same R_m. Typically high values of R_m mean that small length scales become ever more important; in this case however it is not true. The initial fields associated with an optimising time $t \ll 1$ are those of instantaneous magnetic energy instability and are of large scale, a property which is preserved in their swift decay (figure 6.2). For $t \gg 1$, the initial fields approximate the $l = 1$ poloidal decay mode and the $\mathbf{T}_2$ field harmonic dominates the subsequent behaviour, both of which being large scale.

6.3.1 A simple model

We can understand the physical mechanism by considering the simple model

$$\frac{dS}{dt} = -d_P^2 S \tag{6.5}$$

$$\frac{dT}{dt} = R_m S - d_T^2 T \tag{6.6}$$

where the scalars $S(t)$ and $T(t)$ represent the poloidal and toroidal field components respectively, $-d_P^2$ is the slowest poloidal decay rate and similarly for the toroidal decay rate. We model the shearing of poloidal field into toroidal field by the simple source term $R_m S$.

The equations are trivially solved to be

$$S(t) = Ae^{-d_P^2 t} \tag{6.7}$$

$$T(t) = \frac{R_m A}{-d_P^2} e^{-d_P^2 t} - Be^{-d_T^2 t} \tag{6.8}$$

Suppose now we want to find the maximal energy growth and the time at which it is attained. It is clear that the poloidal contribution merely decays and that the field will be dominated (as expected), particularly for large R_m, by the toroidal component T. If the initial field is chosen such that $S(0) = 1$ and $T(0) = 0$ then

$$S(t) = e^{-d_P^2 t} \tag{6.9}$$

$$T(t) = \frac{R_m}{-d_P^2} \left[e^{-d_P^2 t} - e^{-d_T^2 t} \right] \tag{6.10}$$

We seek the maximum of the dominant toroidal component, comprising the difference of two decaying exponentials, the archetypal model of transient growth of section 6.1.1.

It is clear that the maximum toroidal energy scales as R_m^2 and that the maximising time is constant, calculated from

$$\frac{d}{dt} \left\{ e^{-d_P^2 t} - e^{-d_T^2 t} \right\} = 0 \tag{6.11}$$

which is

$$t = \frac{ln\left(\frac{-d_P^2}{-d_T^2}\right)}{-d_T^2 - (-d_P^2)} \approx 0.069 \tag{6.12}$$

Although this model is rather simplistic, it explains two important features of transient growth under the $\mathbf{t}_1$ flow. Firstly, the optimal magnetic energy growth scales as R_m^2; secondly, the optimising time is independent of R_m, for $R_m \gg 1$, which is reasonably close to the numerically calculated value of 0.04. The mechanism responsible is that of shearing by differential rotation, creating toroidal field out of poloidal field. We note that the flow of energy is one way: there is no process sustaining the poloidal field. We do not expect such simple scalings to apply in more complex cases, where both field components interact in a complicated fashion.

6.4 The transient magnetic energy amplification of $\mathbf{s}_2$ flows

6.4.1 Symmetry separation

Figures 6.7(a)–(c) show envelope calculations for the four independent symmetries as before. Red shows $m = 1$ (calculated with $N_{max} = L_{max} = 24$) and blue $m = 0$ (calculated with $N_{max} =$

$L_{max} = 40$). Solid denotes the E^A symmetry, dashed E^S. We use different truncations since the $m = 1$ symmetries contain twice as many harmonics as the $m = 0$ symmetries; however all calculations are satisfactorily converged. We show plots at $R_m = 100$ of (a) the $\mathbf{s}_2$(MDJ) flow, (b) the $\mathbf{s}_2$(KR) flow and (c) the $\mathbf{s}_2$ (IC) flow.

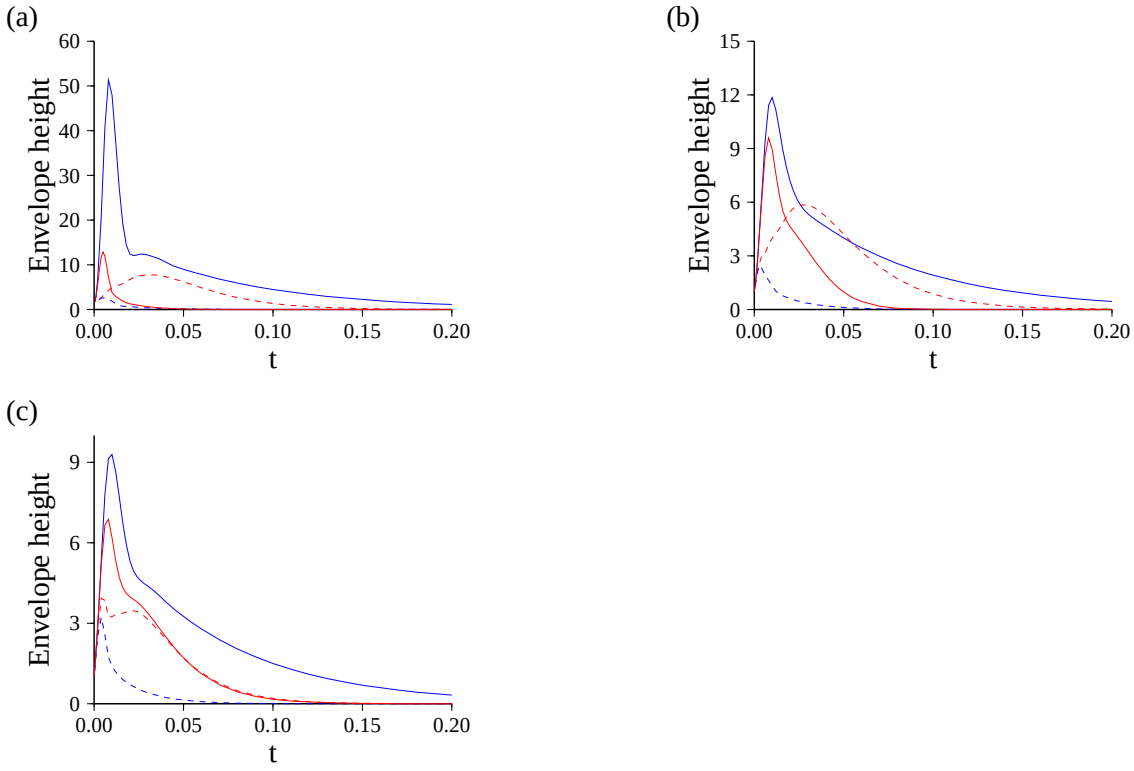


Figure 6.7: Energy envelope calculations for different field symmetries for (a) the $\mathbf{s}_2$ (MDJ) flow (b) the $\mathbf{s}_2$(KR) flow and (c) the $\mathbf{s}_2$ (IC). Blue is the $m = 0$ symmetry, red is the $m = 1$ symmetry, solid is E^A, dashed is E^S. In each case $R_m = 100$.

In each case the $m = 0\, E^A$ field symmetry is the most dominant, identical to the analysis of instantaneous energy growth (section 5.7). There are several 'kinks' in the above curves (e.g. in the solid blue curve of figure 6.7(a) and the dashed red curve of figure 6.7(c)). These are not convergence errors, but can be understood by the envelope being the maximum of two different curves, one describing optimal growth for small times and one for large times. The $\mathbf{s}_2$(MDJ) flow can give a magnetic energy amplification of 52, the KR flow of 12 and the IC flow of 9. It would appear that the behaviour at the origin is of great importance in determining this maximal amplification: the more vigorous the flow at or near the origin, the greater the potential for large growth. Recall that the $\mathbf{s}_2$(KR) flow has not only a stagnation point but vanishing strain rates at the origin, and that $\mathbf{s}_2$(IC) is static for $r \leq 0.35$. Despite the differences in the flows however, since the favoured symmetry is the same, we might expect similar physical growth mechanisms

to apply.

6.4.2 Convergence

We now determine for the dominant $m = 0$ E^A field symmetry, the envelope height for the $\mathbf{s}_2$(MDJ) flow as a function of R_m and truncation level, shown in table 6.2. Unfortunately it appears that out of all the points on the envelope, this maximum is the least well converged. However, if we do manage to show convergence we may have faith that the whole envelope is converged. We note that $R_m = 100$ does not converge as well as we might like, especially considering the generic swift convergence in chapter five of λ_E, and of the $\mathbf{t}_1$ envelopes in section 6.2. We can only be confident that the values are correct to 4 significant figures, sufficient for graphical purposes.

Immediately we see a dramatic difference in behaviour from the envelopes of the $\mathbf{t}_1$ flow. Apparently the initial fields that maximise the magnetic energy growth decrease in length scale as R_m increases, and good convergence is not achieved after $R_m = 46.42$. We see also an increase of envelope height with R_m and a decrease in the associated optimising time.

The convergence can also be illustrated in the power spectra of the relevant initial field. We consider the energy contained in the different harmonics:

$$M_l = \frac{1}{2} \int_{V+\hat{V}} \mathbf{B}_l^2 \, dV \tag{6.13}$$

where $\mathbf{B}_l$ is the magnetic field restricted to spherical harmonics of degree l. We also introduce the surface magnetic energy spectra given by

$$SE_l = \frac{1}{2} \int_{\partial V} \mathbf{B}_l^2 \, d\Omega \tag{6.14}$$

For this flow, the optimal initial conditions are poloidal only. Figure 6.8 shows energy spectra of the optimal initial conditions for the $\mathbf{s}_2$(MDJ) flow, with $R_m = 46$ (solid), $R_m = 100$ (dotted) and $R_m = 464$ (dashed). We note that as R_m increases, the solutions become less and less well converged as small length scales become increasingly important.

We lastly note that this convergence behaviour is seen also in the envelope heights for the other $\mathbf{s}_2$(KR) and $\mathbf{s}_2$(IC) flows.

	Truncation $N_{max} = L_{max}$			
R_m	16	24	32	40
6.813	**1.046157012**	1.046157012	1.046157012	1.046157010
	1.88079E-2	1.88081E-2	1.88064E-2	1.88051E-2
10.00	**1.653049544**	1.653049548	1.653049545	1.653049544
	3.74097E-2	3.74103E-2	3.74084E-2	3.74079E-2
14.68	3.118450364	**3.118450707**	3.118450713	3.118450681
	3.61503E-2	3.61501E-2	3.61495E-2	3.61475E-2
21.54	5.876564026	**5.876590843**	5.876590860	5.876590800
	2.88406E-2	2.88404E-2	2.88406E-2	2.88394E-2
31.62	10.60415170	10.60509048	**10.60509078**	10.60509079
	2.15601E-2	2.15604E-2	2.15604E-2	2.15600E-2
46.42	8.413779906	18.42991952	**18.42995688**	18.42995693
	1.58105E-2	1.58176E-2	1.58173E-2	1.58172E-2
68.12	31.01706535	31.16944751	31.17112047	31.17112651
	1.15211E-2	1.15470E-2	1.15472E-2	1.15468E-2
100.0	50.74456144	51.62364995	51.65488505	51.65540378
	8.3411E-3	8.4093E-3	8.4115E-3	8.4118E-3
146.7	80.63581727	83.897875512	84.185211623	84.19999334
	5.9786E-3	6.1054E-3	6.1136E-3	6.1141E-3
215.4	126.5391395	133.71806103	135.21533500	135.3928391
	4.2691E-3	4.4102E-3	4.4312E-3	4.4334E-3
316.2	U	209.86438334	214.11621378	215.2075859
	U	3.1655E-3	3.2002E-3	3.2063E-3
464.1	U	334.99353930	334.94449783	338.3461971
	U	2.2818E-3	2.2999E-3	2.3113E-3
681.2	U	U	529.52947934	527.2444883
	U	U	1.6572E-3	1.6599E-3
1000.0	U	U	U	833.3477164
	U	U	U	1.1977E-3

Table 6.2: Convergence in the height of the envelope as a function of R_m and the truncation used, for the $m = 0$ E^A field symmetry. Shown are the values of the envelope height and the time it is attained underneath, accurate to 10^{-7}. Bold shows that convergence has been apparently achieved. The R_m are logarithmically equally spaced. U means that calculations did not converge.

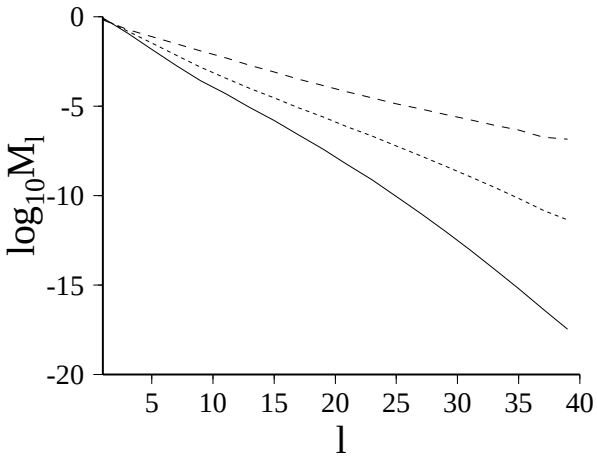


Figure 6.8: The power spectra of the initial fields with unit energy maximising the magnetic energy growth for the $\mathbf{s}_2$(MDJ) flow, $N_{max} = L_{max} = 40$. Solid shows $R_m = 46$, dotted $R_m = 100$ and dashed $R_m = 464$.

6.4.3 The nature of the transient growth

We now investigate the nature of the transient growth and the physical process responsible. For $R_m = 100$, the envelope height is 52 with an optimising time of $t = 0.0084$, corresponding to 1600 years in dimensional terms. We forward propagate the associated initial field through time. Figure 6.9 shows contours of $|\mathbf{B}|$ for a truncation level of $N_{max} = L_{max} = 24$. The field gets advected towards the z axis where it is vertically stretched. Note the scales of the plots.

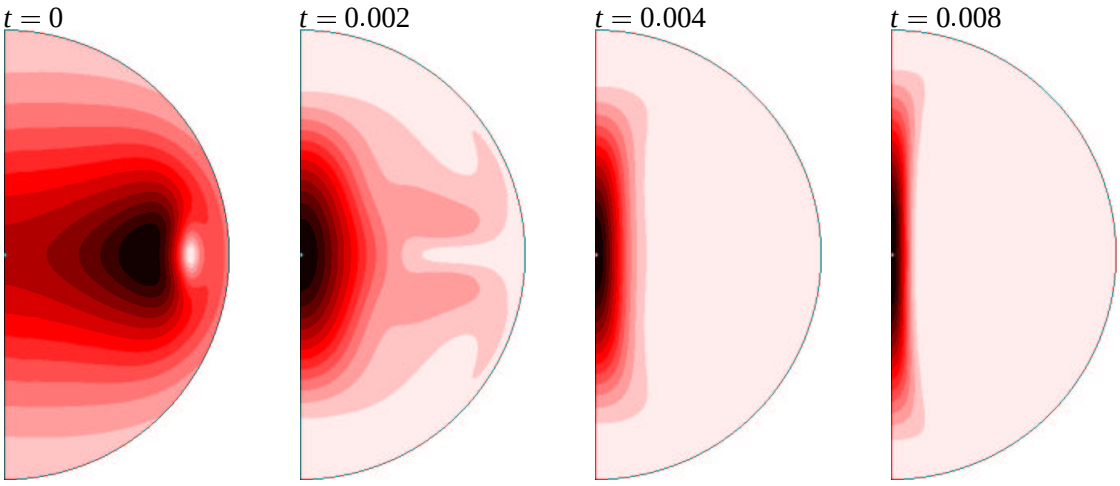


Figure 6.9: Contours of $|\mathbf{B}|$ for times: $t = 0$, 0.002, 0.004, 0.008, spanning the dimensional time period of 1500 years. Contour maxima are, respectively, 0.822, 5.643, 29.817, 114.509. The field is normalised to have unit initial energy.

Figure 6.10 shows the equivalent sequence of images for the $\mathbf{s}_2$ (IC) flow. The initial

field associated with an optimising time of $t = 0.0092$ for $R_m = 100$ is forward propagated. Again, the field is advected into and stretched near the origin, in a similar manner to the $\mathbf{s}_2$(MDJ) flow.

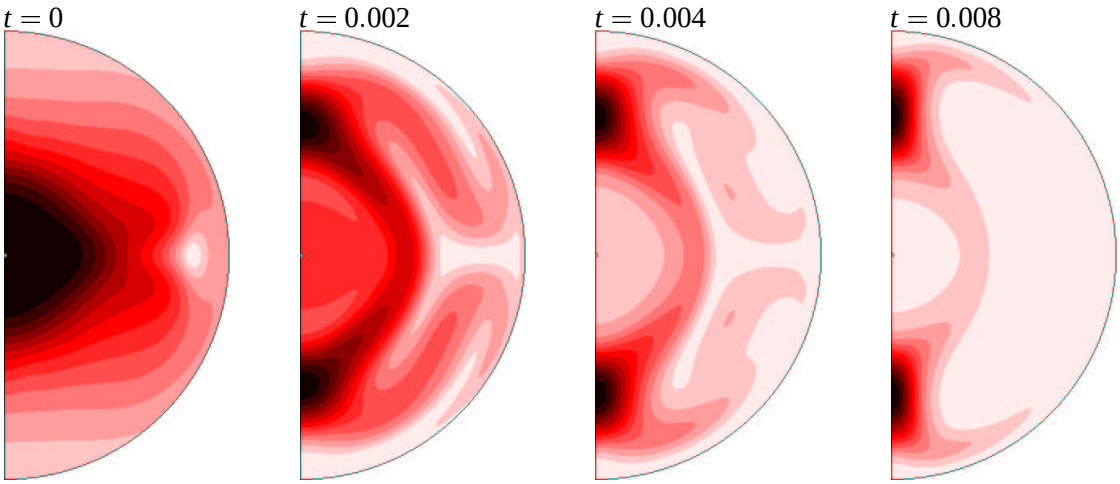


Figure 6.10: Contours of $|\mathbf{B}|$ for times: $t = 0$, 0.002, 0.004, 0.008, spanning the dimensional time period of 1500 years. Contour maxima are respectively:1.397, 2.137, 9.218, 25.346. The field is normalised to have unit initial energy.

Thus the physical mechanism responsible is the same despite the energy amplification difference. Magnetic field is advected towards the origin where it is stretched; this incorporates both finite time advection and the field line stretching of instantaneous magnetic energy growth.

Figure 6.11 shows the time evolved magnetic energy relating exactly to figure 6.9, achieving the maximal energy amplification of 52. The times of $t = 0.004$ and $t = 0.008$ are

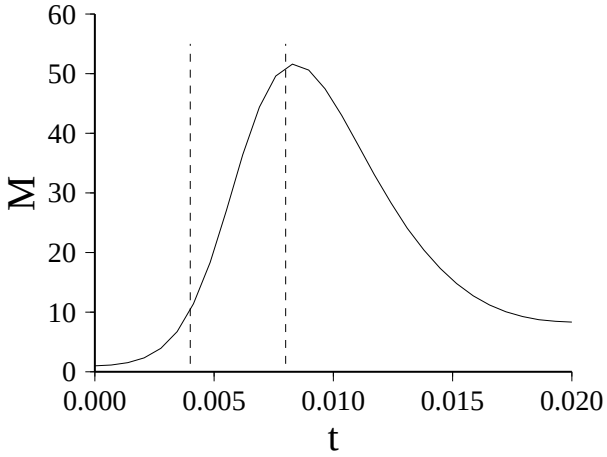


Figure 6.11: The evolution of the total magnetic energy as a function of time. Marked are lines representing $t = 0.004, 0.008$ for reference.

highlighted, since not only do they match with the contour plots of figure 6.9 but mark small and almost maximally large magnetic energy states. We investigate how the magnetic energy spectrum changes between its initial configuration and these times in figure 6.12(a). Figure 6.12(b) shows the evolution of the surface energy at the CMB. The initial field is shown as solid, that at $t = 0.004$ as short-dashed and that at $t = 0.008$ as long-dashed. The two sets of spectra

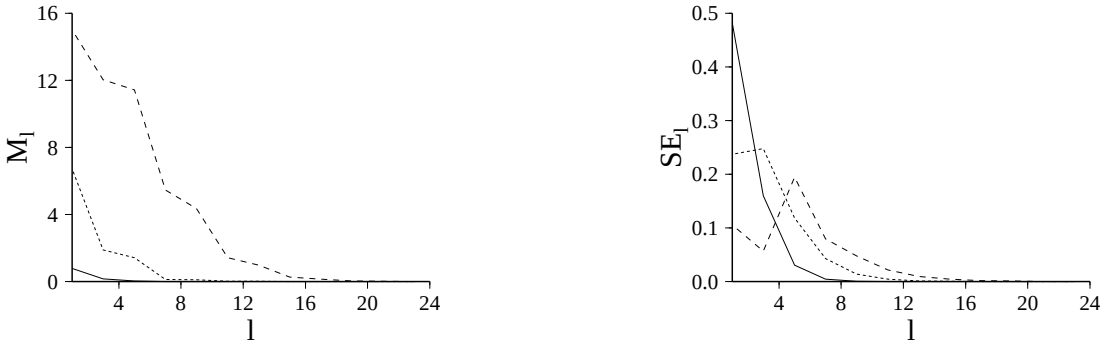


Figure 6.12: The spectrum of the magnetic field as a function of spherical harmonic degree l at $t = 0, 0.004, 0.008$.(a) shows the energy spectrum, M_l and (b) the surface energy at the CMB, SE_l (see equation 6.14). The initial condition is shown as the solid curve, $t = 0.004$ as short-dashed and $t = 0.008$ as long-dashed.

show very different characteristics. The large scales of the total magnetic energy appear to linearly increase, retaining their initial ratios to one another. Additionally, the $l = 1$ harmonics dominate at large times. In contrast, the surface magnetic energy shows a decrease as time evolves which a complex scale-dependent behaviour. Despite the $l = 1$ harmonics dominating initially, as time evolves it seems that the $l = 5$ harmonics become significant. An observer on the Earth's surface would therefore not see any transient magnetic energy growth, since not only does the surface energy actually decrease overall, but seems to be concentrated in $l = 5$, which undergoes moderate geometrical attenuation through the insulating mantle (see Appendix A).

Of course, we have chosen the initial condition to maximise the total magnetic energy growth and not that of the field as sampled in any other way. We may therefore expect that to achieve this favourable energy growth, the field becomes small near the CMB, concentrating its intensity where the flow is large and so being maximally amplified. The field as sampled on the CMB, and indeed that sampled on the Earth's surface, will not behave in the same way as that of the interior.

Figure 6.13 shows a comparison of the field intensity at a variety of locations on the Earth's surface, as a function of time. The field has been scaled to the Earth by its g_1^0 component, see section 6.1.4. Black shows that at Steens Mountain (of colatitude and longitude

$(0.265\pi, -0.659\pi)$), red, Greenwich $(0.214\pi, 0)$, green, the Equator at 0 longitude $(0.5\pi, 0)$ and blue, the South Pole $(\pi, 0)$. It is only after the dimensional optimising time of 1600 years

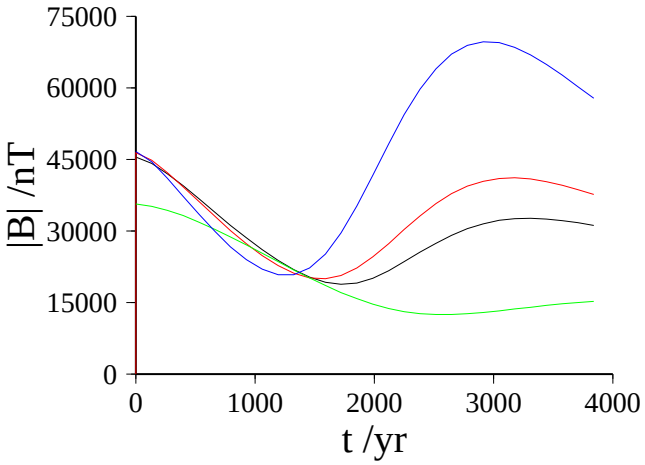


Figure 6.13: The time evolution of field intensity sampled at Steens Mountain (black), Greenwich (red), the Equator at 0 longitude (green) and at the South Pole (blue), for the $\mathbf{s}_2$(MDJ) flow with $R_m = 100$, truncation $N_{max} = L_{max} = 24$ with the $m = 0\, E^A$ field symmetry. The initial field is chosen to maximise the energy envelope (at $t = 0.0084$) and is scaled to the Earth through its g_1^0 component.

that the intensities show any sign of growth; indeed they decay initially whilst the field energy as a whole increases dramatically. After the growth of magnetic energy towards the centre of V has subsided, it diffuses outwards, leading to higher field energy as measured pointwise on the Earth's surface. We note however that this effect is location dependent: at the Equator the measured field intensity decays almost monotonically towards zero. In most locations, observers would have to wait 3000 years to see the maximum of the magnetic field intensity as sampled on the Earth's surface, double the time it takes the total magnetic energy to peak.

Although the above results pertain only to the $\mathbf{s}_2$ (MDJ) flow, the general behaviour as indicated is generic and does not change on altering the form of the $\mathbf{s}_2$ flow.

6.4.4 Scaling of envelope heights

We investigate how the energy envelope height scales with R_m for the three $\mathbf{s}_2$ flows studied, for the dominant $m = 0\, E^A$ field symmetry. Table 6.2 shows that fully converged solutions are not obtained above $R_m = 46$. However, we shall consider up to $R_m = 1000$ and hope that convergence errors are not large enough to impair the general behaviour, for we only need a few significant figures of accuracy for graphical purposes.

Figure 6.14(a) shows plots of the envelope height as a function of R_m for the three $\mathbf{s}_2$

flows; figure 6.14(b) shows the relevant optimising times. We use the highest truncation which is numerically tractable: $N_{max} = L_{max} = 40$. The curves in figure 6.14(a) do not tend to a per-

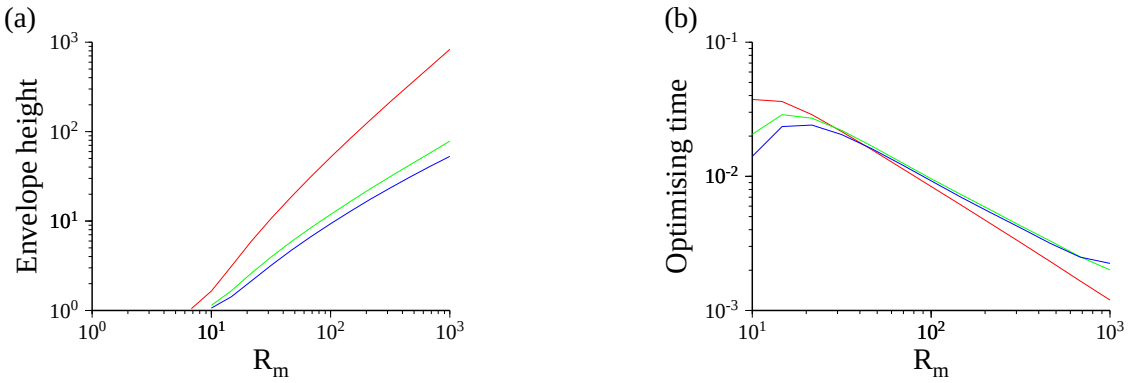


Figure 6.14: For the field of $m = 0$ E^A symmetry and truncation level $N_{max} = L_{max} = 40$, for varying R_m (a) shows the associated envelope height and (b) the optimising time. The three $\mathbf{s}_2$ flows are shown as red (MDJ), green (KR) and blue (IC).

fectly straight line in log-log space, indicating that either numerical inaccuracies at large R_m are important, or that the asymptotic regime (should it exist) has not yet been reached. However, we may fit straight lines through the points located in the range $10^2 \leq R_m \leq 10^3$ as approximations to the asymptotes. The curves in figure 6.14(b) apparently reach linear asymptotes in the range shown, until discrepancies appear at large R_m, due to numerical errors. Indeed, the fact that an asymptote has apparently been reached suggests that similar behaviour should occur in (a) but is affected significantly by numerical errors or truncation inaccuracies.

The approximated asymptotic scalings are shown below in table 6.3.

$\mathbf{s}_2$ flow	Envelope height	Optimising time
MDJ	$0.426R_m^{1.20}$	$0.426R_m^{-0.85}$
KR	$0.313R_m^{0.80}$	$0.249R_m^{-0.70}$
IC	$0.396R_m^{0.71}$	$0.208R_m^{-0.68}$

Table 6.3: Comparison of approximate asymptotic scalings for the envelope height and optimising time as functions of R_m. Computed for the $m = 0$ E^A field symmetry, truncation $N_{max} = L_{max} = 40$.

The $\mathbf{s}_2$ (KR) and $\mathbf{s}_2$ (IC) flows are very similar in their asymptotic scalings; indeed, this behaviour is also identified in figure 6.14(a). The envelope height for the $\mathbf{s}_2$ (MDJ) flow seems to obeys a much more favourable scaling, leading to much greater potential for transient behaviour. This indicates that the presence of a quiescent region in the vicinity of the origin is important, restricting the magnitude of the transient growth allowed, although the exact size does not appear make a great deal of difference.

Despite the differences in envelope height scaling, there are no grounds for supposing that the scalings of the optimising time with R_m are different, bearing in mind the likely errors associated with such approximate asymptotes.

Therefore generically, in the range $0 \leq R_m \leq 1000$, the envelope height scales as $O(R_m^{\gamma})$ where $0.7 \leq \gamma \leq 1.2$, depending on the flow chosen, and the optimising time as $O(R_m^{-3/4})$. We note that these scalings lead to the increase of transient growth with R_m, occurring on ever shorter time periods. This is intuitive: on driving the flow faster we expect that it becomes easier to generate magnetic fields and that any associated mechanism will operate more rapidly.

6.4.5 Initial fields

In each case, the optimal initial fields for magnetic energy growth over any time $t \geq 0$ are poloidal only in structure. Figures 6.15(a) and (b) show field lines for the initial $m = 0$ E^A fields chosen to optimise the energy envelope height for the $\mathbf{s}_2$(MDJ) flow with $R_m = 46$ and $R_m = 464$ respectively. They are extremely similar, although differ slightly in intensity near the origin (as indicated by the size of the arrows plotted). It is interesting to make a comparison with the field

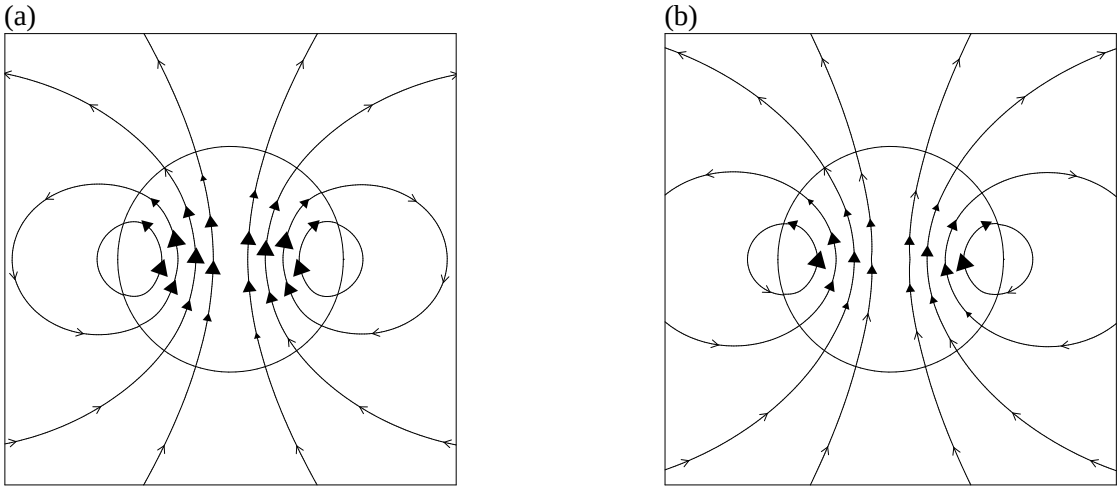


Figure 6.15: The initial field which maximises the envelope for (a) $R_m = 46$ and (b) $R_m = 464$. Computed for the $\mathbf{s}_2$(MDJ) flow and $m = 0$ E^A field symmetry.

which optimises the initial energy growth (as addressed in section 5.7). Field lines in that case are shown in figure 5.10. Although the field looks structurally very similar to that of figure 6.15, if we forward propagate it through time it undergoes vastly suboptimal energy growth as shown in figure 6.16. Energy is amplified by only a factor of 4 rather than the maximum of 52 (for this R_m), despite the initial field being qualitatively similar. We must therefore be wary that maximal growth is not as robust a feature as we might like, and that physically similar fields may not be

amplified in energy equally.

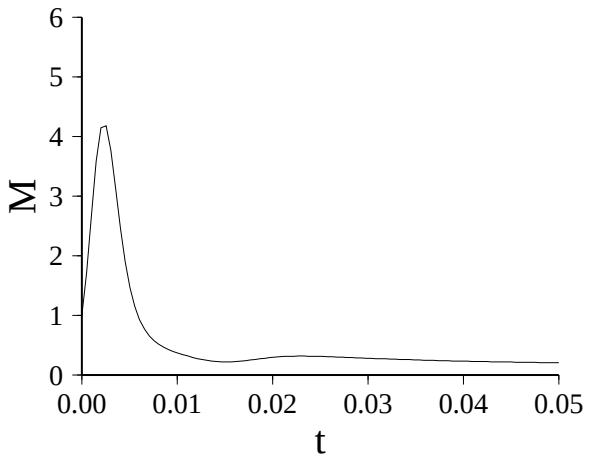


Figure 6.16: The time evolution of the most energetically unstable field, for the $m = 0\ E^A$ field symmetry, $R_m = 100$, $N_{max} = L_{max} = 40$.

6.5 The transient magnetic energy amplification of $\mathbf{t}_1\,\mathbf{s}_2$ flows

6.5.1 Symmetry separation

Figures 6.17(a)–(d) show the magnetic energy envelopes for the various independent symmetries at $R_m = 100$ for the flows: (a) $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 0.5$, (b) $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 1$, (c) $\mathbf{t}_1\,\mathbf{s}_2$(KR) $\tau = 0.5$ and (d) $\mathbf{t}_1\,\mathbf{s}_2$(KR) $\tau = 1$. The truncation used is $N_{max} = L_{max} = 24$. Blue is $m = 0$, red is $m = 1$, solid is E^A dashed is E^S. The $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) $\tau = 0.5$ flow has an unstable eigenvalue ($R_m^c = 55.0$) of field symmetry $m = 1\ E^S$ that dominates the envelopes at large times. Notice though that it is highly suboptimal for $t < 0.04$ (which in dimensional time is 7,700 years). It is unlikely that the flow in the Earth's outer core will be sufficiently stationary to sustain this growing eigenmode over such a time period, so the fact that such an exponentially growing field exists in these calculations is of no real importance.

Again we see that a robust feature of these flows is the dominance of the geophysically relevant $m = 0\ E^A$ field symmetry; this is perhaps not expected since such a symmetry dominates flows consisting of the $\mathbf{t}_1$ and $\mathbf{s}_2$ components taken individually.

Both the $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) flows out-perform the $\mathbf{t}_1\,\mathbf{s}_2$(KR) flows by a factor of four (note the vertical scale). This stems from the increased efficiency of the $\mathbf{s}_2$(MDJ) component relative to the $\mathbf{s}_2$(KR).

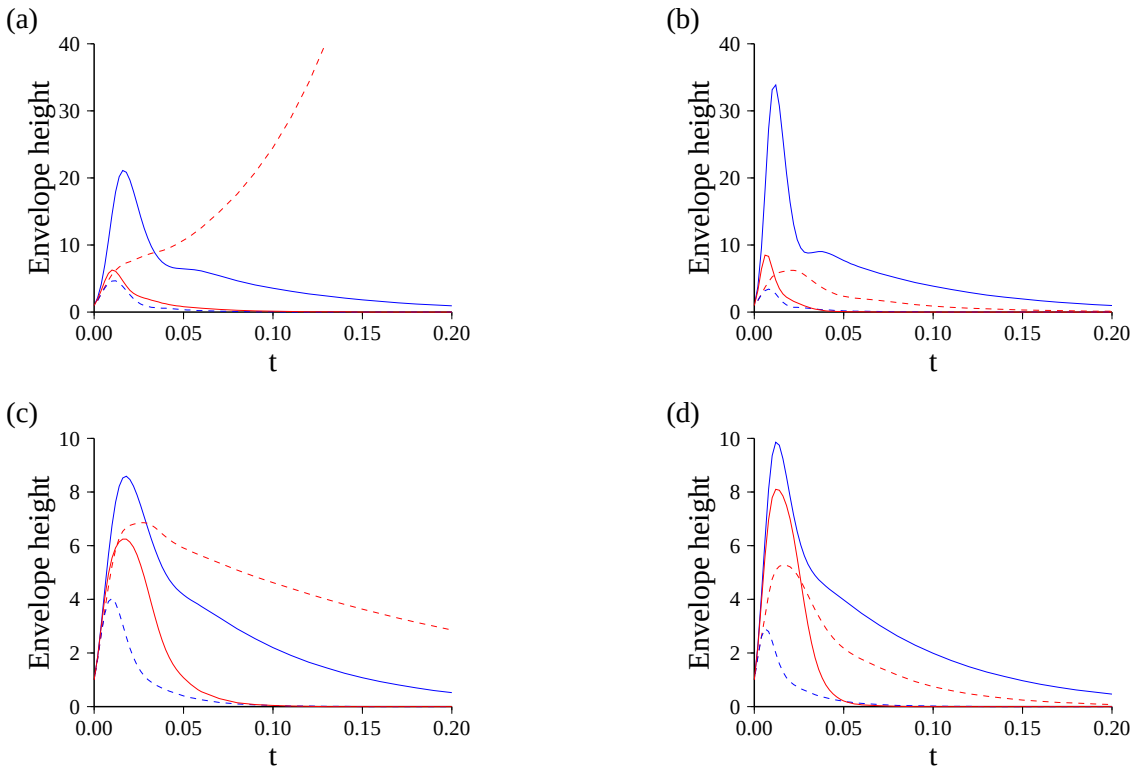


Figure 6.17: The symmetries compared in computations of the magnetic energy envelopes for the (a) $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 0.5$, (b) $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 1$, (c) $\mathbf{t}_1\,\mathbf{s}_2$(KR) $\tau = 0.5$ and (d) $\mathbf{t}_1\,\mathbf{s}_2$(KR),$\tau = 1$. $R_m = 100$; the truncation level is $N_{max} = L_{max} = 24$. Blue is the $m = 0$ symmetry, red is the $m = 1$ symmetry, solid is E^A, dashed is E^S. In (a), a growing eigenmode solution exists of $m = 1\,E^S$ symmetry; in (c), a slowly decaying eigenmode exists of eigenvalue $\lambda = -2.412 \pm 55.42i$ of the same symmetry. The $m = 0\,E^S$ symmetry dominates all non-normal growth in each case.

6.6 Envelope height scalings

Figures 6.18(a) and (b) show the variation of the envelope height for the three different flows with the associated optimising time. We study only the dominant $m = 0\,E^A$ symmetry with a truncation of $N_{max} = L_{max} = 40$. Red shows $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 0.5$, green MDJ $\tau = 1$, blue KR $\tau = 0.5$ and black KR $\tau = 1$ The approximated asymptotic scalings are shown below in table 6.4. All the flows have similar properties, but the two sets of $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) and $\mathbf{t}_1\,\mathbf{s}_2$ (KR) curves look and behave almost alike. Indeed, the analysis does not critically depend on the exact choice of τ, but shows generic behaviour. Of the $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) flows, that with $\tau = 1$ (the equipartition flow) supports greater growth. This is in conflict with the $m = 1\,E^S$ symmetry, where a growing eigenmode exists for $\tau = 0.5$ but not for $\tau = 1$.

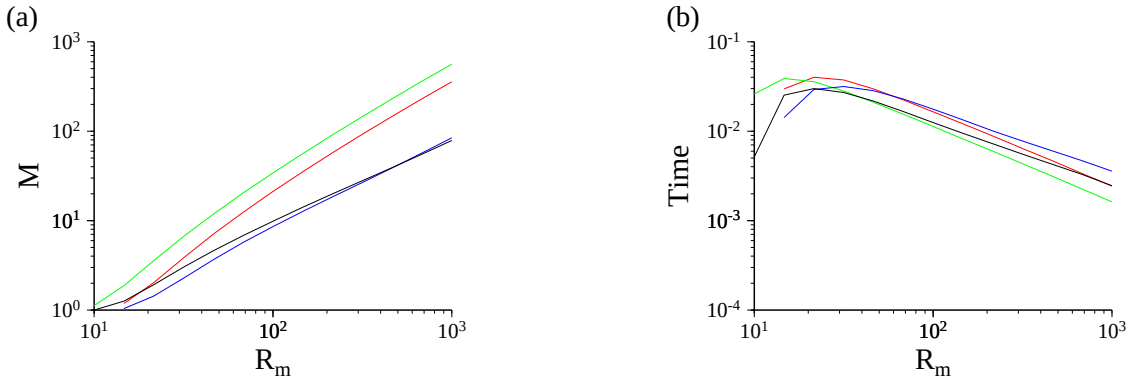


Figure 6.18: Comparison of (a) envelope height and (b) optimising time for $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 0.5$ shown as red, $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 1$ as green, $\mathbf{t}_1\,\mathbf{s}_2$(KR) $\tau = 0.5$ as blue, $\mathbf{t}_1\,\mathbf{s}_2$(KR) $\tau = 1$ as black.

$\mathbf{t}_1\,\mathbf{s}_2$ flow	Envelope height	Optimising time
MDJ $\tau = 0.5$	$0.090R_m^{1.20}$	$0.79R_m^{-0.84}$
MDJ $\tau = 1$	$0.150R_m^{1.20}$	$0.59R_m^{-0.85}$
KR $\tau = 0.5$	$0.097R_m^{0.98}$	$0.44R_m^{-0.70}$
KR $\tau = 1$	$0.166R_m^{0.89}$	$0.28R_m^{-0.68}$

Table 6.4: Comparison of approximate asymptotic scalings for the envelope height and optimising time as functions of R_m. For the $m = 0$ E^A field symmetry, truncation $N_{max} = L_{max} = 40$.

6.7 The magnetic energy amplification of the KR and STW flows

6.7.1 Symmetry separation

Figures 6.19(a) and (b) show the magnetic energy envelopes for (a) the KR and (b) the STW flows, both at $R_m = 100$. We use a truncation level of $N_{max} = L_{max} = 16$ which is just enough to attain sufficient convergence. Since the flows are not axisymmetric, many more harmonics are needed in the symmetry classes, and in this case they involve large matrices of dimension 2304×2304. This is the maximum resolution tractable, since not only are matrix exponentials computationally expensive, but numerical inaccuracy issues and spurious singular values arise from large matrix calculations.

The symmetries shown are: blue, containing even values of m, red, containing odd values of m; solid shows E^A and dashed E^S symmetry.

It is interesting to compare these curves with the equivalent for the $\mathbf{t}_1$ flow in section 6.2 which comprises 99% and 96% of the KR and STW respectively by rms: they look almost identical. This measure is therefore robust in the sense that small flow differences do not change the resulting stability, at least at $R_m = 100$.

We could not compute any asymptotic scalings since calculations with $R_m > 100$ were

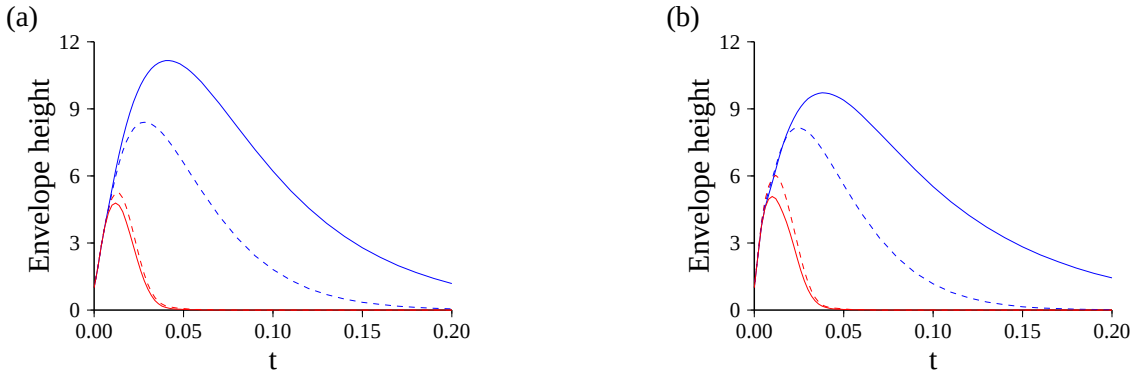


Figure 6.19: Energy envelopes for (a) the KR flow and (b) the STW flow, $R_m = 100$. Blue contains even m field harmonics, red contains odd m harmonics, solid is E^A and dashed is E^S.

	$R_m = 100$		Asymptotic scaling	
Flow	Envelope height	Optimising time	Envelope height	Optimising time
$\mathbf{t}_1$	10.93	0.040	R_m^2	0.0407
KR	11.16	0.042	-	-
STW	9.72	0.038	-	-
$\mathbf{s}_2$(MDJ)	51.66	0.0084	$R_m^{1.2}$	$R_m^{-0.85}$
$\mathbf{s}_2$(KR)	11.88	0.0096	$R_m^{0.80}$	$R_m^{-0.70}$
$\mathbf{s}_2$(IC)	9.36	0.0092	$R_m^{0.71}$	$R_m^{-0.68}$
$\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 0.5$	21.16	0.016	$R_m^{1.20}$	$R_m^{-0.84}$
$\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 1$	34.19	0.011	$R_m^{1.20}$	$R_m^{-0.85}$
$\mathbf{t}_1\,\mathbf{s}_2$(KR) $\tau = 0.5$	8.59	0.018	$R_m^{0.98}$	$R_m^{-0.70}$
$\mathbf{t}_1\,\mathbf{s}_2$(KR) $\tau = 1$	9.88	0.166	$R_m^{0.89}$	$R_m^{-0.68}$

Table 6.5: A comparison of the envelope heights attained by the various flows studied with $R_m = 100$ with the approximate asymptotic scaling. In two cases, convergence issues did not allow the scaling to be calculated, shown by $-$.

unconverged.

6.8 Flow comparison

For each flow, the dominant symmetry for the non-normal growth is $m = 0$ E^A, matching that of the Earth. Table 6.5 shows the differing amount of transient growth possible for all the flows studied for $R_m = 100$, along with their approximate asymptotic scalings.

We have written the flows in groups, all showing similar characteristics. The $\mathbf{t}_1$, KR and STW flows, all having very similar structures amplify an initial magnetic field by a factor of around 10 in energy, for $R_m = 100$. The time at which these maxima are attained are approximately also equal, indicating that the same mechanism of poloidal field line shearing is causing the transient magnetic energy growth. At a higher value of $R_m = 1000$, the $\mathbf{t}_1$ flow may optimally

amplify magnetic energy by $O(1000)$ with the generated field being purely toroidal.

The $\mathbf{s}_2$ flows vary considerably in amplification capability, with the $\mathbf{s}_2$(MDJ) being the most effective. The optimising time though is very similar for $R_m = 100$, and indeed this also shows in the asymptotic scaling. Again the same mechanism is responsible for each, that of advection towards the z-axis followed by stretching. At $R_m = 1000$, the $\mathbf{s}_2$(MDJ) flow may amplify magnetic energy by $O(1000)$, generating purely poloidal field.

The $\mathbf{t}_1\,\mathbf{s}_2$ flows again vary in envelope height, with those containing the $\mathbf{s}_2$(MDJ) component performing the best. Although those containing the $\mathbf{s}_2$(KR) component in fact show a lesser energy amplification than that of $\mathbf{s}_2$(KR) itself, the asymptotic scaling suggests that for larger R_m this efficiency difference will reverse.

Overall there are two main mechanisms for generating transient magnetic fields: that associated with $\mathbf{t}_1$ of field line shearing and that associated with $\mathbf{s}_2$, of field line stretching. All the flows considered are either combinations of these or close to one (in rms). Small flow perturbations may alter the effectiveness of the energy amplification, but the mechanism remains the same. In addition, as R_m increases, the ability of all the flows to grow magnetic fields concomitantly rises. Therefore, not only is this analysis robust, but it agrees with our physical understanding of generation mechanisms.

6.8.1 Geophysical discussion

Transient growth analysis is geophysically robust, in that each flow is dominated by the Earth-like $m = 0$ E^A field symmetry. Certain flows have the ability to support growing eigenmode solutions; however, these might take up to 10,000 years to dominate (depending on the initial field), a time period over which the flow will undoubtedly change since the timescale of core convection turnover time is around 500 years. Given the sensitivity of such solutions to the underlying flow, it is unlikely that they are manifested in the Earth.

Transient growth may explain the recovery of the field intensity and its axially symmetric dipolar component after a magnetic reversal, especially in the case where the flow is dominated by an $\mathbf{s}_2$ convective overturn motion. At $R_m = 1000$, the $\mathbf{s}_2$(IC) flow can amplify magnetic energy by 53 times, corresponding to an intensity increase of 7. The timescale on which this occurs is 500 years or so, a little short but nonetheless consistent with the figure of 10,000 years from paleomagnetic data. We also note that this recovery is unlikely to be manifested immediately on the Earth's surface, and that calculations indicate a significant time delay of several thousand years. This brings the swift timescale computed to be more in line with that geophysically observed.

6.9 Optimal excitation of the most unstable eigenmode

The transient growth discussed above deals with time periods short compared to those on which we would expect the fastest growing (or slowest decaying) eigenmode to dominate. That is, we are more interested in how the superposition of these modes creates magnetic energy rather than studying any mode individually. At large times however, assuming the flow remains stationary, the most unstable eigenmode will dominate the system and is the subject of most of the previous literature on kinematic dynamos. In order to maximise the field energy at large times, we want all of the energy to end up in this eigenmode; however, a slight twist comes about from the non-normality of the system: the optimal initial condition in this case is not that of purely the eigenvector, but its associated adjoint vector. In most cases having an initial condition proportional to the dominant eigenvector is extremely suboptimal. We first introduce this rather counter-intuitive result with a simple example.

6.9.1 A simple example

Let us consider the example given in section 6.3.1 in matrix form. We represent **B** by the vector (S,T) and write the schematic induction equation with the $\mathbf{t}_1$ flow as:

$$\frac{d}{dt}\begin{pmatrix} S \\ T \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ R_m & -2 \end{pmatrix}\begin{pmatrix} S \\ T \end{pmatrix} \tag{6.15}$$

where we represent the poloidal and toroidal decay rates by -1 and -2 respectively. If we now seek solutions of the form $\mathbf{B} \propto e^{\lambda t}$ then we may write

$$\begin{pmatrix} S \\ T \end{pmatrix}(t) = A\begin{pmatrix} 1 \\ R_m \end{pmatrix}e^{-t} + B\begin{pmatrix} 0 \\ 1 \end{pmatrix}e^{-2t} \tag{6.16}$$

If we use the usual Euclidean 2-norm, the two eigenvectors are not orthogonal. Notice that the non-normality of the system increases with R_m: when $R_m = 0$ the eigenvectors are orthogonal. To maximise the energy at large times, we would like the initial condition to have a maximal projection onto the slowest decaying eigenvector, that is, such that A is maximal. Restricting ourselves to initial fields of unit energy, the choice $B = 0$ forces $A = 1/\sqrt{(1+R_m^2)}$ which is vastly suboptimal as we shall see.

The adjoint problem is described in this norm by the transpose of the defining time-

propagation matrix and has eigenvalues -2 and -1 with associated eigenvectors respectively

$$\begin{pmatrix} R_m \\ -1 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{6.17}$$

As with all adjoint vectors, they form a biorthogonal set, that is, each is orthogonal to all the original eigenvectors except in the case where the eigenvalues are complex conjugates of each other, or in this case equal (since they are real). If we want an initial condition to give a maximal value of A, then we consider writing (at $t = 0$), since both sets of vectors are complete:

$$A \begin{pmatrix} 1 \\ R_m \end{pmatrix} + B \begin{pmatrix} 0 \\ 1 \end{pmatrix} = C \begin{pmatrix} 1 \\ 0 \end{pmatrix} + D \begin{pmatrix} R_m \\ -1 \end{pmatrix} \tag{6.18}$$

for some constants C and D. We want to find the field with the largest projection onto A. It is clear that $D = 0$ since the adjoint vector it multiplies is orthogonal to that multiplying A. Therefore, the optimal initial condition is $C = 1$, equating to

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ R_m \end{pmatrix} - R_m \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{6.19}$$

which has unit norm, giving the value $A = 1$.

Thus, two eigenmodes both of norm greater than one (if $R_m > 1$) can superpose to create an initial field of unit norm but with a larger projection onto the most unstable eigenvector than just that of the vector itself. As $R_m \to \infty$, the 'efficiency' of an initial field of the adjoint mode relative to that of just the leading eigenmode itself, is measured by the ratio of A: $1/\sqrt{(1+R_m^2)}$, the ratio of the coefficient multiplying the leading eigenmode at large times. This decreases as $O(R_m^{-1})$ for large R_m.

Physically, in order to generate large magnetic energy growth, we need to start off with a purely poloidal field (the associated adjoint vector) which gets sheared by the flow, creating large toroidal energy growth. The most unstable eigenvector is mostly toroidal (at least for $R_m \gg 1$) and so an initial configuration in this form cannot be sheared nearly as much; such an initial configuration is therefore highly suboptimal.

6.9.2 A theoretical derivation

Suppose a vector system is governed by the equation

$$\frac{d\mathbf{v}}{dt} = \mathsf{L}\,\mathbf{v} \tag{6.20}$$

where in our case $\mathbf{v}$ represents $\mathbf{B}$ with respect to the Chebyshev basis and L is the discretised induction operator. As discussed in section 5.1.1 we may work in the Euclidean 2-norm. It follows simply from its definition that the adjoint in this norm is $\mathsf{L}^\dagger = \mathsf{L}^T$.

If we diagonalise $\mathsf{L} = \mathsf{H}\Lambda\mathsf{H}^{-1}$ then

$$\mathsf{L}^\dagger = \left[\mathsf{H}^{-1}\right]^T \Lambda \mathsf{H}^T \tag{6.21}$$

where H contains the eigenvectors of L as columns and Λ is a diagonal matrix of the eigenvalues.

Since the exponential series expansion can be written

$$e^{\mathsf{H}\Lambda\mathsf{H}^{-1}t} = \sum_{i=0} \left(\mathsf{H}\Lambda\mathsf{H}^{-1}t\right)^i / i! = \mathsf{H}\left[\sum_{i=0} \left(\Lambda t\right)^i / i!\right]\mathsf{H}^{-1} = \mathsf{H}e^{\Lambda t}\mathsf{H}^{-1} \tag{6.22}$$

the solution to the time-dependent problem is

$$\mathbf{v} = e^{\mathsf{L}t}\mathbf{v}(0) = \mathsf{H}e^{\Lambda t}\mathsf{H}^{-1}\mathbf{v}(0) \tag{6.23}$$

The exponential of a diagonal matrix is also diagonal with entries $\left[e^{\Lambda t}\right]_{ii} = e^{[\Lambda]_{ii}t}$. Suppose that Λ_{11} is the unique eigenvalue with maximal real part, then the long term behaviour is governed only by the sub-matrix of $e^{\mathsf{L}t}$:

$$\mathsf{H}(:,1)\,e^{\Lambda_{11}t}\left[\mathsf{H}^{-1}\right](1,:) \tag{6.24}$$

where the notation $\mathsf{H}(:,1)$ means the matrix of identical size to H but containing only the the first column of H; $\mathsf{H}(1,:)$ denotes a similar matrix but only containing the first row. This matrix is already in singular value decomposition (up to a normalisation) and so the optimal initial condition maximising the energy at $t \gg 1$ is given by the right singular vector (section 4.8.1). On inspecting equation (6.21) this turns out to be nothing more than the adjoint eigenvector associated with the eigenvalue Λ_{11}. Denoting this adjoint vector by $\mathbf{a}$ and the associated eigenvector of L by $\mathbf{v}$ (both column vectors), we can write for $t \gg 1$, $e^{\mathsf{L}t} = \mathbf{v}e^{\Lambda_{11}t}\mathbf{a}^T$ and the ratio of the

energies of the two time-propagated initial vectors of **a** and **v** is

$$\frac{|e^{\mathsf{L}t}\mathbf{v}|^2}{|e^{\mathsf{L}t}\mathbf{a}|^2} = \frac{|\mathbf{v}e^{\Lambda_{11}t}\mathbf{a}^T\mathbf{v}|}{|\mathbf{v}e^{\Lambda_{11}t}\mathbf{a}^T\mathbf{a}|} = \frac{\mathbf{a}\cdot\mathbf{v}}{1} \tag{6.25}$$

where both **v** and **a** are normalised to have unit Euclidean 2-norm. Reverting back to our energy inner product, the efficiency of having an initial condition as the most unstable eigenvector relative to that of the associated adjoint vector is $(\mathbf{a},\mathbf{v})$ which typically scales as $R_m^{-\gamma}$ for some $\gamma > 0$.

6.9.3 Scalings of efficiency

We firstly analyse the $\mathbf{t}_1$ flow with a field of $m = 0$ E^A symmetry, which we simplistically represented as an introductory example in section 6.9.1. Figure 6.20(a) shows the (energy) inner product of **v** and **a**. The scaling R_m^{-1} is shown as the dashed line, in agreement with that of the simple model. The adjoint mode is purely poloidal as already discussed, which is maximally sheared by the flow before its eventual decay. Figure 6.20(b) shows an equivalent plot but for the $\mathbf{s}_2$(MDJ) flow with a field of identical symmetry; the dashed line is the scaling $R_m^{-0.37}$.

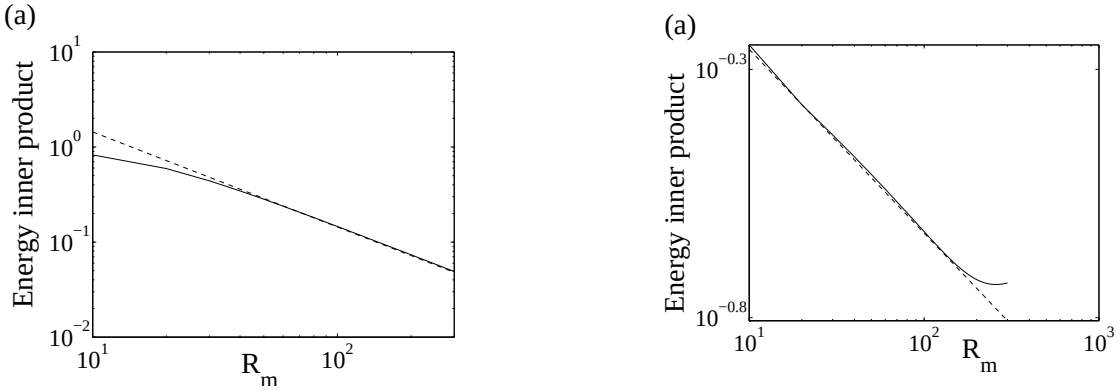


Figure 6.20: Energy inner product between the most unstable eigenvector and its associated adjoint vector, as a function of R_m. Dashed shows the linear asymptote (in log-log space); the truncation level is $N_{max} = L_{max} = 24$. (a) shows the $\mathbf{t}_1$ flow, having an associated scaling of R_m^{-1}, (b) shows the $\mathbf{s}_2$ (MDJ), having a scaling $R_m^{-0.37}$.

Let us now consider the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) $\tau = 0.5$ flow which supports an exponentially growing eigenmode with $R_m^c = 55.0$. Figure 6.21(a) shows the energy inner product for the relevant $m = 1$ E^S field symmetry; dashed is the fit $R_m^{-0.4}$ (the same index as that found for the $m = 0$ E^A symmetry for this flow). The behaviour is unaffected going through the linear stability transition at $R_m = 55.0$. The eigenvectors are complex in this case, but we use only the physically relevant real part.

In figure 6.21(b) we show the time propagated initial fields of the most unstable eigenvector (solid) and that of the associated adjoint mode (dashed), with $R_m = 60 > R_m^c$ for the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow, $\tau = 0.5$. At the end of the time series plotted, the latter curve takes on an exponential form but with a much greater amplitude.

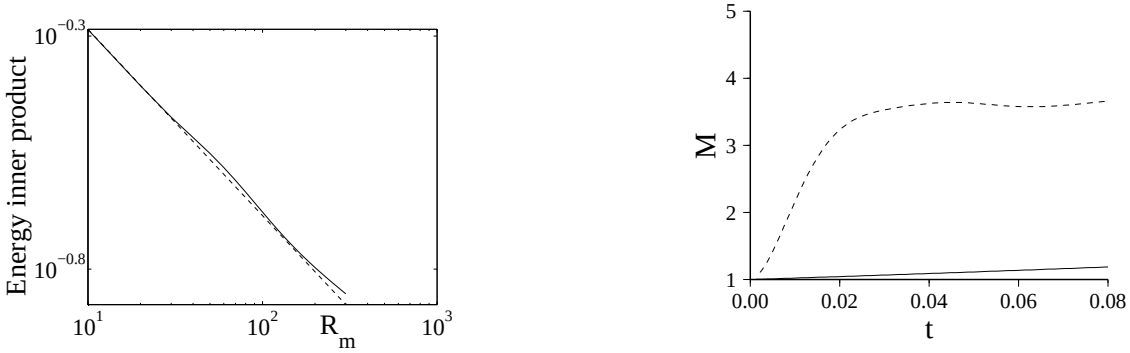


Figure 6.21: For the $\mathbf{t}_1\,\mathbf{s}_2$ (MDJ) flow with $\tau = 0.5$, $N_{max} = L_{max} = 24$, $m = 1$ E^S field symmetry (a) shows the energy inner product between the most unstable eigenvector and its associated adjoint mode. Dashed shows the fit $R_m^{-0.4}$. (b) The time evolution of the magnetic energy of the most unstable eigenvector (solid) and its associated adjoint mode (dashed), both normalised to have unit initial energy.

6.9.4 Summary of the adjoint–eigenvector analysis

In each flow considered above due to the non-normality of the system, it is always far more efficient to excite the dominant eigenmode by taking an initial field of its associated adjoint vector, rather than the eigenvector itself. Most flows seem to have an 'efficiency' scaling of R_m^γ where $\gamma \approx -0.4$, apart from the $\mathbf{t}_1$ flow which scaled as R_m^{-1}, the most severe. In their study of non-normal magnetic energy growth in a cylindrical geometry, *Farrell and Ioannou* (1999a) found a similar scaling of $R_m^{-1/2}$, despite the different geometry.

6.10 Minimum time calculations

If we seek a final steady state in fully non-linear calculations, the time period over which such computer models are run must be sufficiently long so that all transient processes have decayed. The question is, exactly how long is such a period?

When solving the joint system of the induction equation with the Navier Stokes equations, any ephemeral behaviour would in general couple both the flow and magnetic field in a non-linear fashion. We set a lower bound on the model time required by considering only the linear process of transient magnetic field growth in stationary flows. We study the same flows as

before but now solve for the maximum time needed for the energy to fall back down to its initial value of 1. This corresponds to intersections of the energy envelope with the curve $M = 1$ for non-zero time, which we locate using a bisection method. These values will in general depend on R_m, so we fix it initially at $R_m = 500$ being a geophysically relevant value. We choose the $m = 0$ E^A field symmetry since this dominates in each flow. Table 6.6 shows the strictly positive non-dimensional time for which the energy envelope intersects the curve $M = 1$ (second column). We scale this to the number of dipole diffusion times (third column) and the dimensional time in years (fourth column). No converged solutions we found for the $\mathbf{s}_2$(MDJ) flow with $R_m = 500$

Flow	t	$\pi^2 t$	Time/(10^3 yrs)
$\mathbf{t}_1$	0.357	3.52	69
$\mathbf{s}_2$ (KR)	0.185	1.83	36
$\mathbf{t}_1\,\mathbf{s}_2$(MDJ), $\tau = 0.5$	0.282	2.78	54
$\mathbf{t}_1\,\mathbf{s}_2$(MDJ), $\tau = 1$	0.266	2.63	51
$\mathbf{t}_1\,\mathbf{s}_2$(KR) , $\tau = 0.5$	0.260	2.57	50

Table 6.6: Intersection of the energy envelope curves with $M = 1$ for $t > 0$, for various flows. Computed with a truncation level of $N_{max} = L_{max} = 24$ for the $m = 0$ E^A field symmetry, $R_m = 500$. The third column shows the number of dipole diffusion times, the fourth the associated dimensional time in years.

nor for the KR and STW flows with $R_m > 100$.

At a higher value of $R_m = 1000$, more transient growth is possible and the associated values of t become larger. Table 6.7 shows results from a few flows for comparison. For

Flow	t	$\pi^2 t$	Time/(10^3 yrs)
$\mathbf{t}_1$	0.4274	4.32	82
$\mathbf{t}_1\,\mathbf{s}_2$(MDJ), $\tau = 0.5$	0.3249	3.21	62
$\mathbf{t}_1\,\mathbf{s}_2$(KR), $\tau = 0.5$	0.3145	3.10	60

Table 6.7: Intersection of the energy envelope curves with $M = 1$ for $t > 0$ for various flows. Computed with a truncation of $N_{max} = L_{max} = 24$ for the $m = 0$ E^A field symmetry, $R_m = 10^3$. The third column shows the number of dipole diffusion times, the fourth the associated dimensional time in years.

geophysical values of $500 \leq R_m \leq 1000$ the minimum computational time required for a run to discount all transient effects is 4.32 dipole diffusion times (82,000 years). Such bounds depend on the particular flow studied: this figure is associated with the $\mathbf{t}_1$ flow. It is of note that this exceeds the bound of *Backus* (1957) of 4 dipole diffusion times corresponding to axisymmetric poloidal-only fields, avoiding the conditions of the theorem by generating large toroidal field energy. Generically we require at least three dipole diffusion times for a computational run, in line with most current simulations. Of course fully non-linear transient behaviour may last longer

than these kinematic results, and consequently they only place a lower bound on the necessary model time for the full problem.

6.11 Virtual geomagnetic pole paths for transient events

So called VGP diagrams are a good way of visualising the behaviour of the magnetic field as sampled at a particular site. This is especially useful for paleomagnetic data, where good spatial coverage is often lacking and global field models cannot be constructed. A VGP is the location of the geomagnetic north pole of a purely dipole field, consistent with measurements taken at the site. If the data allows, these VGP locations trace out a curve as a function of time, the so called VGP path.

If the field was purely dipolar, then such diagrams would exactly capture the path of the geomagnetic pole. However, in general the field morphology is much more complex and VGP paths are site dependent. Of particular interest are VGP paths recording the nature of geomagnetic reversals, especially those showing extremely rapid field changes, for example, that seen in the data from Steens Mountain. In general the paths wander in both longitude and latitude as time evolves.

We now consider how the transient magnetic field growth as computed in the preceding sections would look as VGP paths, computed from observations taken on the Earth's surface. As has already been shown in figure 6.13, the behaviour of the magnetic field intensity on the Earth's surface may not indicate in any way the generation processes going on deep within the core. We investigate whether the same is true here.

To construct VGP paths at a given site, we require measurements of the field inclination I and declination D. The methodology used is detailed in Appendix D.

6.11.1 Computed VGP paths

In each of the axisymmetric flows studied, the dominant transient magnetic energy growth symmetry is $m = 0$ E^A. This means that the VGP paths lie on a line of constant longitude identical to that of the sampling location. Geophysically this does not give much insight since observed VGP paths in general vary in longitude with time. Instead we form plots associated with the KR and STW flows, which are non-axisymmetric and have a dominant field symmetry that is E^A and contains harmonics for which m is even.

We propagate through time initial fields which maximise the magnetic energy growth for $R_m = 100$, for which calculations are sufficiently converged using a manageable truncation

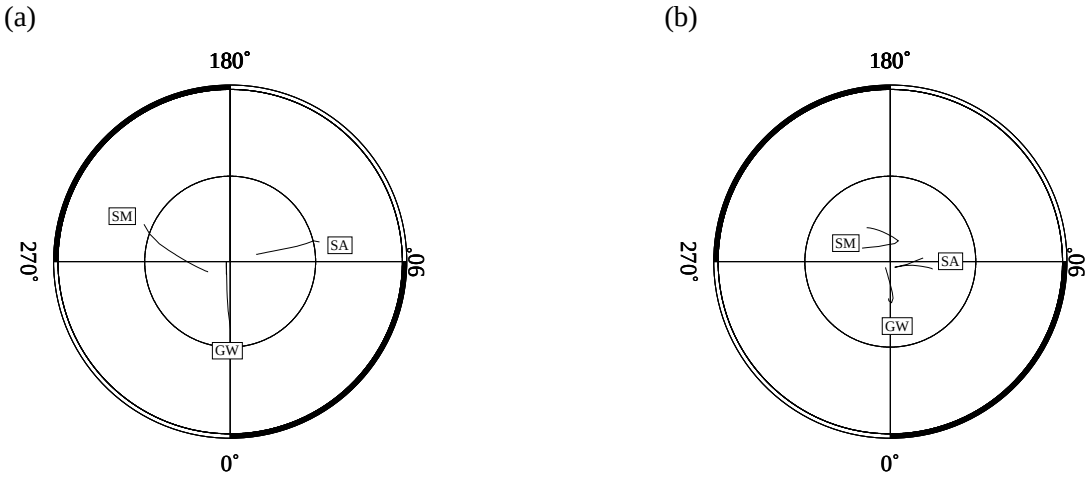


Figure 6.22: VGP paths for the (a) KR and (b) STW flows, viewed in a polar projection from above the north pole. The initial field maximising the magnetic energy growth for $R_m = 100$ is forward propagated through time. The latitudinal grid is 10°.

level of $N_{max} = L_{max} = 16$. For a sequence of times $0 \leq t \leq 0.2$ in this interval, we compute the VGP path for various chosen locations on the Earth's surface.

Figures 6.22(a) and (b) show VGP paths for three locations: SM labels Steens Mountain $(\theta = 0.265\pi, \phi = -0.659\pi)$, SA: offshore south Africa $(\theta = 0.722\pi, \phi = 0.278\pi)$ and GW: Greenwich $(\theta = 0.214\pi, \phi = 0)$. The VGP paths for both the KR and STW flows are very close to the geographic north pole and do not extend far in latitude (the scale is 10°). In KR, each path starts furthest away and moves toward the north pole. In STW, at SA and SM the VGP paths start heading toward the geographic north pole and then turn away, in contrast to that at GW, which heads straight towards it.

Crucially, over a non-dimensional time scale of 0.2 (38,000 years), the VGPs really do not extend very far in latitude or longitude, so that in particular, the huge variation in the VGP at Steens Mountain over a course of a week or so cannot be explained by such a model. We note however that for these flows, the largest component of energy growth is toroidal which is not seen at the Earth's surface. The VGPs point predominantly northwards, indicating that the $\mathbf{S}_1$ component dominates all other poloidal components. However, as time proceeds (for example seen at SM in figure 6.22(b)), the axisymmetric dipole eventually decays and the VGP moves slightly away from the geographic north pole.

6.12 The Lorentz force

In this chapter we have so far addressed the issue of transient magnetic energy growth. The kinematic regime on which this work is based is only justified when the Lorentz force is small, so that the induction equation may be studied in isolation. Indeed, our aim is to study the energy growth of small magnetic fields until this assumption breaks down. We do not seek exponential growth in particular, but growth on any timescale such that this occurs. Although we have already seen substantial potential for transient energy growth, no comments have yet been made on the likely effect of the associated Lorentz force on the underlying flow pattern. It is plausible, for example, that the Lorentz force, proportional to $(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$, is precisely zero because the generated current lies parallel to the associated field lines. Such a circumstance does not arise in general, but it is nonetheless important to quantify its strength relative to the forces which might be responsible for driving the flow.

The Earth is believed to be in a so called *magnetostrophic* balance (e.g. *Gubbins and Roberts*, 1987), where to leading order, the Coriolis force ($2\rho\boldsymbol{\Omega} \times \mathbf{u}$, where Ω is the angular rotation vector of the Earth and ρ is the density), the Lorentz force ($\mu_0^{-1}(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$), pressure gradients and buoyancy are in equilibrium. In our kinematic framework, we assume that initially the magnetic field is sufficiently small that the Lorentz force can be neglected. In this situation then, the Coriolis force, buoyancy and pressure are in a *geostrophic* equilibrium, of which we assume $\mathbf{u}$ to be a stationary solution. Of course, such flows are unknown to us, so we generally prescribe them to describe some physical mechanism, such as convection. Should the magnetic field and therefore the Lorentz force grow to a magnitude comparable with any of the geostrophic forces, the kinematic framework then breaks down. We may therefore compute the (dimensional) ratio of the Coriolis force to the Lorentz force, as a measure of the veracity of the kinematic assumption.

We scale the magnetic field to the Earth by its largest external dipole component (in most cases $g_1^0 = -29616.13$ nT). Explicitly, the rms dimensional Lorentz force is $\frac{\mathscr{B}^2}{\mathscr{L}\mu_0}(\boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B})_{rms}$ where $\mathscr{B}$ is the magnetic field scaling factor and $\mathscr{L}$ is the dimensional length scale, taken to be the radius of the CMB. The dimensional Coriolis force is $2\rho\mathscr{U}\Omega(\hat{\mathbf{z}} \times \mathbf{u})_{rms}$ where the typical velocity scale is $\mathscr{U} = \frac{R_m}{\mathscr{L}\mu_0\sigma}$. The other parameters are: the angular rotation rate of the Earth $\Omega = 7.3 \times 10^{-5}\mathrm{s}^{-1}$, the density of the outer core $\rho = 10^4 \mathrm{kg\,m}^{-3}$ (*Dziewonski and Anderson*, 1981), the conductivity of the outer core $\sigma = 5 \times 10^5 \mathrm{S\,m}^{-1}$ and permeability of free space $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{H\,m}^{-1}$.

We compute the non-dimensional rms values using identical techniques to those pre-

sented in chapter four and noting that $\hat{\mathbf{z}} = \mathbf{S}_1$, with $S_1(r) = r^2/2$. Additionally, we may compute the Ohmic dissipation $\int_V (\mu_0^{-1}\boldsymbol{\nabla} \times \mathbf{B})^2/\sigma \, dV$ associated with the transient energy growth. This represents a drain on the magnetic field: electric currents induce Ohmic heating and consequently magnetic energy is lost. The main contribution to this dissipation is small scale fields, thus the main field would be spatially 'smoothed' over time if no generation mechanism was present. If the Ohmic heating is too large, too much energy is lost and the dynamo will cease to operate. Geophysical values of 1–2 TW for the Ohmic dissipation were proposed by *Roberts et al.* (2003) as an upper bound, while (*Buffett*, 2002a) suggested 0.1–0.5 TW, although larger values were not discounted. We can use these to restrict our choice and flow and the size of R_m, for we do not expect the numerically evaluated dissipation to vastly exceed these geophysical values.

Another quantity of relevance to the geodynamo is the work done by the Lorentz force on the flow, non-dimensionally given by $R_m \mathbf{u} \cdot (\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B})$.

In section 2.3.3, we saw that if non-slip boundary conditions were obeyed by an incompressible flow, this work is equal and opposite to that done by the flow on the magnetic field. Indeed, we showed that

$$R_m \int_V \mathbf{B} \cdot \boldsymbol{\nabla} \times (\mathbf{u} \times \mathbf{B}) \, dV = -R_m \int_V \mathbf{u} \cdot (\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B} \, dV = R_m \int_V \mathbf{B} \cdot \mathsf{e}\mathbf{B} \, dV \tag{6.26}$$

Thus a spatial average of these rates of work is equal in magnitude to the spatial average of the amount of field line stretching. In a pointwise sense however, these work rates give important clues as to where and how the flow pattern might be significantly affected by the Lorentz force.

Since most of the existing literature on kinematic dynamos is concerned with eigenmode instability, we compare the Lorentz force and other relevant quantities associated with such growth, with those from transience calculations. In this way, we can determine whether there is much difference, if any, in the physical effects of these two solution types.

In what follows we write $\mathbf{J} = \boldsymbol{\nabla} \times \mathbf{B}$ for the non-dimensional electric current density, differing by a factor of μ_0^{-1} from the dimensional quantity, and $(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$, the Lorentz force, differing by an identical factor from its dimensional value. Unless specified to the contrary, all quantities will be non-dimensional. Although we present results for a few flows individually, the results relating to all the various flows are compared in section 6.12.4.

6.12.1 The $\mathbf{t}_1$ flow

We analyse the $\mathbf{t}_1$ flow with a truncation level of $N_{max} = L_{max} = 24$ with the dominant $m = 0$ E^A field symmetry. When $R_m = 100$, the optimising time for the maximum energy growth is $t = 0.0397$ and has an associated energy amplification of 10.93, corresponding to a rise in $\mathbf{B}_{rms}$ of $\sqrt{10.93} = 3.31$.

Figure 6.23 shows the evolution of $\mathbf{B}_{rms}$ (long dashed), $(\boldsymbol{\nabla} \times \mathbf{B})_{rms}$ (short-dashed) and $((\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B})_{rms}$ (solid), where rms denotes the root mean squared average over the unit sphere; the initial field is normalised to have unit energy. The rms value of the field increases by

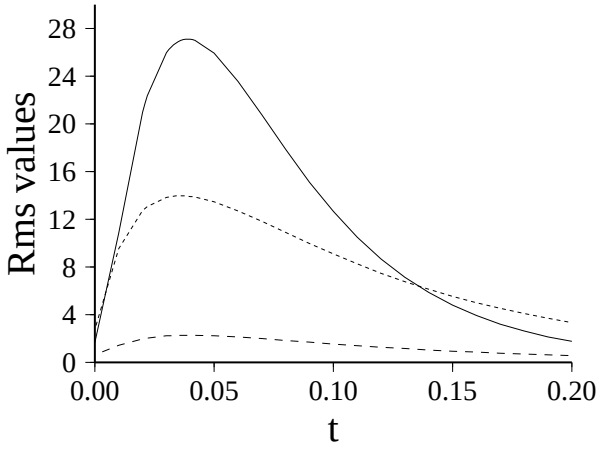


Figure 6.23: The evolution of the rms values of $\mathbf{B}$ (long dashed), $\boldsymbol{\nabla} \times \mathbf{B}$ (short-dashed) and $(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$ (solid). The rms field increases by a factor of 3.31 (from an initial value of 0.62), the current by 4.5 (initially 2.78) and the Lorentz force by 16 (initially 1.66).

a factor of 3.31, while the Lorentz force increases by a factor of 16. This much larger factor is explained by the rise in current of around 4.5 times. In this case, the rms value of the Lorentz force is precisely the product of the rms values of its constituent components: the current and the magnetic field, indicating that they are orthogonal and so we can write

$$\int_V |(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}|^2 dV = \int_V |(\boldsymbol{\nabla} \times \mathbf{B})|^2 |\mathbf{B}|^2 dV \tag{6.27}$$

which suggests, if the components behave suitably

$$|(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}|_{rms} \approx |\boldsymbol{\nabla} \times \mathbf{B}|_{rms} |\mathbf{B}|_{rms} \tag{6.28}$$

This is simply explained since the transiently grown field is predominately $\mathbf{T}_2$ (section 6.2.4), having only a $\hat{\boldsymbol{\phi}}$ component. Its associated current $\mathbf{J}$ comprises mainly $\mathbf{S}_2$, which has no $\hat{\boldsymbol{\phi}}$

component. Thus in this case the current and the magnetic field are pointwise orthogonal. The magnetic field grows only in the main body of the flow, whereas the current is roughly constant except near $r = 1$ where it is heavily influenced by the boundary conditions which bring it sharply to zero. This suggests that equation 6.28 is sufficiently satisfied in order to explain our results.

Figure 6.23 shows that the current scales linearly with the magnetic field, and indicates that the extra $\boldsymbol{\nabla}\times$ operation on $\mathbf{B}$, forming $\mathbf{J}$, gives no extra time dependent behaviour, i.e. the length scales relevant are the same at any given instant. This is in agreement with the results of section 6.2.2 where we found that the field energy spectra did not change through time, being dominated by the $\mathbf{T}_2$ component. It follows that the peaks of the Lorentz force, magnetic field and the current occur almost simultaneously as the figure shows, and that since the energy envelope height scales as R_m^2 for this flow, the rms Lorentz force scales similarly as R_m^2.

From a geophysical point of view, it is the ratio of the Coriolis force and the dimensional Lorentz force which is of importance. In particular, if this ratio is $O(1)$ then the field has a significant effect on the flow and the stationary-flow kinematic problem is no longer valid. As before, to calculate the importance of this we rescale the g_1^0 component of the field to its Earth-like value; details can be found in section 6.1.4. The ratio for the case $R_m = 100$, takes a maximum value of 0.088. In this regime then, the Coriolis force is dominant, even though the Lorentz force has increased by a factor of 16 from its original value. Since the rms Coriolis force scales linearly in R_m, it follows that the ratio also scales as R_m. Higher flow rates then mean that the kinematic assumption becomes increasingly less valid in the event of transient growth.

Figure 6.24 shows (a) contours of $|\mathbf{B}|$, (b) contours of $|\boldsymbol{\nabla}\times\mathbf{B}\times\mathbf{B}|$ and (c) contours of $R_m\mathbf{u}\cdot\boldsymbol{\nabla}\times\mathbf{B}\times\mathbf{B}$, the work done by the Lorentz force on the flow. The snapshot is taken at $t = 0.04$ corresponding to the magnetic energy peak; red is positive, blue negative. The predominantly negative contour values in figure 6.24(c) show that the Lorentz force acts mainly against the flow, as we might expect from Lenz's law (stating that any growing field will act against the movement generating it), and that it is concentrated at locations of large magnetic energy growth. The total work done is $\int_V R_m\mathbf{u}\cdot\boldsymbol{\nabla}\times\mathbf{B}\times\mathbf{B}\,dV = -403.9754$, the negative value indicating that overall the Lorentz force does work against the flow at this time. In the fully dynamical system, the flow velocity near the regions of strong field line shearing would be slowed, and consequently the the scope for field growth would be reduced. Notice though that there are regions for which the Lorentz force acts upon the flow favourably (near $r = 1$ in particular).

We can also compute the dimensional Ohmic dissipation: $\int_V \frac{\mathbf{J}^2}{\sigma}\,dV$. For a value of $R_m = 1000$, its maximum is $339.29GW$, whereas for $R_m = 100$, it is about 10^2 times lower: $3.44GW$. These are well within the geophysical bounds of *Roberts et al.* (2003) of 1–2 TW.

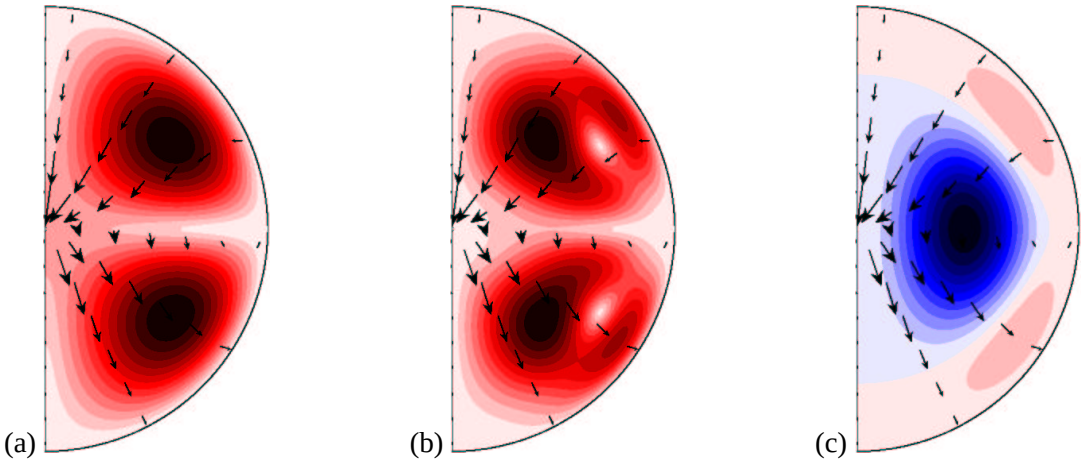


Figure 6.24: At $t = 0.04$, contours of (a) $0.1130 \leq |\mathbf{B}| \leq 4.2285$, (b) $0 \leq |\boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B}| \leq 5.1292$, (c) $-134.00 \leq R_m \mathbf{u} \cdot \boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B} \leq 24.76$. Arrows show the direction of the field. Computed for the $\mathbf{t}_1$ flow, $m = 0\, E^A$ field symmetry, $R_m = 100$. The initial field is normalised to have unit energy. The snapshot is taken at $t = 0.04$.

6.12.2 The $\mathbf{s}_2$(MDJ) flow

We now investigate the Lorentz force associated with the transient magnetic energy growth under the action of the $\mathbf{s}_2$(MDJ) flow. Again we use $R_m = 100$, and we take as an initial field that which maximises the magnetic energy growth, associated with an optimising time of $t = 0.0084$ and energy amplification of 52. We use the dominant $m = 0\, E^A$ field symmetry and a truncation level of $N_{max} = L_{max} = 24$.

Figure 6.25(a) shows the evolution of $\mathbf{B}_{rms}$ (dotted, right axis) and $\mathbf{J}_{rms} = (\boldsymbol{\nabla} \times \mathbf{B})_{rms}$ (solid, left axis). Note the disparity of the magnitudes: although $\mathbf{B}_{rms}$ has increased by a factor of about 8, the current shows an increase (in rms) of 29. This means that the characteristic length scale must decrease by a factor of 3 during the time evolution shown. That this is so is not surprising considering how the spatial scales evolve in time as the field is stretched by the flow (section 6.4.3).

Figure 6.25(b) shows the evolution of $((\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B})_{rms}$ (solid, left axis) and $\int_V R_m \mathbf{u} \cdot \boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B}\, dV$ (dotted, right axis). The Lorentz force grows by a factor of 2275 and on average, opposes the motion of the fluid (since the integral above is negative). It reaches its rms maximum at roughly the same time as those of the magnetic field and current. That the Lorentz force increases by 2275, the magnetic field strength by 8 and the current by 29 is not in error! Although an *a priori* scaling of Lorentz force is linear with the magnetic field and current, these can locally superpose to create very strong Lorentz forces, making the global average large while the global average of its constituent components might still be relatively weak. Using the

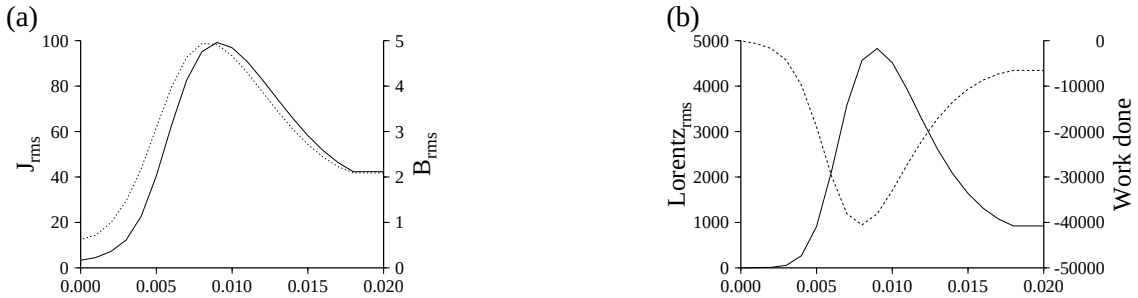


Figure 6.25: The evolution of various quantities associated with the initial field chosen to maximise the energy growth at $t = 0.0084$ for $R_m = 100$. The rms values plotted are those of (a) $\boldsymbol{\nabla} \times \mathbf{B}$ (solid, left hand axis) and $\mathbf{B}$ (dotted, right hand axis). (b) $(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$ (solid, left hand axis) and $\int_V R_m \mathbf{u} \cdot \boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B} \, dV$ (dotted, right hand axis).

Cauchy-Schwarz inequality:

$$\int_V |\mathbf{J} \times \mathbf{B}|^2 \, dV \leq \int_V |\mathbf{J}|^2 \, |\mathbf{B}|^2 \, dV \leq \left\{ \int_V |\mathbf{J}|^4 \, dV \right\}^{\frac{1}{2}} \left\{ \int_V |\mathbf{B}|^4 \, dV \right\}^{\frac{1}{2}} \tag{6.29}$$

and so it does not follow that the rms Lorentz force is bounded by the product of the rms values of the magnetic field and current, unless either the current or magnetic field intensity is constant in locations when the other varies, enabling us to replace the first inequality by two separate integrals.

Scaling the relevant quantities to their geophysical values, the maximum rms ratio of the Lorentz force to the Coriolis force is 21.33. Hence in this regime, the flow is very strongly influenced by the field. The time at which the Lorentz force is maximised is around 0.009. Figure 6.26 shows (a) contours of $|\boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B}|$ and (b) contours of $R_m \mathbf{u} \cdot \boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B}$, where we see the Lorentz force strongly suppressing the flow near the z-axis. To see the effect of changing R_m on the Lorentz force, we compare the two values $R_m = 100$ (a lower estimate for the Earth) and $R_m = 1000$ (an upper estimate). In each case we choose the initial condition which maximises the energy envelope, which for $R_m = 1000$ occurs at an optimising time of 0.0012. Table 6.8 shows the change in behaviour for two different values of R_m. Increasing R_m by an order of

R_m	$\mathbf{B}_{rms}$	$(\boldsymbol{\nabla} \times \mathbf{B})_{rms}$	$((\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B})_{rms}$	$R_m \mathbf{u} \cdot (\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$
100	4.94 (0.008)	99.25 (0.009)	4827.62(0.009)	-40500(0.008)
1000	19.95 (0.004)	1198.26 (0.004)	638428.58 (0.004)	-6002500 (0.004)

Table 6.8: The maximum value of different quantities as they evolve in time; the time at which the maxima occurred are shown in parentheses. The initial field is normalised to have unit energy.

magnitude increases $\mathbf{B}_{rms}$ by a factor of five, and decreases the length scale on which $\mathbf{B}$ varies

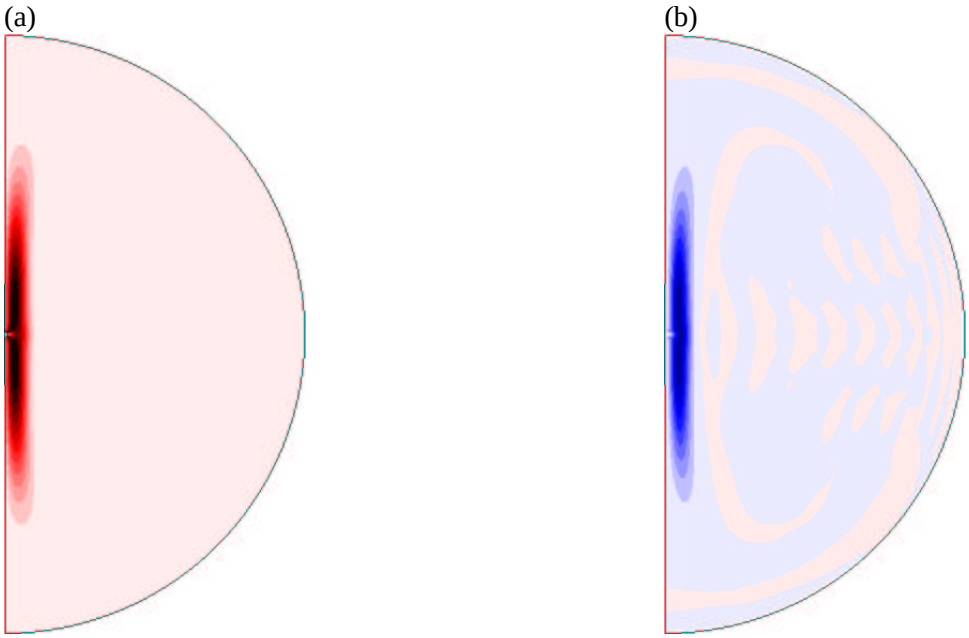


Figure 6.26: At $t = 0.009$, contours of (a) $0 \leq |\nabla \times \mathbf{B} \times \mathbf{B}| \leq 2526.154$,
(b) $-47807.40 \leq R_m \mathbf{u} \cdot \nabla \times \mathbf{B} \times \mathbf{B} \leq 5449.552$, for the initial field with unit energy maximising the magnetic energy at $t = 0.0084$, for the $\mathbf{s}_2$(MDJ) flow with $m = 0$ E^A symmetry, $R_m = 100$.

from $1/20$ to $1/60$ of the core radius. The Lorentz force for $R_m = 1000$ is huge, corresponding to 334 times the Coriolis force. The Ohmic dissipation associated with this is 27.18 TW, more than geophysical constraints allow. Its large magnitude is due primarily to the small length scales involved (compare with the value of 154 GW for $R_m = 100$). These two geophysical constraints restrict the choice of R_m, should this process occur to around $R_m \leq 100$. However, other $\mathbf{s}_2$ flows are not so good at generating transient magnetic energy growth and we might expect that the associated Ohmic dissipation and Lorentz force are concomitantly lower, since the field will not be stretched on such a tiny length scale. Therefore, such a stringent bound on R_m might not apply. Indeed, even the geophysical bound on dissipation rates may be avoided if the growth is only transient, for example, recovering the field after a magnetic reversal. Even though for a short while the Ohmic heating might be high, its contribution to the total energy budget may be small.

6.12.3 $\mathbf{t}_1\mathbf{s}_2$ flows

Similar calculations have been carried out on $\mathbf{t}_1\mathbf{s}_2$ flows and the results are shown in the flow comparison of section 6.12.4. However, of particular interest is the nature of the Lorentz force associated with eigenmode growth: how does it compare to the transience calculations?

We investigate the $\mathbf{t}_1\mathbf{s}_2$ (MDJ) flow with $\tau = 0.5$ which is eigenvalue unstable with

field symmetry $m = 1\, E^S$ and $R_m^c = 55.0$. The eigenvector is unique up to a solid body rotation, and we fix the orientation by demanding that the largest scale Chebyshev basis component of $\mathbf{S}_1^{1c}(1)$ vanishes. We rescale the h_1^1 component associated with $S_1^{1s}(1)$ to have the geophysical value of 5185.72 nT.

Again we use $R_m = 100$ for comparison, and choose the initial condition to optimise the field energy at $t = 0.2$, in order that it has an opportunity to dominate. This has the effect of putting all the energy into the relevant adjoint vector; the field evolves quickly into the eigenvector although at a higher energy level than if this was chosen as the initial state. The most unstable eigenvalue is $8.2961 \pm 56.6541\, i$ and we would expect the growth to take on the form of an exponentially growing solution at large times. Unless the Lorentz force and current identically vanish, they must also grow exponentially in time; however, we must determine their initial amplitude.

Figure 6.27(a) shows the evolution of $\mathbf{B}_{rms}$ (dotted, right axis) and $(\boldsymbol{\nabla} \times \mathbf{B})_{rms}$ (solid, left axis). After an initial transient, both the field and current grow exponentially. Note though that the scale of the current is 10 times that of the field, indicating that the length scale of the current is a tenth of that of the magnetic field. Figure 6.27(b) shows the associated rms Lorentz force (solid, left axis) and its rate of working on the flow (dotted, right axis).

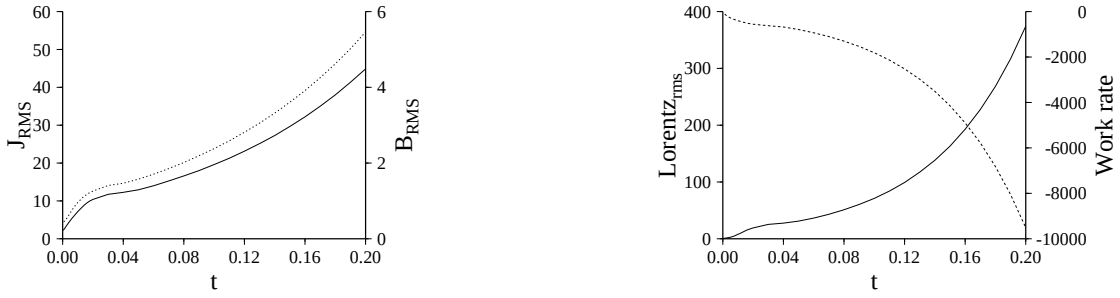


Figure 6.27: For the $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 0.5$ flow, $R_m = 100$, the evolution of the most unstable eigenmode solution in rms values of (a) $\boldsymbol{\nabla} \times \mathbf{B}$ (solid, left hand axis) and $\mathbf{B}$ (dotted, right hand axis); (b) $(\boldsymbol{\nabla} \times \mathbf{B}) \times \mathbf{B}$ (solid, left axis) and $R_m \int_V \mathbf{u} \cdot \boldsymbol{\nabla} \times \mathbf{B} \times \mathbf{B}\, dV$ (dotted, right axis).

The Lorentz force, on average, opposes the motion of the fluid (since the work rate above is negative). The ratio of the maximum rms dimensional Lorentz force with the rms Coriolis force in this case is 0.060, despite the field having grown to a much larger energy. The Ohmic dissipation rate is 0.036 GW, tiny compared to other previously calculated values. The work done by the Lorentz force on the flow is -8.08 GW, small compared to transient growth calculations. This can be explained by the the fact that the eigenmode assumption relies on having the same growth rate everywhere. Therefore, in regions of strong stretching, the field must

be weaker than in those regions of lesser dynamo capability. Thus the work done in stretching field lines is less than in the optimal transient cases.

We must also take into account that the field has been normalised using a different (and smaller) dipole component than g_1^0, used for other flows. Although this gives an Earth-like comparison, it means that quantities rescaled by it may be a little low. We deliberately chose to normalise the field such that the largest scale contribution to S_1^{1c} vanished on $r = 1$, so that $S_1^{1s}(1)$ would be larger than $S_1^{1c}(1)$, enabling us to match to the largest equatorial external dipole coefficient. By such a procedure, we scaled the field to be as large as geophysically possible, to try to compensate for the discrepancy of a factor of 6 between g_1^0 and h_1^1. Despite our efforts, we could plausibly include an additional factor of 36 in to the Ohmic heating and Lorentz force, increasing the values in line with the other results. The Ratio of Lorentz to Coriolis forces is also rendered approximately unity by this (maximal) scaling. Nonetheless, if we restrict ourselves to the geophysical values, then all the computed dimensional quantities are very small in comparison to those of transience analysis.

6.12.4 Comparison of flows

We now show a summary of the various quantities computed for the suite of flows studied. Table 6.9 shows details of the initial fields used and the maximum fractional increase in the rms values of field, current and Lorentz force (these may not take place at precisely the same time). Table 6.10 shows the maximal rms values of the quantities of table 6.9, along with the non-dimensional work done by the Lorentz force on the flow.

The procedure was as follows. For $R_m = 100$, the height of the envelope curve for the dominant field symmetry was calculated. For the associated optimising time, we used the pertaining initial field and forward propagated it through time, at intervals of 0.001 non-dimensional time units. For each time, we computed the various quantities listed in the table. The symbol Δ indicates the maximal gain in the associated rms quantity relative to its initial value. Note that $\Delta \mathbf{B}_{rms}$ is not exactly the same as the square root of the energy envelope height, since this rms is taken only over V and the energy extends over $V + \hat{V}$.

The symbol † indicates a truncation level of $N_{max} = L_{max} = 16$ rather than $N_{max} = L_{max} = 24$ was used for computational reasons. Values of $R_m > 100$ are not amenable to this comparative study since the associated matrices required to obtain convergence for the KR and STW flows were too large, introducing instead numerical errors and spurious results. The symbol * indicates that the solution was an eigenmode; see section 6.12.3 for details.

Column four of table 6.10 shows the rms size of the Lorentz force associated with

Flow	Symmetry	Opt time	$\Delta\mathbf{B}_{rms}$	$\Delta(\nabla\times\mathbf{B})_{rms}$	$\Delta((\nabla\times\mathbf{B})\times\mathbf{B})_{rms}$
$\mathbf{t}_1$	$m=0\ E^A$	0.0397	3.65	5.03	16.29
KR†	$m=0\ E^A$	0.042	3.71	5.25	17.60
STW†	$m=0\ E^A$	0.038	3.18	6.47	24.45
$\mathbf{s}_2$(MDJ)	$m=0\ E^A$	0.0084	7.94	29.37	2275.29
$\mathbf{s}_2$(KR)	$m=0\ E^A$	0.0096	3.81	10.21	198.80
$\mathbf{s}_2$(IC)	$m=0\ E^A$	0.0093	3.37	10.17	203.25
$\mathbf{t}_1\mathbf{s}_2$(MDJ)($\tau=0.5$)	$m=0\ E^A$	0.0165	5.15	15.04	504.08
$\mathbf{t}_1\mathbf{s}_2$(MDJ)($\tau=0.5$)*	$m=1\ E^S$	0.2	12.52	20.49	526.73
$\mathbf{t}_1\mathbf{s}_2$(MDJ)($\tau=1$)	$m=0\ E^A$	0.0127	6.58	22.74	1399.02
$\mathbf{t}_1\mathbf{s}_2$(KR)($\tau=0.5$)	$m=0\ E^A$	0.0176	3.28	7.50	68.65
$\mathbf{t}_1\mathbf{s}_2$(KR)($\tau=1$)	$m=0\ E^A$	0.0125	3.50	8.92	128.48

Table 6.9: Initial field details for the calculations, with maximum relative growth in rms of the magnetic field, current and Lorentz force. * indicates that the solution was an eigenmode, † indicates a truncation level of $N_{max}=L_{max}=16$ rather than $N_{max}=L_{max}=24$. Computed for $R_m=100$.

Flow	Maximal non-dimensional values of			
	$\mathbf{B}_{rms}$	$(\nabla\times\mathbf{B})_{rms}$	$((\nabla\times\mathbf{B}\times\mathbf{B}))_{rms}$	$R_m(\mathbf{u}\cdot\nabla\times\mathbf{B}\times\mathbf{B})_{rms}$
$\mathbf{t}_1$	2.28 (0.040)	13.97 (0.036)	27.10 (0.039)	-982 (0.021)
KR†	2.31 (0.042)	14.23 (0.038)	27.84 (0.041)	-996 (0.022)
STW†	2.02 (0.045)	15.00 (0.042)	26.52 (0.044)	-964 (0.029)
$\mathbf{s}_2$(MDJ)	4.94 (0.008)	99.25 (0.009)	4827.62(0.009)	-40542(0.008)
$\mathbf{s}_2$(KR)	2.38 (0.010)	31.59 (0.012)	402.18 (0.014)	-4038 (0.008)
$\mathbf{s}_2$(IC)	2.11 (0.009)	28.88 (0.012)	367.26 (0.013)	-3416 (0.008)
$\mathbf{t}_1\mathbf{s}_2$(MDJ)($\tau=0.5$)	3.17 (0.017)	41.45 (0.018)	773.22 (0.018)	-7010 (0.016)
$\mathbf{t}_1\mathbf{s}_2$(MDJ)($\tau=0.5$)*	7.72 (0.2)	63.47 (0.2)	748.40 (0.2)	-19014 (0.2)
$\mathbf{t}_1\mathbf{s}_2$(MDJ)($\tau=1$)	4.02 (0.012)	67.11 (0.013)	2178.60(0.13)	-18504 (0.011)
$\mathbf{t}_1\mathbf{s}_2$(KR)($\tau=0.5$)	1.99 (0.020)	19.83 (0.021)	95.52 (0.025)	-1614 (0.015)
$\mathbf{t}_1\mathbf{s}_2$(KR)($\tau=1$)	2.16 (0.013)	25.36 (0.015)	219.8(0.017)	-2652 (0.011)

Table 6.10: Non-dimensional values of the quantities given for $R_m=100$. The figure in parenthesis is the time at which the maximum occurred, discretised every 0.001 units.

each flow. The largest entries are for $\mathbf{s}_2$(MDJ) and $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau=1$, approximately ten times larger than most other values. We may contrast this with column one, showing the maximum rms field. The eigenmode, indicated by *, grows the most in rms over $0\le t\le 0.2$, but has a small associated Lorentz force. It therefore does not follow that large magnetic fields and large Lorentz forces go hand in hand. Clearly the Lorentz force always scales quadratically in the field, although the important multiplicative factor seems to vary tremendously depending on the situation.

The first three flows listed are $\mathbf{t}_1$, KR and STW which all behave similarly. Indeed both tables show that the absolute size and relative amplification of the Lorentz force is approximately equal to that of the field and current, also apparently true for the eigenmode solution. In all other

flows, the rms Lorentz force is significantly larger than the product of that of the field and current. In the cases where the flows contain a large component of $\mathbf{s}_2$(MDJ) with $\tau \geq 1$, it is $O(10)$ bigger.

The differences between columns two and three of table 6.10 show the effect of an extra spatial derivative. In the $\mathbf{t}_1$, KR and STW flows there is a discrepancy of about a factor of 10, indicating that the important length scale is one tenth of the radius of the core. This is also shown in results for the $\mathbf{t}_1\,\mathbf{s}_2$ flows with $\tau = 0.5$; however all the $\mathbf{s}_2$ flows (and to some extent the $\mathbf{t}_1\,\mathbf{s}_2$ flows with $\tau = 1$) show a length scale of 1/20. This difference is unsurprising when we bear in mind the very different mechanisms of field generation, that of either shearing by toroidal flow or stretching by poloidal flow.

Column five of table 6.10 shows the non-dimensional work rate of the Lorentz force on the flow, being equal and opposite to the rate of work of the flow on the magnetic field. In a dynamical system, this gives the rate of energy loss from the velocity into the magnetic field which must be matched by flow-generative terms to retain a steady state. In the Earth, the energy source is buoyancy, whereas in dynamo experiments the flows are externally driven by propellers. It is therefore of interest to quantify the efficiency ratio, defined as the ratio of the energy loss rate and the amount of field actually generated. In comparing column five to column two of table 6.10, it is clear that $\mathbf{t}_1$, KR and STW flows are in this sense the most efficient, despite not actually generating the most magnetic field. The most inefficient are the $\mathbf{s}_2$(MDJ) and $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 1$ flows.

Lastly, we note that the time at which the maxima of the rms magnetic field, current and Lorentz force are attained, are in all cases approximately the same. It is noteworthy however that the rate of working of the Lorentz force on the flow always peaks before all the other quantities, sometimes by a substantial fraction. This can be explained at least in comparison to the rms magnetic field and in the case of constant Ohmic dissipation, in the following way. Recall that the equation for magnetic energy is

$$\frac{dM}{dt} = R_m \int_V \mathbf{B}\cdot \mathsf{e}\,\mathbf{B}\, dV - \int_V |\boldsymbol{\nabla}\times\mathbf{B}|^2\, dV \tag{6.30}$$

so that $\mathbf{B}_{rms}$ grows only when the rate of work of the Lorentz force exceeds that of Ohmic dissipation. The peak of the magnetic energy, before its decay, corresponds to an exact balance between the two terms. At some earlier time, the rate of working of the Lorentz force had to be larger in order to increase M. Therefore the maximum of the field line stretching must occur before the maximum of $\mathbf{B}_{rms}$.

Table 6.11 shows dimensional values of the ratio of the Lorentz to the Coriolis force,

Flow	Max rms **B** (mT)	Max rms Lorentz (mN)	Max Ratio	Max work rates (GW) Lorentz	Ohmic heating
$\mathbf{t}_1$	2.22	5.89	0.088	-18.07	3.44
KR†	2.12	5.41	0.081	-16.40	3.19
STW†	1.60	3.82	0.058	-11.77	2.62
$\mathbf{s}_2$(MDJ)	4.54	931.03	21.33	-661.88	153.79
$\mathbf{s}_2$(KR)	2.16	75.86	1.74	-64.47	15.24
$\mathbf{s}_2$(IC)	1.90	68.83	1.58	-54.19	12.65
$\mathbf{t}_1\mathbf{s}_2$(MDJ)$(\tau = 0.5)$	2.71	128.79	2.05	-98.86	23.18
$\mathbf{t}_1\mathbf{s}_2$(MDJ)$(\tau = 0.5)^*$	1.14	3.76	0.060	-8.08	1.64
$\mathbf{t}_1\mathbf{s}_2$(MDJ)$(\tau = 1)$	3.33	343.84	6.10	-247.23	57.57
$\mathbf{t}_1\mathbf{s}_2$(KR)$(\tau = 0.5)$	1.63	14.53	0.23	-20.78	4.84
$\mathbf{t}_1\mathbf{s}_2$(KR)$(\tau = 1)$	1.88	37.76	0.67	-38.57	8.95

Table 6.11: Geophysically scaled values of the maximum quantities given for computations at $R_m = 100$. The ratio is that of the dimensional Lorentz force to the dimensional Coriolis force. The Lorentz force work rate and Ohmic heating rate are given in GW.

the maximum value of the work done by the Lorentz force on the flow, and the maximum Ohmic heating rate. We first draw the reader's attention to the eigenmode solution $(*)$, which has been scaled to the Earth using the external dipole coefficient h_1^1 rather than g_1^0. Recall that this is a factor of six smaller. Therefore, plausibly, the Lorentz force and Ohmic heating could be 36 times larger if we rescaled in a different way. Nonetheless, we see that $\mathbf{B}_{rms}$ is 1.14 mT, a similar magnitude to that for the other flows, and that it has the smallest rms Lorentz force shown. On these grounds then, we might speculate that it might be possible to see eigenmode solutions in the Earth of this form, assuming the associated $\mathbf{t}_1\,\mathbf{s}_2$ flow is a stationary solution of the geostrophic balance, since the flow would to first order remain unperturbed by the Lorentz force and hence remain stationary.

Column six shows the dimensional Ohmic heating; in each case, it is well within the geophysical values of 1–2 TW. However, in the preceding sections we have seen that for higher values of R_m, such bounds may be exceeded. For example, the $\mathbf{s}_2$(MDJ) flow with $R_m = 1000$ has an associated Ohmic heating rate of 27.18 TW; this solution could therefore be excluded on physical grounds. In practice however, if the growth was ephemeral (e.g. in field recovery after reversals), the energy dissipated as heat would contribute little to the global energy budget of the Earth and it would be difficult to justify barring a particular flow type from being important in the Earth's dynamo. Flows containing a substantial fraction by rms of $\mathbf{s}_2$(MDJ) have the highest Lorentz forces, their associated working rates and Ohmic heating rates.

Perhaps the geophysically most important results are in column four, where we list the

ratios of the dimensional Lorentz force and Coriolis force. Here we see that the $\mathbf{s}_2$(MDJ), $\mathbf{s}_2$(KR), $\mathbf{s}_2$(IC) and $\mathbf{t}_1\,\mathbf{s}_2$(MDJ) $\tau = 0.5{,}1$ flows all have a ratio of order unity or greater. This means that in the fully dynamical system, the flow would be significantly altered by the generated transient magnetic field growth, even at a geophysically low value of $R_m = 100$. The $\mathbf{t}_1$, KR and STW flows have a tiny ratio, indicating that if the flow was dynamically self-consistent, it would be almost unchanged by the presence of the magnetic field that it generated.

The ratio of Ohmic heating and the rate of work of the Lorentz force lies in the range $0.19 - 0.24$, thus appearing to be independent of flow. If we use the value of 2 TW for dissipation in the Earth, then if these values carry over, the flow in the outer core must do work at a rate of about 0.4 TW on the magnetic field. This agrees in order of magnitude to the total heat flux leaving the core, usually taken to be about 7 TW (e.g. *Roberts et al.*, 2003).

6.13 Concluding remarks

In this chapter, we have studied the ability of various flows to support transiently (and in one case exponentially) growing magnetic field solutions. Without exception, the dominant symmetry in the non-normal growth was the geophysically relevant $m = 0\ E^A$. This is in direct conflict with Cowling's theorem, but nonetheless may go some way in explaining the large presence of the axial dipole field in the Earth. We also saw that in many cases, especially between the $\mathbf{t}_1$, KR and STW flows that this growth was robust: small changes to the flow taking little effect in the generated fields. The dominant growth mechanism in these flows was shearing of the poloidal field into toroidal field by strong differential rotation. This was in contrast to the flows containing a substantial $\mathbf{s}_2$ component, whose generation mechanism was advection towards the z-axis following by field line stretching, increasing the poloidal field energy. For $R_m = 1000$, transient energy growth of $O(1000)$ was possible by both the $\mathbf{t}_1$ and $\mathbf{s}_2$(MDJ) flows, the relevant mechanisms explained by simple physical processes.

Kinematic dynamo theory tries to explain the growth of magnetic fields under stationary flows until the Lorentz force becomes important. We showed that a large rms field strength did not necessarily couple with a large rms Lorentz force, so that the boundary of the kinematic regime could not simply be dictated by 'large' magnetic fields. We also investigated the ratio of the associated Lorentz force with the Coriolis force, and found it in five cases to be $O(1)$ or greater, indicating the breakdown of kinematic theory, at least at $R_m = 100$. Therefore, in such cases, this would explain the form of magnetic field growth in the kinematic regime without appealing to eigenmode solutions, whose growth might be doubtful given their extreme sensitivity

to the flow. In other cases, particularly in the $\mathbf{t}_1$, KR and STW flows, the ratio of Lorentz to Coriolis forces was found to be tiny, and although these were not able to show the most transient magnetic energy growth, these flows would remain almost untouched by the Lorentz force.

Bringing together the approximate asymptotic scalings of the energy envelope height of the form R_m^γ for the various flows and the results of the previous section, we see that if $\gamma > 1$ then the ratio of the Lorentz to Coriolis forces will asymptotically increase, the latter being linear in R_m. If $\gamma < 1$ however (true for flows with $\mathbf{s}_2$(KR) or $\mathbf{s}_2$(IC) components), if R_m is large enough then the Lorentz force will always be of second order importance compared to the Coriolis force.

From an observational point of view, we found that transient energy growth was not able to explain the records at Steens Mountain of swift directional and intensity changes over a period of weeks. Indeed, the VGP plots that we computed over 38,000 years showed little movement from the geographic north pole. However, transient growth could account for the the fast recovery of field intensity after magnetic reversals occurring over a typical time period of 10,000 years, particularly with flows having a large poloidal overturn component. Even though the field growth was rapid (on a timescale of less than 1000 years), there was a significant time delay in the field reaching the Earth's surface, increasing the time taken for the signal to be recorded in the paleomagnetic data. This brings the field energy growth timescale more in line with the paleomagnetic data.

Chapter 7

Concluding remarks

7.1 Discussion of work carried out

In this thesis, we have investigated the instability of magnetic fields in geophysically motivated stationary spherical flows of conducting fluid. The study has been two-fold: firstly we addressed the onset problem of total magnetic energy growth. Using a variational method, we were able to formulate the relevant equations that could be solved numerically using a Galerkin method, showing excellent convergence properties. We were able to find which magnetic field structures grew in magnetic energy the fastest, their associated growth rates and the detail of the physical mechanism that was applicable.

Secondly, we looked at finite time transient total magnetic energy growth in the same stationary flows and investigated how large the energy can grow before its eventual decay (should infinite growth not be supported). In the light of the many anti-dynamo theorems in the literature, should any fields which are barred from being infinitely sustainable exhibit growth, this may force us to re-evaluate the relevance of such principles, since over the course of the lifetime of the instability, the flow pattern may change and therefore the fields may not get an opportunity to enter their decaying phase. This situation was first considered by *Backus* (1958) who studied a cycle of flows, none of which could support dynamo action individually, but when spliced together in a time sequence could sustain magnetic fields indefinitely. Therefore, for example the geophysically dominant axisymmetric field, barred by Cowling's theorem from being sustainable by a dynamo mechanism, could still be of great importance in the geodynamo. In studying transiently growing magnetic fields, we may also not only quantify such things as the magnitude and effect of the associated Lorentz force, but again what physical mechanisms are responsible for the growth.

The starting point for our so called kinematic study is the magnetic induction equation in a spherical geometry, describing the time evolution of a magnetic field under the influences of the conducting outer core moving at velocity **u** and magnetic diffusion. Ideally, we would substitute the true structure of the flow field into this equation; unfortunately this is unknown and probably will remain so for the foreseeable future. As an alternative, we may attempt to solve the Navier-Stokes equation, giving the fully dynamical time dependence of **u**. This is formidably complex, and in the parameter regime relevant to the Earth, numerically intractable at the current time. Therefore, we motivate our choice of flow by physical processes: differential rotation (represented by $\mathbf{t}_1$) and axisymmetric convection (represented by $\mathbf{s}_2$). Geophysically we hope these flows are stationary solutions of the magnetostrophic balance, where buoyancy, pressure and the Coriolis force are in equilibrium. If a magnetic field grows and its Lorentz force becomes comparable in magnitude to any of these geostrophic components, then the flow will change and we enter a non-linear regime, signalling the demise of the kinematic assumption. We can never hope to capture all the features of the flow by simple models; indeed, we hope that only the large scale flow components are important. Therefore we must make sure that any dynamo processes that we come up with are robust, that is, not critically dependent on the choice of flow.

Historically, in attempts to find a generative mechanism inside the Earth's core that could explain the presence of the geomagnetic field, workers have applied linear eigenvalue methods to the induction equation. In some flows, exponentially growing eigenmodes were found, the existence of which have been attributed to favourable meridional circulation (*Roberts*, 1972b) and flow helicity (*Gubbins et al.*, 2000a). It was also found that a conducting layer at the top of the core favoured dynamo action (*Hutchenson and Gubbins*, 1994). However, the main drawback of such models is their extreme sensitivity to the precise choice of stationary flow. We saw in the pseudospectra of chapter five that a change in the flow of a root mean squared value of a fraction of 1% could result in the cessation of dynamo action; that is, the unstable eigenvalues are shifted into the stable half-plane. This means that we cannot attribute a physical mechanism to the generation of magnetic fields in this manner, since a minute change in the flow would not alter our description of the dynamo process, but it might switch the dynamo off. Magnetic fields in nature are a robust phenomenon: not only has the Earth had a field for 3 billion years or so, but other planets and stars (including the sun) possess working dynamos too. We would therefore not expect the details of the flow to be of critical importance.

Our study of non-eigenmode instabilities is motivated by the success in the fluid dynamics literature of using techniques, associated with non-normal problems, to explain transi-

tion to turbulence of (non-magnetic) plane Poiseuille and Couette flows. In these cases, flow perturbations transiently grew large enough to push the system into a non-linear regime, before entering their eventual decaying phase. Non-normal problems also raise the issue of numerical difficulties when computing eigenvalues. This is due to their expanded pseudospectra (section 5.1.1) which means that larger regions of the complex plane than our convergence tolerance allows behave as 'almost eigenvalues', that is, $|\mathsf{A}\,\mathbf{v} - \sigma\,\mathbf{v}| \approx 0$ for the controlling matrix A and some value σ, not within the convergence tolerance of the true eigenvalue λ. In the induction equation, the non-normality is caused by the velocity-field interaction, scaling linearly in R_m. Therefore as R_m increases, it becomes more and more difficult to find accurate eigenvalues, not only because higher truncation levels are required but due to the expanded pseudospectra. This means we must be cautious when locating the eigenvalues for large R_m.

In the onset problem of magnetic energy instability we found robust results, in that small changes in the flow did not significantly affect the outcome. In particular, the $\mathbf{t}_1$, KR and STW flows, all dominated by differential rotation and differing only by a few percent in rms, showed almost identical characteristics, in contrast to linear stability analysis showing that only KR and STW could support growing eigenmodes. The critical field excited in these cases was $m = 1\ E^S$, the associated generation mechanism being field line stretching. In the case of convectively driven flows represented by $\mathbf{s}_2$ components, all three particular choices, including that containing a quiescent region representing the conducting solid inner core, showed similar energetic instabilities. These were all of $m = 0\ E^A$ symmetry and were explained by field line stretching at the location of radial upwelling, near the origin. At first sight we might have expected the presence of an inner core to disrupt this process, since the flow is effectively pushed away from the origin and squashed towards the CMB. However, despite the large scale flow pattern being slightly altered, the mechanism of instability remains unchanged. This means that this magnetic instability can occur with or without an inner core, lending weight to the robustness of this mechanism. When combining together the differentially rotating and convective flows forming a $\mathbf{t}_1\,\mathbf{s}_2$ configuration, we found that if the ratio of poloidal to toroidal flow energies was above 0.2, then the $m = 0\ E^A$ field symmetry became favoured. This is a significant finding with respect to the geodynamo, since we may possibly explain the identical dominant symmetry by a flow pattern in the outer core which is predominately poloidal, a likely situation since the main energy source for the geodynamo is compositional convection.

Additional findings from the onset study included the improvement of the bound of *Proctor* (1977a) on the minimum R_m required for dynamo action, by a factor of 5–14 times, depending on the flow. Generically we require $R_m = O(10)$, based on the rms flow velocity, to

get magnetic energy growth. We also investigated the magnitude of the instantaneous magnetic energy growth as R_m increased, and found it to scale asymptotically linearly in R_m. Physically this indicates that the higher the flow speed, the easier it is to grow magnetic fields, at least instantaneously. This holds, in particular, for the $m = 0$ symmetries in the case of flows with a large poloidal component. Therefore although axisymmetric fields may not be infinitely sustainable (by applying Cowling's theorem), their instantaneous growth is asymptotically unbounded. We compare this asymptotic behaviour to that found in eigenvalue analysis, in which stable eigenvalues often decrease in real part as R_m increases. This implies somewhat counter-intuitively that it becomes increasingly more difficult to generate field, with more vigorous flows.

In the study of finite time growth of magnetic energy in chapter six, we found without exception that the dominant symmetry giving rise to the most transient growth was the geophysically relevant $m = 0$ E^A. Although changes in the flow affected the amount of energy amplification possible, the physical mechanisms remained the same. Flows dominated by a $\mathbf{t}_1$ flow component sheared poloidal field lines into toroidal field, whereas those dominated by $\mathbf{s}_2$ components advected field towards the locations of radial upwelling near the origin where it was radially stretched, similarly to the instantaneous growth mechanism. Although large transient energy growth of $O(1000)$ was possible with $R_m = 1000$, the initial fields were optimally chosen and we would not expect to see such amplification in the Earth. However, since the mechanisms were robust, it is clear that substantial growth could easily take place; this in particular could explain the swift recovery of the field after magnetic reversals. That the dominant symmetry in this study was axisymmetric is important as it seemingly directly opposes Cowling's theorem, barring all fields of this type from being infinitely sustained. Although such fields will eventually decay, the Lorentz force may have grown sufficiently to push the system into a non-linear regime before the energy can begin to decrease. Therefore, neglecting axisymmetric fields on the grounds of non-infinite sustainability may be folly, since these transient mechanisms may be fundamentally important in the geodynamo.

We computed the Lorentz forces associated with this transient energy growth and found that large magnitudes were not necessarily associated with large magnetic fields. Although the Lorentz force scales quadratically in the magnetic field, the all important multiplying factor seems to vary enormously, depending on the particular situation. In particular, those flows dominated by a $\mathbf{t}_1$ component, although not showing the largest transient effects for $R_m = 100$ had the smallest associated Lorentz force. Indeed, its maximum value was about a tenth of the size of the geophysically scaled Coriolis force, thus flows of this type would not lead to transition to a non-linear state in the Earth, at least with a geophysical choice of R_m. On the other

hand, flows dominated by an $\mathbf{s}_2$ component showed moderate to large transient growth and very high Lorentz forces for $R_m = 100$, in most cases being of a similar magnitude to the Coriolis force when appropriately scaled. This observation does not depend on the particular flow choice involved and again gives a robust result. Therefore, if such transient growth took place in a convectively driven flow, it would almost certainly lead to a regime where the Lorentz force was important, signalling the breakdown of the kinematic regime. Another geophysical constraint that we can apply is that of Ohmic dissipation, the rate at which magnetic energy is lost through Ohmic heating of the outer core. *Roberts et al.* (2003) give an upper bound of 1–2 TW that we can test against our computed fields. For $R_m = 100$, all solutions, when scaled to the Earth, have a maximal Ohmic heating of O(1–100) GW, well within this bound. For $R_m = 1000$, the $\mathbf{s}_2$(MDJ) flow gives a heating rate of 27.18 TW and therefore could potentially be excluded on energetic grounds. However, we note that in this case, there is also a very high Lorentz force, in magnitude 334 times that of the Coriolis force, so that the flow pattern would not remain stationary and the energy growth may not be as great as the computations suggest. Additionally if the growth was ephemeral, for example in a recovery after a magnetic reversal, although the Ohmic dissipation is high it will only be so for a short period and its contribution to the global energy budget will be small.

From a geomagnetic observational viewpoint, transient magnetic field growth can explain the rapid recovery of the field after an intensity low during a reversal, typically of around 10–20% of the pre-transitional field (*Merrill and McFadden*, 1990). Paleomagnetic records indicate in addition that the post-transitional field is in most cases more stable than that before the transition (*Bogue*, 2001), suggesting that the axisymmetric poloidal dipole component dominates this growth. We find that our calculations can explain both of these phenomena if the flow has a significant poloidal component, for example the $\mathbf{s}_2$(MDJ) flow giving an rms increase in field of about 30 for $R_m = 1000$. Although the $\mathbf{t}_1$ flow also gives this energy amplification, the generated field is principally toroidal which would not be seen at the Earth's surface. The timescale on which this growth takes place is around 1000 years, roughly the same as typical estimates of 1000–10,000 years (*Valet and Meynadier*, 1993), and sufficiently rapid that the flow pattern might be modelled as stationary, given the turnover time of the core of around 500 years. We have also shown that observations on the Earth's surface may be delayed by several thousand years relative to the actual growth occurring inside the core. This means that we may be able to reduce the observational upper limit of 10,000 to 8,000 years or so, more in line with our computed values.

Our results cannot explain extremely swift (over days or weeks) field directional and

intensity changes, such as those seen in the data from Steens Mountain (*Prévot et al.*, 1985; *Mankinen et al.*, 1985), at least with a geophysical value of R_m. Indeed, VGP paths associated with transient magnetic energy growth have also been computed, and we found little overlap with those from field transition data (e.g. *Valet and Herrero-Bervera*, 2003). Nonetheless, if the flow patterns were more complicated and indeed time dependent, it is likely that other non-axisymmetric field components would be of importance and thus the same physical mechanism may, with a different velocity field, be able to provide VGP paths similar to observations, the virtual geomagnetic pole wandering on a timescale of 100–1000 years.

It is interesting to note the comparison of whole sphere convective flows and those containing an inner core. In both the onset and transient growth studies, neither the dominant field symmetry nor the physical mechanism responsible for field growth is affected. However, in both cases we find a lesser capacity for instability with an inner core, due primarily to the reduced flow region. The eigenvalue study of *Sarson and Gubbins* (1996) showed that the inclusion of the inner core had significant effects; however, they also found that it may play a role in symmetry selection, something that our results do not indicate. Our conclusion is that a dynamo may operate more efficiently without an inner core. *Gubbins et al.* (2003) argue on energetic grounds that the inner core must be at least three billion years old in order to supply the compositional convection required to drive the dynamo. Convective flows without an inner core would be driven by thermal convection, attributed to the slow cooling of the Earth after its formation. Our calculations indicate that although this method of convection is far more inefficient than that of composition, that its dynamo capability is greater and so we may be able to reduce the age of the inner core. This may explain other studies (e.g. *Labrosse et al.*, 2001), based on thermodynamic calculations that suggest that the inner core is only one billion years old.

Since the toroidal field can never escape from the outer core through the insulating mantle, its magnitude may never be known. If transient mechanisms are important in regenerating the field after a magnetic reversal however, the poloidal flows causing the growth will generate mainly poloidal field energy. If during a reversal the field strength was only 10% of its pre-transitional value, then after the recovery around 90% of the field will be poloidal. This possibly makes the field initially stable preventing other immediate reversals, in agreement with paleomagnetic studies (*Bogue and Paul*, 1993). There may well be significant poloidal decay as the field eventually settles down to its time averaged state, and the field may plausibly be somewhere in the region of 30–90% poloidal in energy.

7.2 Extensions of this study

To extend this study, following *Hutchenson and Gubbins* (1994) and *Sarson and Gubbins* (1996) we could undertake the inclusion of a thin conducting region on top of the core, representing a stratified layer of light buoyant material or the D'' layer at the base of the mantle *Garnero et al.* (1998) that could have large conductance *Holme* (1998). *Bullard and Gubbins* (1977) found that the insulating mantle was partially responsible for the difficulty in finding growing kinematic dynamo field solutions, trapping the current in thin sheets near the boundary leading to large Ohmic dissipation. Subsequent studies have indicated that including a thin conducting region is favourable to the eigenmode instability of magnetic fields, either by reducing the critical magnetic Reynolds number or allowing fields which otherwise decayed to grow. We could apply the analyses of instantaneous energy growth and finite time transience by including a thin quiescent region just below the CMB in our flows, rescaling R_m to account for the change in radius of effective dynamo region. Indeed, we could vary the depth of this layer to see if there was an optimal thickness, as predicted in the eigenmode case by *Kaiser and Tilgner* (1999). A further extension would be to see the effect of an insulating inner core. This could be simply accommodated by extending our basis functions radially inwards from the flow region, in the same way as we did into the mantle and using flows defined only in the outer core.

We have suggested at various points in this thesis that by some mechanism, transient growth or instantaneous instability of magnetic fields might be continuously excited in the outer core. It is plausible that stochastic forcing could provide such a stimulus, driven by the turbulent underlying flow. Mean field dynamo theory is concerned with a similar problem, but in which the effect of the small scale processes, beyond the resolution of the model, are parameterised by the so called α-effect, represented by the spatially dependent tensor α (e.g. *Moffatt*, 1978). This is a useful concept, although in general α will fluctuate in time and typical formulations do not capture this important aspect. Such a stochastically driven study has already been carried out in a cylindrical geometry (*Farrell and Ioannou*, 1999b) who find that axisymmetric fields can be sustained by such a mechanism. In fact, it is the fields which dominate the non-normal analysis of the large scale flow which are physically manifested, seemingly unaware of the small scales providing the forcing. Understanding the effect of small scales is an important issue, since it is unlikely that fully dynamical models which can cope with the huge scale disparity of the geodynamo will be available, at least in the foreseeable future. Although the spherical geometry of the Earth differs from the cylindrical geometry of *Farrell and Ioannou* (1999a), their transience calculations show similar results to those that we found, such as the dominant

$m = 0$ field symmetry and the scaling of the energy envelope height of $R_m^{2/3}$, which in our case was R_m^{γ} with $0.7 \leq \gamma \leq 2$, depending on the flow. These indicate that a similar analysis applied to the Earth may show an identical mechanism.

We found in chapter five that the $\mathbf{s}_2$(MDJ) flows supported magnetic energy instability with the lowest value of R_m, but is this the 'best' possible flow configuration? We can envisage an analysis on the same lines as *Love and Gubbins* (1996b) and *Holme* (2003) in addressing such a problem as an optimisation over the choice of the defining flow scalars. This may also include simultaneously minimising the Ohmic dissipation and possibly the Lorentz force associated with such fields. Additionally, the $\mathbf{s}_2$(MDJ) flow supported the most transient magnetic energy growth, at least in the geophysically appropriate range $R_m \leq 1000$. We could apply a similar study to investigate whether a more 'efficient' flow exists for these values of R_m, incorporating similar Ohmic dissipation or Lorentz force constraints. Such optimisation schemes give an alternative way of choosing the flow pattern that we study, instead of being via the fairly arbitrary though physically motivated manner used here. A different though attractive scheme would be to study a flow produced by either non-magneto rotating convection (e.g. *Zhang*, 1992) or by a fully dynamic magneto-convection model (e.g. *Kuang and Bloxham*, 1997) and see if the magnetic instabilities predicted agreed with those of our simple flow models. In the latter case, it would be also possible to compare the true subsequent time behaviour of the magnetic field with that as computed by transience calculations, to see if similarities can be found.

References

Abramowitz, M., and I. A. Stegun, *Pocketbook of mathematical functions*, Verlah Harri Deutsch, 1984.

Acheson, D., *Elementary Fluid Dynamics*, OUP, 1990.

Backus, G., The axisymmetric self-excited fluid dynamo, *Ann. Phys.*, **125**, 500–522, 1957.

Backus, G., A Class of Self-Sustaining Dissipative Spherical Dynamos, *Ann. Phys.*, **4**, 372–447, 1958.

Backus, G., R. Parker, and C. Constable, *Foundations of Geomagnetism*, CUP, 1996.

Bloxham, J., and A. Jackson, Time-dependent mapping of the magnetic field at the core-mantle boundary, *J. Geophys. Res.*, **97**, 19,537–19,563, 1992.

Bogue, S., Geomagnetic field behavior before and after the Kauai reverse-normal polarity transition, *J. Geophys. Res.*, **106**, 447–461, 2001.

Bogue, S., and H. Paul, Distinctive field behavior following geomagnetic reversals, *Geophys. Res. Lett.*, **20**, 2399–2402, 1993.

Borba, D., K. Riedel, W. Kerner, G. Huymans, M. Ottaviani, and P. Schmid, The pseudospectrum of the resitive magnetohydrodynamics operator: Resolving the resistive Alvén paradox, *Phys. Plasmas*, **1**, 3151–3160, 1994.

Boyd, J., *Chebyshev and Fourier Spectral Methods*, Dover, 2001.

Braginsky, S., Theory of the hydromagnetic dynamo, *Soviet Phys. JETP*, **20**, 1462–1471, 1965.

Browder, F., The dirichlet problem for linear elliptic equations of arbitrary even order with variable coefficients, *Proc. Nat. Aced. Sci.*, **38**, 230–235, 1952.

Browder, F., On the eigenfunctions and eigenvalues of the general linear elliptic differential operator, *Proc. Nat. Aced. Sci.*, **39**, 433–439, 1953.

Brummell, N., F. Cattaneo, and S. Tobias, Linear and nonlinear dynamo action, *Phys. Lett. A*, **249**, 437–442, 1998.

Buffett, B., Estimates of heat flow in the deep mantle based on the power requirements for the geodynamo, *Geophys. Res. Lett.*, **29**, 7.1–7.4, 2002a.

Buffett, B., Modelling of nutation and precession: Effects of electromagnetic coupling, *J. Geophys. Res.*, **107**, 5.1–5.15, 2002b.

Bullard, E., The magnetic field within the earth, *Proc. R. Soc. Lond. A*, **197**, 433–453, 1949.

Bullard, E., and H. Gellman, Homogeneous dynamos and terrestrial magnetism, *Phil. Trans. R. Soc. Lond. A*, **247**, 213–278, 1954.

Bullard, E., and D. Gubbins, Generation of magnetic fields by fluid motions of global scale, *Geophys. Astrophys. Fluid Dyn.*, **8**, 43–56, 1977.

Busse, F., A necessary condition for the Geodynamo, *J. Geophys. Res.*, **80**, 278–280, 1975.

Busse, F., Convective flows in rapidly rotating spheres and their dynamo action, *Phys. Fluids A*, **14**, 1301–1314, 2002.

Camps, P., M. Prévot, and R. Coe, Revisiting the initial sites of geomagnetic field impulses during the Steens Mountain polarity reversal, *Geophys. J. Int.*, **123**, 484–506, 1995.

Camps, P., R. Coe, and M. Prévot, Transitional geomagnetic impulse hypothesis: Geomagnetic fact or rock-magnetic artifact?, *J. Geophys. Res.*, **104**, 17,747–17,758, 1999.

Canuto, C., M. Hussaini, A. Quarteroni, and T. Zang, *Spectral Methods for Fluid Dynamics*, Springer-Verlag, 1988.

Chandrasekhar, S., *Hydrodynamic and hydromagnetic stability*, OUP, 1961.

Chapman, S., Subcritical transition in channel flows, *J. Fluid Mech.*, **451**, 35–97, 2002.

Childress, S., *Théorie Magnétohydrodynamique de l'effet Dynamo*, 1969, Lecture notes.

Childress, S., and A. Gilbert, *Stretch, Twist, Fold: The Fast Dynamo*, vol. 37 of *Lecture Notes in Physics: Monographs*, Springer-Verlag, 1995, lecture notes.

Coe, R., M. Prévot, and P. Camps, New evidence for extraordinary rapid change of the geomagnetic field during a reversal, *Nature*, **374**, 687–692, 1995.

Constable, C., On rates of occurence of geomagnetic reversals, *Phys. Earth Planet. Int.*, **118**, 181–193, 2000.

Cowling, T., The magnetic field of sunspots, *Mon. Not. R. Astr. Soc.*, **94**, 39–48, 1933.

Davidson, P., *An introduction to magnetohydrodynamics*, CUP, 2001.

Dormy, E., J.-P. Valet, and V. Courtillot, Numerical models of the geodynamo and observational constraints, *Geochem. Geophys. Geosyst.*, **1**, 1–42, 2000.

Dudley, M., and R. James, Time-dependent kinematic dynamos with stationary flows, *Proc. R. Soc. Lond. A*, **425**, 407–429, 1989.

Dziewonski, A., and D. Anderson, Preliminary reference Earth model, *Phys. Earth Planet. Int.*, **25**, 297–356, 1981.

Elsasser, W., Induction effects in Terrestrial magnetism I, *Phys. Rev.*, **69**, 106–116, 1946.

Elsasser, W., Induction effects in Terrestrial magnetism III, *Phys. Rev.*, **72**, 821–833, 1947.

Farrell, B., and P. Ioannou, Generalized stability theory, part i: Autonomous operators, *J. Atmos. Sci.*, **53**, 2025–2040, 1996.

Farrell, B., and P. Ioannou, Optimal excitation of magnetic fields, *Astr. J.*, pp. 1079–1087, 1999a.

Farrell, B., and P. Ioannou, Stochastic dynamics of field generation in conducting fluids, *Astr. J.*, pp. 1088–1099, 1999b.

Fearn, D., Hydromagnetic flow in planetary cores, *Rep. Prog. Phys.*, **61**, 175–235, 1998.

Fornberg, B., *A practical guide to pseudospectral methods*, CUP, 1998.

Fowler, C., *The Solid Earth: An Introduction to Global Geophysics*, CUP, 1990.

Fuchs, H., K.-H. Rädler, and M. Rheinhardt, On self-killing and self-creating dynamos, *Astron. Nachr.*, **3**, 129–133, 1999.

Gailitis, A., et al., Detection of a Flow Induced Magnetic Field Eigenmode in the Riga Dynamo Facility, *Phys. Rev. Lett.*, **84**, 4365–4368, 2000.

Gailitis, A., et al., Magnetic Field Saturation in the Riga Dynamo Experiment, *Phys. Rev. Lett.*, **86**, 3024–3027, 2001.

Garnero, E., J. Revenaugh, Q. Williams, T. Lay, and L. Kellogg, Ultralow Velocity Zone at the Core-Mantle Boundary, *in The Core-Mantle Boundary Region, Ed. Gurnis et.al.*, **28**, 319–334, 1998.

Gerard-Varet, D., Amplification of small perturbations in a Hartmann layer, *Phys. Fluids A*, **14**, 1458–1467, 2002.

Gibbons, J., *Dynamo Models for the Earth's Magnetic Field*, 1998, Thesis.

Gibson, R., and P. Roberts, Some comments on the theory of homogeneous dynamos, *in Magnetism and the Cosmos, Ed. Hindmarsh et al. 1967*, pp. 108–120, 1967.

Glatzmaier, G., and P. Roberts, A three-dimensional self-consistent computer simulation of a geomagnetic field reversal, *Nature*, **377**, 203–209, 1995.

Gubbins, D., Numerical solutions of the kinematic dynamo problem, *Phil. Trans. R. Soc. Lond. A*, **274**, 493–521, 1973.

Gubbins, D., and J. Bloxham, Morphology of the geomagnetic field and implications for the geodynamo, *Nature*, **325**, 509–511, 1987.

Gubbins, D., and S. Gibbons, Three-Dimensional Dynamo Waves In A Sphere, *Geophys. Astrophys. Fluid Dyn.*, pp. 481–498, 2002.

Gubbins, D., and P. Roberts, Magnetohydrodynamics of the Earth's Core, *in Geomagnetism, Ed. Jacobs, J.A.*, **2**, 1–183, 1987.

Gubbins, D., and G. Sarson, Geomagnetic field morphologies from a kinematic dynamo model, *Nature*, **368**, 51–55, 1994.

Gubbins, D., and K. Zhang, Symmetry properties of the dynamo equations for paleomagnetism and geomagnetism, *Phys. Earth Planet. Int.*, **75**, 225–241, 1993.

Gubbins, D., C. Barber, S. Gibbons, and J. Love, Kinematic dynamo action in a sphere: I. Effects of differential rotation and meridional circulation on solutions with axial dipole symmetry, *Proc. R. Soc. Lond. A*, **456**, 1333–1353, 2000a.

Gubbins, D., C. Barber, S. Gibbons, and J. Love, Kinematic dynamo action in a sphere: II. Symmetry selection, *Proc. R. Soc. Lond. A*, **456**, 1669–1683, 2000b.

Gubbins, D., D. Alfè, G. Masters, D. Price, and M. Gillan, Can the Earth's Dynamo run on Heat alone?, *Geophys. J. Int.*, **155**, 609–622, 2003.

Herzenberg, A., Geomagnetic dynamos, *Phil. Trans. R. Soc. Lond. A*, **250**, 543–583, 1958.

Hide, R., and T. Palmer, Generalisation of Cowling's Theorem, *Geophys. Astrophys. Fluid Dyn.*, **19**, 301–309, 1982.

Hollerbach, R., and C. Jones, Influence of the Earth's inner core on geomagnetic fluctuations and reversals, *Nature*, **365**, 541–543, 1993.

Holme, R., Electromagnetic core-mantle coupling II: Probing deep mantle conductance, *in The Core-Mantle Boundary Region, Ed. Gurnis et.al.*, **28**, 139–151, 1998.

Holme, R., Optimised axially-symmetric kinematic dynamos, *Phys. Earth Planet. Int.*, p. in press, 2003.

Holme, R., and J. Bloxham, The treatment of attitude errors in satellite geomagnetic data, *Phys. Earth Planet. Int.*, **98**, 221–233, 1996.

Hutchenson, K., and D. Gubbins, Kinematic magnetic-field morphology at the core-mantle boundary, *Geophys. J. Int.*, **116**, 304–320, 1994.

Ivers, D., and R. James, Axisymmetric antidynamo theorems in compressible non-uniform conducting fluids, *Phil. Trans. R. Soc. Lond. A*, **312**, 179–218, 1984.

Jackson, A., Storm in a lava flow?, *Nature*, **377**, 685–686, 1995.

Jackson, A., Intense equatorial flux spots on the surface of the Earth's core, *Phys. Earth Planet. Int.*, **424**, 760–763, 2003.

Jacobs, J., *Geomagnetism*, vol. 2, Academic Press, 1987.

Kaiser, R., and A. Tilgner, Kinematic dynamos surrounded by a stationary conductor, *Phys. Rev. E*, **60**, 2949–2952, 1999.

Kallenrode, M.-B., *Space Physics*, 2nd ed., Springer-Verlag, 2001.

Kerswell, R., and A. Davey, On the linear instability of elliptic pipe flow, *J. Fluid Mech.*, **316**, 307–324, 1996.

Kono, M., and P. Roberts, Recent geodynamo simulations and observations of the geomagnetic field, *Rev. Geophys.*, **40**, 1013, doi:10.1029/2000RG00,102, 2002.

Kuang, W., and J. Bloxham, An Earth-like numerical dynamo model, *Nature*, **389**, 371–374, 1997.

Kumar, S., and P. H. Roberts, A three-dimensional dynamo problem, *Proc. R. Soc. Lond. A*, **344**, 235–258, 1975.

Labrosse, S., J.-P. Poirer, and J. Le Mouël, The age of the inner core, *Earth Planet. Sci. Lett.*, **190**, 111–123, 2001.

Laj, C., A. Mazaud, and R. Weeks, Geomagnetic reversal paths, *Nature*, **351**, 447, 1991.

Lanczos, C., *Linear differential operators*, D. Van Nostrand Co., 1961.

Langel, R., The Main Field, *in Geomagnetism, Ed. Jacobs, J.A.*, **1**, 249–512, 1987.

Larmor, J., How could a rotating body such as the sun become a magnet?, *Rept. Brit. Assoc. Adv. Sci.*, pp. 159–160, 1919.

Lathrop, D., W. Shew, and D. Sisan, Laboratory experiments on the transition to MHD dynamos, *Plasma Phys. Control. Fusion*, **43**, A151–A160, 2001.

Lilley, F., On kinematic dynamos, *Proc. R. Soc. Lond. A*, **316**, 153–167, 1970.

Livermore, P., and A. Jackson, On magnetic energy instability in spherical stationary flows, *Proc. R. Soc. Lond. A*, p. In press, 2004.

Lloyd, D., and D. Gubbins, Toroidal fluid motion at the top of the Earth's core, *Geophys. J. Int.*, **100**, 455–467, 1990.

Lortz, D., Exact solutions of the hydromagnetic dynamo problem, *Plasma Phys.*, **10**, 967–972, 1968.

Love, J., and D. Gubbins, Dynamos driven by poloidal flow exist, *Geophys. Res. Lett.*, **23**, 857–860, 1996a.

Love, J., and D. Gubbins, Optimized Kinematic Dynamos, *Geophys. J. Int.*, **124**, 787–800, 1996b.

Macrobert, T., *Spherical Harmonics*, vol. 3, Pergamon Press, 1967.

Malin, S., Historical Introduction to Geomagnetism, *in Geomagnetism, Ed. Jacobs, J.A.*, **1**, 1–49, 1987.

Mankinen, E., M. Prévot, C. Grommé, and R. Coe, The Steens Mountain (Oregon) Geomagnetic Polarity Transistion 1. Directional History, Duration of Episodes, and Rock Magnetism, *J. Geophys. Res.*, **90**, 10,393–10,416, 1985.

Merrill, R., and P. McFadden, Paleomagnetism and the Nature of the Geodynamo, *Science*, **248**, 345–350, 1990.

Merrill, R., and P. McFadden, Geomagnetic polarity transitions, *Rev. Geophys.*, pp. 201–226, 1999.

Merrill, R., M. M.W., and P. McFadden, *The Magnetic Field of the Earth*, vol. 63 of *International Geophysics Series*, Academic Press, 1998.

Moffatt, H., *Magnetic field generation in electrically conducting fluids*, CUP, 1978.

Molar, C., and C. Van Loan, Nineteen Dubious Ways to Compute the Exponential of a Matrix, *SIAM Rev.*, **20**, 801–836, 1978.

Nakajima, T., and M. Kono, Kinematic dynamos associated with large scale fluid motions, *Geophys. Astrophys. Fluid Dyn.*, **60**, 177–209, 1991.

Nataf, H.-C., Dynamo and convection experiments, *in Earth's core and lower mantle, Ed. Jones, C.A. et al.*, pp. 153–179, 2003.

Núñez, M., Plasma velocity in hydromagnetic dynamos, *J. Math. Phys.*, **43**, 3202–3206, 2002.

Olsen, N., A model of the geomagnetic field and its secular variation for epoch 2000 estimated from Ørsted data, *Geophys. J. Int.*, **149**, 454–462, 2002.

Olsen, N., et al., Ørsted Initial Field Model, *Geophys. Res. Lett.*, **27**, 3607–3610, 2000.

Parkinson, W., *Introduction to Geomagnetism*, Scottish Academic Press, 1983.

Peffley, N., A. Goumilevski, A. Cawthrone, and D. Lathrop, Characterization of experimental dynamos, *Geophys. J. Int.*, **142**, 52–58, 2000a.

Peffley, N., A. Goumilevski, A. Cawthrone, and D. Lathrop, Toward a self-generating magentic dynamo: The role of turbulence, *Phys. Rev. E*, **61**, 5287–5294, 2000b.

Pekeris, C., Y. Accad, and B. Shkoller, Kinematic dynamos and the Earth's magnetic field, *Phil. Trans. R. Soc. Lond. A*, **275**, 425–461, 1973.

Peyronneau, J., and J. Poirer, Electrical conductivity of the Earth's lower mantle, *Nature*, **342**, 537–539, 1989.

Ponomarenko, Y., Theory of the hydromagnetic generator, *J. Appld. Mech. Tech. Phys.*, **14**, 775–778, 1973.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipies in Fortran 77*, 2nd ed., CUP, 1992.

Prévot, M., E. Mankinen, C. Grommé, and R. Coe, How the geomagnetic field vector reverses polarity, *Nature*, **316**, 230–234, 1985.

Proctor, M., On Backus' necessary condition for dynamo action in a conducting sphere, *Geophys. Astrophys. Fluid Dyn.*, **9**, 89–93, 1977a.

Proctor, M., The role of Mean Circulation in Parity Selection by Planetary Magnetic Fields, *Geophys. Astrophys. Fluid Dyn.*, **8**, 311–324, 1977b.

Reddy, S., and D. Henningson, Energy growth in viscous channel flows, *J. Fluid Mech.*, **252**, 209–238, 1993.

Rikitake, T., and Y. Hagiwara, Non-steady Bullard-Gellman Dynamo Model (1), *J. Geomagn. Geoelectr.*, **20**, 57–65, 1968.

Roberts, G., Spatially periodic dynamos, *Phil. Trans. R. Soc. Lond. A*, **266**, 535–558, 1970.

Roberts, G., Dynamo action of fluid motions with two-dimensional periodicity, *Phil. Trans. R. Soc. Lond. A*, **271**, 411–454, 1972a.

Roberts, P., *An introduction to magnetohydrodynamics*, Longmans, 1967.

Roberts, P., Kinematic dynamo models, *Phil. Trans. R. Soc. Lond. A*, **272**, 663–698, 1972b.

Roberts, P., Dynamo Theory, *Annu. Rev. Fluid Mech.*, **24**, 459–512, 1992.

Roberts, P., Fundamentals of dynamo theory, *in Lectures on Solar and Planetary Dynamos, Ed. M.R.E. Proctor and A.D. Gilbert*, 1994.

Roberts, P., C. Jones, and A. Calderwood, Energy fluxes and ohmic dissipation in the earth's core, *in Earth's core and lower mantle, Ed. Jones, C.A. et al.*, 2003.

Sarson, G., *Kinematic Dynamo Calculations for Geomagnetism*, 1994, Thesis.

Sarson, G., Kinematic dynamos driven by thermal wind flows, *Proc. R. Soc. Lond. A*, **459**, 1241–1259, 2003.

Sarson, G., and F. Busse, The kinematic dynamo action of spiralling convective flows, *Geophys. J. Int.*, **133**, 140–158, 1998.

Sarson, G., and D. Gubbins, Three-dimensional kinematic dynamos dominated by strong differential rotation, *J. Fluid Mech.*, **306**, 223–265, 1996.

Sarson, G., and C. Jones, A convection driven geodynamo reversal model, *Phys. Earth Planet. Int.*, **111**, 3–20, 1999.

Schmid, P., and D. Henningson, Optimal energy density growth in Hagen-Poiseuille flow, *J. Fluid Mech.*, **277**, 197–225, 1994.

Schmid, P., and D. Henningson, *Stability and Transition in Shear Flows*, Springer, 2001.

Shankland, T., J. Peyronneau, and J.-P. Poirer, Electrical conductivity of the Earth's lower mantle, *Nature*, **366**, 453–455, 1993.

Shure, L., R. Parker, and G. Backus, Harmonic splines for geomagnetic modelling, *Phys. Earth Planet. Int.*, **28**, 215–229, 1982.

Stern, D., A Millenium of Geomagnetism, *Rev. Geophys.*, **40**, 1–30, 2002.

Stieglitz, R., and U. Müller, Experimental demonstration of a homogeneous two-scale dynamo, *Phys. Plasmas*, **13**, 561–564, 2001.

Sweet, D., E. Ott, T. Antonsen Jr., D. Lathrop, and J. Finn, Blowout bifurcations and the onset of magnetic dynamo action, *Phys. Plasmas*, **8**, 1944–1952, 2001.

Takeuchi, H., and Y. Shimazu, On a self-exciting process in magneto-hydrodynamics, *J. Geophys. Res.*, **58**, 497–518, 1953.

Tilgner, A., Onset of dynamo action in an axisymmetric flow, *Phys. Rev. E*, **66**, 017,304–1, 2002.

Tough, J., Nearly Symmetric Dynamos, *Geophys. J. R. Astr. Soc.*, **13**, 393–396, 1967.

Trefethen, L., Pseudospectra of linear operators, *Siam Rev*, **39**, 383–406, 1997.

Trefethen, L., A. Trefethen, S. Reddy, and T. Driscoll, Hydrodynamic stability without eigenvalues, *Science*, **261**, 578–584, 1993.

Ultré-Guérard, P., and J. Achache, Core flow instabilities and geomagnetic storms during reversals: The Steens Mountain impulsive field variations revisited, *Earth Planet. Sci. Lett.*, **135**, 91–99, 1995.

Valet, J.-P., and E. Herrero-Bervera, Some characteristics of geomagnetic reversals from detailed volcanic records, *C.R. Geoscience*, **335**, 79–90, 2003.

Valet, J.-P., and L. Meynadier, Geomagnetic field intensity and reversals during the past four million years, *Nature*, **366**, 234–238, 1993.

Valet, J.-P., J. Brassart, X. Quidelleur, V. Soler, P.-Y. Gillot, and L. Hongre, Paleointensity variations across the last geomagnetic reversal at La Palma, Canary Islands, Spain, *J. Geophys. Res.*, **104**, 7577–7598, 1999.

Waleffe, F., Transition in shear flows. Nonlinear normality versus non-normal linearity, *Phys. Fluids A*, **7**, 3060–3066, 1995.

Weir, A., Axisymmetric convection in a rotating sphere. Part 1. Stress-free surface., *J. Fluid Mech.*, **75**, 49–79, 1976.

Zhang, K., Spiralling columnar convection in rapidly rotating spherical fluid shells, *J. Fluid Mech.*, **236**, 535–556, 1992.

Zhang, K., and F. Busse, Finite amplitude convection and magnetic field generation in a rotating spherical shell, *Geophys. Astrophys. Fluid Dyn.*, **44**, 33–53, 1988.

Appendix A

An inverse model of the Earth's magnetic field from satellite data

A.1 Introduction

This appendix details the inversion method used to map satellite data from the Ørsted mission (1999-2000) to the field on the core-mantle boundary (CMB), being the edge of the dynamo region, for a specific date: January 1st 2000 (epoch 2000). The procedure followed is similar to that of many authors (eg. *Shure et al.*, 1982; *Bloxham and Jackson*, 1992; *Olsen et al.*, 2000; *Olsen*, 2002). The data set consists of measurements of the three components of the magnetic field (B_x, B_y, B_z) (relative to the satellite's orientation), its intensity $F = \sqrt{(}B_x^2 + B_y^2 + B_z^2)$ along with other information such as the local time, position and attitude. The satellite's magnetic instruments were stationed at the end of an 8m long boom, in order to minimise noise from the main body. They comprise a flux-gate magnetometer (to measure the field vector), an Overhauser magnetometer to record the intensity (although mainly used to calibrate the flux-gate magnetometer) and a star-imager. In general three angles are needed to fix the satellite's orientation: two to describe the direction of the fixed axis of the imager and one of rotation about that axis. A slight complication comes about from the errors determining these angles: it is more difficult to establish the rotational attitude than the directional orientation of the star-imager.

A.2 The magnetic field model

In what follows we model the mantle and lower atmosphere (at least up to satellite altitude) as electrical insulators; the CMB (radius $c = 3485$km) and the surface of the Earth (radius

6371.2km) are assumed to be perfectly spherical. From Maxwell's equations, taking the current $\mathbf{J} = \mathbf{0}$ gives

$$\boldsymbol{\nabla} \times \mathbf{B} = \mathbf{0} \tag{A.1}$$

The field then admits the potential form

$$\mathbf{B} = -\boldsymbol{\nabla} V \tag{A.2}$$

with $\nabla^2 V = 0$ from the solenoidal condition on $\mathbf{B}$. Working in spherical polar coordinates $(r, \theta\, \phi)$ we can separate the variables to find, in the region $r \geq c$, that part of the potential associated with the internal field:

$$V = a \sum_{l=1}^{L} \sum_{m=0}^{l} \left(\frac{a}{r}\right)^{l+1} \left[\left(g_l^m \cos(m\phi) + h_l^m \sin(m\phi)\right] P_l^m(\cos\theta)\right. \tag{A.3}$$

where $P_l^m(\cos\theta)$ are the associated Legendre polynomials of degree l and order m (see *Macrobert*, 1967, for example or Appendix B.1); we use standard Schmidt quasi-normalisation. In general equation (A.3) contains a sum involving contributions from the external field, however we have excluded it here since we do not solve for it: instead we use an existing model and subtract the signal from the data. Given the Gauss coefficients g_l^m and h_l^m it is straightforward to compute $\mathbf{B}$ from the internal potential:

$$\begin{aligned}
B_r &= \sum_{l=1}^{L} \sum_{m=0}^{l} (l+1) \left(\frac{a}{r}\right)^{l+2} \left\{g_l^m \cos(m\phi) + h_l^m \sin(m\phi)\right\} P_l^m(\cos\theta) \\
B_\theta &= -\sum_{l=1}^{L} \sum_{m=0}^{l} \left(\frac{a}{r}\right)^{l+2} \left\{g_l^m \cos(m\phi) + h_l^m \sin(m\phi)\right\} \frac{dP_l^m(\cos\theta)}{d\theta} \\
B_\phi &= \frac{1}{\sin\theta} \sum_{l=1}^{L} \sum_{m=0}^{l} m \left(\frac{a}{r}\right)^{l+2} \left\{g_l^m \sin(m\phi) - h_l^m \cos(m\phi)\right\} P_l^m(\cos\theta)
\end{aligned} \tag{A.4}$$

Note that in general the summations extend over all values of l, so there will be an infinity of unknown coefficients. We truncate the expansion of the internal field at $L = 19$ in this study, excluding the monopolar $l = 0$ term. Although the coefficients are computed up to degree 19, harmonics of degree greater than 14 are thought to contain significant crustal signals which are not accounted for in this model (see *Langel*, 1987).

A.3 Anisotropic errors

The errors in the attitude measurements determined by the star-imager are anisotropic: it transpires that determining the direction axis of the star-imager is more accurate than the measurement of the rotation about that axis. Values of these errors are $60''$ (60 arc seconds) for the rotation and $10''$ for the direction angles (*Olsen et al.*, 2000). The nature of these non isotropic errors mean that the covariance matrix of the field data in the Earth's frame of reference is not diagonal: errors in the different field components are correlated. To proceed we must first rotate the field readings into a frame in which the errors are independent. Such a geometry is described by the eigenvectors of the covariance matrix, a full treatment of which can be found in *Holme and Bloxham* (1996). In addition we augment the error budget by an isotropic error of 2.25nT to account for instrumentation inaccuracies and unmodelled crustal and transient external magnetic fields.

A.4 Reduction of measured data

In addition to the signal of the internal field, the Ørsted data contains contributions from the external field and noise. Because we are trying to fit a model for one particular time (January 1st 2000), we must also take account of the fact that the internal field changes slightly over time (the secular variation). In this respect we use the same model used in *Olsen et al.* (2000) to subtract the secular variation from the signal.

A.4.1 Disturbance indexes

The Earth's external magnetic field is associated with the magnetospheric equatorial 'ring' current which is fed by charged particles from the sun. During magnetic storms this field is much enhanced and its effect can felt on the Earth's surface. This effect is quantified by the *Disturbance storm time* or Dst index, which measures the excursion of the horizontal magnetic field compared with quiet times at fixed observatories.

Another index frequently used is Kp, which is calculated from ground based observations and measures a global average of magnetic activity from external sources every three hours. These are then standardised to produce a scale of increasing excitation:

$$0o, 0+, 1-, 1o, 1+, 2-, 2o, 2+, \cdots, 8+, 9-, 9o$$

For more details on these effects see *Kallenrode* (2001).

A.4.2 Data Selection

Data from geomagnetic quiet time conditions, between December 18, 1999 and January 21, 2000, were selected according to the following criteria:

- $K_p \leq 1+$ for time of observation and $K_p \leq 2o$ for the previous 3 hour interval
- $|Dst| \leq 10$nT and $\frac{d(Dst)}{dt} \leq 3$nT/hr
- Local time about 22:00 (local night time data)
- Vector data of geomagnetic colatitude θ_d such that $40^\circ \leq \theta_d \leq 140^\circ$
- Intensity data for $|90^\circ - \theta_d| \geq 50^\circ$

Local night-time data were used to reduce contributions from ionospheric currents. Geomagnetic colatitude is similar to the usual geographic colatitude except that it is measured from the geomagnetic north pole. Near the geomagnetic poles field aligned currents induce small magnetic fields which really only contaminate only the vector data, leaving the intensities relatively unperturbed. Therefore within 40° of the geomagnetic poles we use only intensity measurements.

In total, 13869 data were used, comprising 3907 locations sampling the three magnetic field components, and 2148 locations giving intensity data.

A.4.3 Removal of the external field

Using the model of *Olsen et al.* (2000), we write the external field potential in the following way:

$$V_{ext} = a \left\{ \sum_{l=1}^{2} \sum_{m=0}^{l} \left(\frac{r}{a}\right)^l \left[q_l^m \cos(m\phi) + s_l^m \sin(m\phi)\right] P_l^m(\cos\theta) + Dst \left[\frac{r}{a} + Q_1 \left(\frac{a}{r}\right)^2\right] \cdot \left[\tilde{q}_1^0 P_1^0(\cos\theta) + \left(\tilde{q}_1^1 \cos\theta + \tilde{s}_1^1 \sin\phi\right) P_1^1(\cos\theta)\right] \right\} \quad \text{(A.5)}$$

The first sum is the external (approximately dipolar) field due to the quiet-time equatorial ring current; the second, proportional to the Dst index, is the contribution from high solar activity. The factor Q_1 represents the magnetic field internal to the Earth induced by the external field; this can occur because the mantle is not a perfect electrical insulator and allows small currents.

The model parameters for the external field are shown in table A.1. To subtract the external field from the measured vector data is a trivial task. However, reducing the intensity (F) data is a non-linear process and requires knowledge of the field vector which is not available

l	m	q_l^m	s_l^m	$\tilde{q}_l^m$	$\tilde{s}_l^m$
1	0	22.43		−0.59	
1	1	0.84	−3.73	0.04	0.10
2	0	1.57			
2	1	0.29	-0.32		
2	2	-0.52	-0.04		

Table A.1: Model parameters for the external field

at such locations. We use the field model of *Olsen et al.* (2000) to compute the unknown field at these locations. Equations (A.6a)-(A.6b) show the reduction; the suffix 'm' denoting measured field, 'red' the reduced field and 'ext' the external field.

$$\mathbf{B}_{\text{red}} = \mathbf{B}_{\text{m}} - \mathbf{B}_{\text{ext}} \tag{A.6a}$$

$$F_{\text{red}}^2 = |\mathbf{B}_{\text{red}}|^2 = F_{\text{m}}^2 - 2\mathbf{B}_{\text{m}} \cdot \mathbf{B}_{\text{ext}} + F_{\text{ext}}^2 \tag{A.6b}$$

Although the secular variation of the field will be small over a matter of days either side of epoch 2000 when the measurements were taken, we use the model of *Olsen et al.* (2000) to remove the effect. This was done in an identical manner to the removal of the external field.

A.5 Fitting the model

We are now in a position to fit the data to the model of the magnetic field. Strong distortion results in downwards propagation of a field model due to the high degree harmonics being multiplied by a factor $\left(\frac{a}{r}\right)^{l+1}$. This means that small errors in the field model fitted at satellite altitude will be amplified on the CMB. Indeed, in general, errors beyond the resolution of the satellite instrumentation could in theory be magnified to such an extent as to dwarf all other signal (if l was large enough). In order to resolve this, we use only harmonics up to degree 19 and facilitate convergence by a damping scheme.

Let us define a *model* of the internal field by the vector

$$\mathbf{m} \;=\; \left\{g_1^0, g_1^1, h_1^1, g_2^0, g_2^1, g_2^2, h_2^1, h_2^2, \ldots, h_{19}^{19}\right\} \tag{A.7}$$

which completely determines $\mathbf{B}$ by relation (A.4). Let $\boldsymbol{\gamma}$ denote a set of observations of the internally generated field and $\tilde{\boldsymbol{\gamma}}$ the corresponding set of predictions, a function of the model parameters $\mathbf{m}$; we seek to minimise the difference. In addition we introduce a measure of complexity of the model which we incorporate into the method. The problem of fitting the data is

one of weighted least squares of the form:

$$\text{Minimise} \qquad \Phi = \underbrace{\left[\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}\right]^T \mathbf{C_e^{-1}} \left[\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}\right]}_{weighted\ \ error} + \lambda \underbrace{\mathbf{m}^T \mathbf{C_m^{-1}} \mathbf{m}}_{complexity} \tag{A.8}$$

The matrices $\mathbf{C_e^{-1}}$ and $\mathbf{C_m^{-1}}$ are those of the covariance and complexity norm. The value λ is a weighting, changing the penalisation of the error or the complexity. If large, the resulting field will be 'simple' (one with a low complexity), which will not necessarily fit the data very well. If small, we will obtain a good fit to the data but the field may be very complex. The choice of λ is far from trivial and is discussed presently.

A.6 The measure of field complexity

Since we will be interested in determining the field on the CMB (radius c), we introduce a measure of complexity which reflects the field variation on this surface. Clearly the notion of variation is subjective, in particular it could imply any number of spatial derivatives of any of the field components. Since we will be mainly interested in B_r, we use the definition:

$$C = \oint_{\text{CMB}} \left|\nabla_h B_r\right|^2 d\Omega \tag{A.9}$$

where ∇_h is the horizontal projection of the gradient operator and $d\Omega = \sin\theta\, d\theta\, d\phi$ is the element of solid angle.

The complexity may be written, on substituting the form for **B**:

$$C = \frac{1}{c^2} \sum_{l=1}^{19} \sum_{m=0}^{l} \frac{4\pi}{2l+1} l(l+1)^3 \left(\frac{a}{c}\right)^{2l+4} \left[(g_l^m)^2 + (h_l^m)^2\right] \tag{A.10}$$

This is of the form $\mathbf{m}^T \mathbf{C_m^{-1}} \mathbf{m}$ with

$$\mathbf{C_m^{-1}} = \text{diag}\left(\frac{4\pi}{2l+1} \left(\frac{a}{c}\right)^{2l} l(l+1)^3\right) \tag{A.11}$$

where the other multiplicative constants: $\frac{1}{c^2}\left(\frac{a}{c}\right)^4$ have been absorbed into λ. Notice that the complexity of a harmonic of degree l is $O\left(\frac{a}{c}\right)^{2l}$ which strongly damps spatially complex terms. Other (equally plausible) complexity norms are shown in table A.2.

Norm	Entries on diagonal of $\mathbf{C_m^{-1}}$
$\int B_r^2 d\Omega$	$\frac{4\pi}{2l+1}\left(\frac{a}{c}\right)^{2l}(l+1)^2$
$\int \lvert\nabla_h^2 B_r\rvert^2 d\Omega$	$\frac{4\pi}{2l+1}\left(\frac{a}{c}\right)^{2l} l^2(l+1)^4$

Table A.2: Other equally plausible damping norms and their associated matrix entries

A.7 The Covariance Matrix

The form of equation (A.8) may be directly compared to that of multivariate normal analysis. The matrix $\mathbf{C}_e^{-1}$ is the covariance matrix, measuring the interaction of errors within the set of observables. If we deal only with those which are independent (rotating the vector measurements where necessary) then $\mathbf{C}_e$, and therefore $\mathbf{C}_e^{-1}$, takes on a diagonal form:

$$\mathbf{C}_e^{-1} = \text{diag}\left(\frac{1}{\sigma_{x_1}^2}, \frac{1}{\sigma_{x_2}^2}, \cdots, \frac{1}{\sigma_{x_N}^2}\right) \tag{A.12}$$

The constant $\sigma_{x_i}^2$ is the variance of observable X_i (either a field vector component or the intensity) and is assumed to be a known. In the case of Ørsted the variances come from an *a posteriori* estimate (see *Olsen et al.*, 2000).

A.8 Choosing the value of λ

A measure of how well the data fit the model is given by the misfit χ, defined by

$$\chi^2 = \frac{1}{N}\left[\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}\right]^T \mathbf{C_e^{-1}}\left[\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}\right] \tag{A.13}$$

where N denotes the number of data. On varying the value of λ we derive a 'trade-off' curve, see figure A.1. Plotted is the misfit against the field complexity parameterised by λ. Low values of λ mean that the field is of high complexity but fits the data well (thus the misfit is low). On the other hand, high values penalise spatially complex fields and so the resulting field is simple but has a higher misfit. Any of the models represented in figure A.1 are possible candidates so we need some way of choosing between them. Multivariate statistics lead us to believe that the misfit should be distributed about unity with a variance which decreases with N. Thus we choose the model derived from the value of λ which gives a misfit of unity. In this case, the location coincides with the 'knee' of the curve which separates the two minimisation regimes which also can be used as method for determining λ.

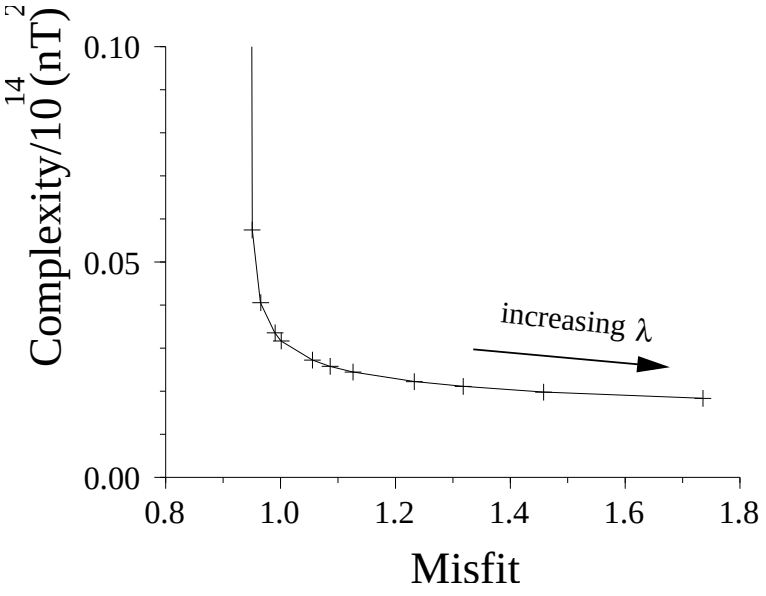


Figure A.1: The trade-off curve for the epoch 2000 model: complexity against misfit.

A.9 Finding the minimising model

Setting the derivative of (A.8) with respect to m_i, the i^{th} model parameter to zero, and denoting $\tilde{\boldsymbol{\gamma}} = \mathbf{f}(\mathbf{m})$ gives

$$0 = -\frac{\partial \mathbf{f}^T}{\partial m_i}\mathbf{C}_e^{-1}\left[\boldsymbol{\gamma} - \mathbf{f}(\mathbf{m})\right] - \left[\boldsymbol{\gamma} - \mathbf{f}(\mathbf{m})\right]^T \mathbf{C}_e^{-1}\frac{\partial \mathbf{f}}{\partial m_i} + \lambda\left(\mathbf{m}^T\mathbf{C}_m^{-1}\right)_i + \lambda\left(\mathbf{C}_m^{-1}\mathbf{m}\right)_i \qquad \text{(A.14)}$$

Since all the above are scalars hence equal to their transpose we have in vector form:

$$\lambda\mathbf{C}_m^{-1}\mathbf{m} = \mathbf{A}^T\mathbf{C}_e^{-1}\left[\boldsymbol{\gamma} - \mathbf{f}(\mathbf{m})\right] \qquad \text{(A.15)}$$

where $\mathbf{A}_{ij} = \frac{\partial f_i}{\partial m_j}$ are the Frechet derivatives. Note that since the measurements are rotated so that the covariance matrix is diagonal, the same treatment must be given to the Frechet derivatives. In the case where the data are linearly related to the model coefficients (this is true for the vector but not the intensity data), $\mathbf{f}(\mathbf{m}) = \mathbf{A}\,\mathbf{m}$ so

$$\mathbf{m} = \left(\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A} + \lambda\mathbf{C}_m^{-1}\right)^{-1}\mathbf{A}^T\mathbf{C}_e^{-1}\boldsymbol{\gamma} \qquad \text{(A.16)}$$

If the data set contains some observables which are nonlinear in the model **m**, then we are forced to adopt an iterative process. Let $\hat{\mathbf{m}}$ be the minimising model and **m** be the current

approximation, so that $\hat{\mathbf{m}} = \mathbf{m} + \delta\mathbf{m}$ and so, correct to first order in $\delta\mathbf{m}$:

$$\mathbf{f}(\mathbf{m} + \delta\mathbf{m}) = \mathbf{f}(\mathbf{m}) + \mathbf{A}\delta\mathbf{m} \tag{A.17}$$

Rewriting equation (A.15) as an iterative procedure:

$$\lambda\mathbf{C}_m^{-1}(\mathbf{m} + \delta\mathbf{m}) = \mathbf{A}^T\mathbf{C}_e^{-1}\left[\boldsymbol{\gamma} - \mathbf{f}(\mathbf{m} + \delta\mathbf{m})\right] \tag{A.18}$$

we obtain

$$\left[\lambda\mathbf{C}_m^{-1} + \mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A}\right]\delta\mathbf{m} = \mathbf{A}^T\mathbf{C}_e^{-1}\left[\boldsymbol{\gamma} - \mathbf{f}(\mathbf{m})\right] - \lambda\mathbf{C}_\mathbf{m}^{-1}\mathbf{m} \tag{A.19}$$

This leads to the iterative procedure:

$$\mathbf{m}_{i+1} = \mathbf{m}_i + \left[\lambda\mathbf{C}_m^{-1} + \underbrace{\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A}}_{\mathbf{e}}\right]^{-1}\left[\underbrace{\mathbf{A}^T\mathbf{C}_e^{-1}\left(\boldsymbol{\gamma} - \mathbf{f}(\mathbf{m}_i)\right)}_{\mathbf{c}} - \lambda\mathbf{C}_m^{-1}\mathbf{m}_i\right]$$

where $\mathbf{A}$ will depend on the model $\mathbf{m}$ in general. The matrix $\mathbf{e}$ above is known as the *normal equations matrix* and $\mathbf{c}$ is the *right hand side vector*. These two terms take most of the computation time to calculate; adding in the damping terms and inverting to find the next model approximation is relatively quick. We solve the system of linear equations using the LU decomposition (see *Press et al.*, 1992).

A.10 The model for epoch 2000

Using the above methodology, the following model was produced with a misfit of unity. Figure A.2 shows the radial component of the field at the CMB. This is the only field component guaranteed to be continuous across the CMB interface as current sheets would introduce discontinuities in the tangential components (see *Gubbins and Roberts*, 1987). Orange is positive and blue negative; continents are shown for reference. The field is principally an axial dipole; field lines mainly leave the core in the southern hemisphere and return in the northern hemisphere. Reversed flux patches complicate the picture: two such strong areas are located beneath the south Atlantic. Patches of equatorially symmetric strong flux intensities have been linked to the presence of axially aligned flow structures (*Gubbins and Bloxham*, 1987) although the actual field morphology is far more complex. More recently, *Jackson* (2003) has speculated that the intense flux patches in equatorial regions could be the manifestation of dynamo waves in the core.

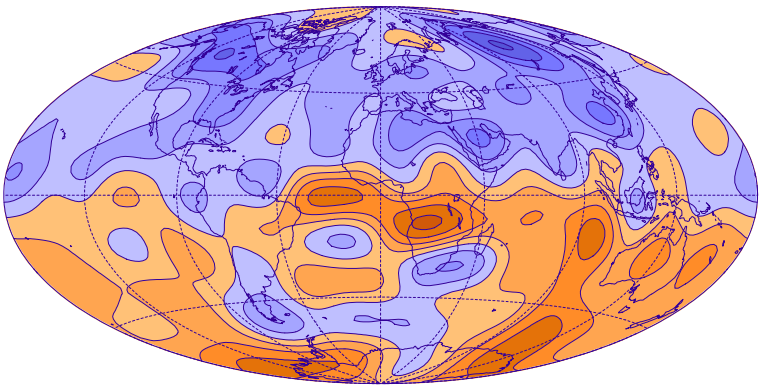


Figure A.2: The radial component of the Earth's magnetic field at epoch 2000. Contour interval is 200 μT. Orange is positive, blue is negative.

Table A.3 shows the model coefficients, which indicate the strength of the corresponding field harmonic at the Earth's surface. In particular, g_1^0 dominates the model by about an order of magnitude, as would be expected, since the observed field is dominated by an axially symmetric dipole. The other $l = 1$ harmonics are equatorially aligned dipoles of longitudinal orientations $\phi = \pi/2$ and $\phi = 0$ associated with the h_1^1 and g_1^1 coefficients respectively.

l	m	g_l^m(nT)	h_m^l(nT)	l	m	g_l^m(nT)	h_m^l(nT)
1	0	-29616.31		11	0	2.42	
1	1	-1729.10	5185.72	11	1	-1.61	0.41
2	0	-2268.36		11	2	-1.69	1.15
2	1	3068.88	-2481.78	11	3	1.35	-0.87
2	2	1670.70	-457.54	11	4	-0.14	-2.50
3	0	1340.00		11	5	0.23	0.90
3	1	-2288.37	-227.94	11	6	-0.76	-0.62
3	2	1250.03	293.14	11	7	0.80	-2.61
3	3	714.03	-491.36	11	8	1.63	-0.89
4	0	932.15		11	9	0.08	-0.98
4	1	786.60	273.32	11	10	0.93	-1.77
4	2	249.82	-231.60	11	11	3.86	-0.41
4	3	-403.40	119.48	12	0	-2.02	
4	4	111.24	-303.64	12	1	-0.38	-0.40
5	0	-216.95		12	2	0.29	0.28
5	1	352.07	42.73	12	3	0.77	2.03
5	2	222.09	171.26	12	4	-0.21	-2.30
5	3	-130.46	-132.83	12	5	0.74	0.49
5	4	-168.41	-39.38	12	6	-0.42	0.24
5	5	-12.82	106.45	12	7	0.21	0.07
6	0	71.38		12	8	-0.14	-0.03
6	1	67.39	-16.91	12	9	-0.48	0.25
6	2	74.21	64.26	12	10	-0.23	-0.91
6	3	-160.82	65.35	12	11	-0.12	-0.35
6	4	-5.64	-61.00	12	12	0.15	0.40
6	5	16.89	0.86	13	0	-0.20	
6	6	-90.30	43.81	13	1	-0.48	-0.44
7	0	78.98		13	2	0.21	0.11
7	1	-73.70	64.93	13	3	0.05	0.97
7	2	-0.05	-24.78	13	4	-0.27	-0.25
7	3	32.98	6.16	13	5	0.72	-0.47
7	4	9.12	23.99	13	6	-0.23	-0.03
7	5	6.99	14.86	13	7	0.36	0.31
7	6	7.07	-25.31	13	8	-0.16	0.09
7	7	-1.34	-5.72	13	9	0.20	0.21
8	0	24.08		13	10	-0.04	0.11
8	1	6.05	12.24	13	11	0.12	-0.11
8	2	-9.12	-21.01	13	12	0.18	-0.20
8	3	-7.68	8.58	13	13	-0.15	-0.38
8	4	-16.60	-21.38	14	0	-0.07	
8	5	8.99	15.29	14	1	0.06	0.08
8	6	6.96	8.77	14	2	-0.04	-0.16
8	7	-7.87	-14.86	14	3	-0.17	0.04
8	8	-7.10	-2.42	14	4	-0.05	0.05
9	0	5.13		14	5	0.03	-0.03
9	1	9.63	-19.92	14	6	-0.02	0.08
9	2	2.87	12.99	14	7	0.01	0.03
9	3	-8.51	12.50	14	8	0.03	0.04
9	4	6.29	-6.22	14	9	0.02	0.05
9	5	-8.70	-8.24	14	10	0.09	-0.01
9	6	-1.46	8.36	14	11	-0.06	-0.01
9	7	9.13	3.91	14	12	0.01	0.05
9	8	-4.19	-8.22	14	13	0.04	-0.01
9	9	-8.17	4.86	14	14	0.06	0.01
10	0	-2.94					
10	1	-6.43	1.89				
10	2	1.64	0.26				
10	3	-2.92	4.14				
10	4	-0.29	4.90				
10	5	3.63	-5.86				
10	6	1.10	-1.13				
10	7	2.00	-2.86				
10	8	4.41	0.24				
10	9	0.42	-2.02				
10	10	-0.90	-7.51				

Table A.3: Model parameters for epoch 2000, up to degree 14.

Appendix B

Spherical harmonics

B.1 Derivation

B.1.1 Laplace's equation

Spherical harmonics arise from solutions of Laplace's equation $\nabla^2 V = 0$. In spherical polar coordinates we need to solve:

$$\frac{1}{r}\frac{\partial^2(rV)}{\partial r^2} + \frac{1}{r^2}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2}\frac{1}{\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2} = 0 \tag{B.1}$$

If we look for separated solutions $V(r,\theta,\phi) = R(r)\Theta(\theta)\Phi(\phi)$ and find that:

$$\begin{aligned} R(r) &= r^l \text{ or } r^{-(l+1)} \\ \Theta(\theta) &= P_l^m(\cos\theta) \\ \Phi(\phi) &= \sin(m\phi) \text{ or } \cos(m\phi) \end{aligned} \tag{B.2}$$

where l and m are integers with $m \leq l$. $P_l^m(z)$ is an associated Legendre function, being a solution of

$$(1-z^2)\frac{d^2P_l^m}{dz^2} - 2z\frac{dP_l^m}{dz} + \left[l(l+1) - \frac{m^2}{1-z^2}\right]P_l^m = 0 \tag{B.3}$$

l is called the degree and m the order.

B.1.2 Associated Legendre functions

The Legendre polynomials satisfy equation (B.3) but with $m = 0$. They can be written in closed form as

$$P_n(z) = \frac{1}{2^n n!} \frac{d^n (z^2 - 1)^n}{dz^n} \tag{B.4}$$

(e.g. *Abramowitz and Stegun*, 1984). The associated functions are closely related by

$$P_l^m(z) = (1 - z^2)^{\frac{m}{2}} \frac{d^m P_l(z)}{dz^m} \tag{B.5}$$

and satisfy the orthogonality relation:

$$\int_0^\pi P_l^m(\cos\theta) P_k^m(\cos\theta) \sin\theta d\theta = \begin{cases} \frac{2}{2l+1} \frac{(l+m)!}{(l-m)!} & l = k \\ 0 & l \neq k \end{cases} \tag{B.6}$$

B.1.3 Properties

For clarity we rewrite equation (B.5) for an associated Legendre function as

$$P_l^m(\cos\theta) = \sin^m\theta \frac{d^m P_l(z)}{dz^m} \tag{B.7}$$

Since $P_l(-z) = (-1)^l P_l(z)$ using equation (B.4), the following properties hold:

$$P_l^m(\theta + \pi) = (-1)^l P_l^m(\theta) \tag{B.8}$$

$$P_l^m(\pi - \theta) = (-1)^{(l+m)} P_l^m(\theta) \tag{B.9}$$

B.2 Spherical harmonics

We define a spherical harmonic of degree l and order m to be of the form

$$Y_l^m(\theta, \phi) = A_l^m P_l^m(\cos\theta) \begin{Bmatrix} \sin(m\phi) \\ \cos(m\phi) \end{Bmatrix} \tag{B.10}$$

where A_l^m are constants, yet to be determined. Integrating over a spherical shell, we see that they are orthogonal, firstly in l (by the associated Legendre functions) and in m by Fourier theory.

To ease notation, we shall sometimes use Greek subscripts, for example Y_α, to denote a particular spherical harmonic.

B.2.1 Normalisation

One of the issues that often leads to confusion is that of normalisation. As seems to be standard in geomagnetism, we use Schmidt quasi-normalised harmonics, defining the remaining multiplicative constants as

$$A_l^m = \begin{cases} 1 & m = 0 \\ \left[2\frac{(l-m)!}{(l+m)!}\right]^{\frac{1}{2}} & m > 0 \end{cases} \tag{B.11}$$

then

$$\int_0^{\pi}\int_0^{2\pi} Y_\alpha(\theta,\phi)\,Y_\beta(\theta,\phi)\sin\theta\,d\phi\,d\theta = \begin{cases} 0 & \alpha \neq \beta \\ \frac{4\pi}{2l_\alpha+1} & \alpha = \beta \end{cases} \tag{B.12}$$

Table B.1 shows associated Legendre functions with this choice of normalisation up to degree 3.

l	m	$A_l^m P_l^m(\cos\theta)$
1	0	$\cos\theta$
1	1	$\sin\theta$
2	0	$\frac{1}{2}(3\cos^2\theta - 1)$
2	1	$\sqrt{3}\cos\theta\sin\theta$
2	2	$\frac{\sqrt{3}}{2}\sin^2\theta$
3	0	$\frac{1}{2}(5\cos^3\theta - 3\cos\theta)$
3	1	$\frac{\sqrt{3}}{\sqrt{8}}\sin\theta(5\cos^2\theta - 1)$
3	2	$\frac{\sqrt{15}}{2}\sin^2\theta\cos\theta$
3	3	$\frac{\sqrt{5}}{\sqrt{8}}\sin^3\theta$

Table B.1: The associated Legendre functions with Schmidt quasi-normalisation up to degree 3.

B.3 Integral identities

We now derive a few useful identities, making use of the surface form of Gauss's divergence theorem (see *Backus et al.*, 1996, p.344). This uses the surface gradient operator $\boldsymbol{\nabla}_1 = (0, \frac{\partial}{\partial\theta}, \frac{1}{\sin\theta}\frac{\partial}{\partial\phi})$ and states

$$\iint_S \boldsymbol{\nabla}_1 \cdot \mathbf{v}\,d\Omega = \oint_{\partial S} \mathbf{v}\cdot\mathbf{dl} \tag{B.13}$$

where S is an origin-concentric spherical surface with boundary ∂S, $d\Omega$ and $\mathbf{dl}$ are respectively elements of solid angle and of the path length and $\mathbf{v}$ is some vector.

B.3.1 Identity 1

$$\int_0^{2\pi}\int_0^{\pi}\left(\frac{\partial Y_\alpha}{\partial\theta}\frac{\partial Y_\beta}{\partial\phi}-\frac{\partial Y_\alpha}{\partial\phi}\frac{\partial Y_\beta}{\partial\theta}\right)d\theta d\phi=0 \tag{B.14}$$

The proof follows from writing it in the form

$$\oint_S \boldsymbol{\nabla}_1(Y_\alpha)\cdot\boldsymbol{\nabla}_1\times\left[Y_\beta\hat{\mathbf{r}}\right]d\Omega \tag{B.15}$$

This can be expressed as

$$\oint_S \boldsymbol{\nabla}_1\cdot\left(\boldsymbol{\nabla}_1(Y_\alpha)\times(Y_\beta)\hat{\mathbf{r}}\right)dS \tag{B.16}$$

which vanishes since S has a null boundary curve (which has zero length).

B.3.2 Identity 2

$$\int_0^{2\pi}\int_0^{\pi}\left(\frac{\partial Y_\alpha}{\partial\theta}\frac{\partial Y_\beta}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial Y_\alpha}{\partial\phi}\frac{\partial Y_\beta}{\partial\phi}\right)\sin\theta d\theta d\phi=\frac{4\pi}{2l_\alpha+1}\delta_{\alpha\beta}l_\alpha(l_\alpha+1) \tag{B.17}$$

The left hand side of the identity above may be written

$$\oint_S \boldsymbol{\nabla}_1 Y_\alpha\cdot\boldsymbol{\nabla}_1 Y_\beta\,d\Omega=\int\left[\boldsymbol{\nabla}_1\cdot(Y_\alpha\boldsymbol{\nabla}_1 Y_\beta)-Y_\alpha\nabla_1^2 Y_\beta\right]d\Omega \tag{B.18}$$

and reduces to

$$l_\beta(l_\beta+1)\oint_S Y_\alpha Y_\beta d\Omega \tag{B.19}$$

since the surface S has a null boundary curve and Y_l^m is an eigenvector of $\nabla_1^2=-L^2$ (where L^2 is the angular momentum operator of section 3.4.1) with eigenvalue $-l(l+1)$. The result follows by our choice of normalisation.

B.3.3 Identity 3

$$\int_1^{\infty}\int_0^{2\pi}\int_0^{\pi}\left|\boldsymbol{\nabla}\left(\frac{Y_l^m}{r^{l+1}}\right)\right|^2 r^2\sin\theta d\theta d\phi dr=(l+1)\frac{4\pi}{2l+1} \tag{B.20}$$

To show this, first consider

$$\boldsymbol{\nabla}\left(\frac{Y_l^m}{r^{l+1}}\right)=r^{-(l+2)}\left(-(l+1)Y_l^m,\frac{\partial Y_l^m}{\partial\theta},\frac{1}{\sin\theta}\frac{\partial Y_l^m}{\partial\phi}\right) \tag{B.21}$$

Taking the modulus and using identity 2 above, we have

$$\begin{aligned}&\left[(l+1)^2+l(l+1)\right]\int_1^\infty r^{-(2l+2)}dr\int[Y_l^m]^2\,d\Omega\\&=\frac{(2l+1)(l+1)}{(2l+1)}\int[Y_l^m]^2\,d\Omega=(l+1)\int[Y_l^m]^2\,d\Omega\\&=(l+1)\frac{4\pi}{2l+1}\end{aligned}\tag{B.22}$$

B.4 Orthogonality of the decay modes

That the decay modes are orthogonal over all space is already known since the diffusion operator is self-adjoint (section 3.8.2) and as such has orthogonal eigenvectors (appealing to standard theory). However, we present here a self-contained derivation of this property.

The decay modes are derived in section 3.4.2 and are all everywhere of time dependence $e^{-d_\alpha^2 t}$ where $-d_\alpha^2$ is the decay rate of mode $\mathbf{B}_\alpha$, and additionally in V:

$$\frac{\partial\mathbf{B}_\alpha}{\partial t}=\nabla^2\mathbf{B}_\alpha\tag{B.23}$$

These may be written explicitly in V for the poloidal and toroidal cases respectively:

$$\mathbf{S}_\alpha=\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\left[r\,j_l(d_P r)\,\hat{\mathbf{r}}\right]\;e^{-d_P^2 t}\tag{B.24}$$

$$\mathbf{T}_\alpha=\boldsymbol{\nabla}\times\left[r\,j_l(d_T r)\,\hat{\mathbf{r}}\right]\;e^{-d_T^2 t}\tag{B.25}$$

where d_T and d_P are the roots of some related spherical Bessel function. In $\hat{V}$ the toroidal modes vanish and the poloidal modes are represented there by a potential.

Consider two decay modes, $\mathbf{B}_\alpha$ and $\mathbf{B}_\beta$. We prove their orthogonality over all space using a similar method to *Backus* (1958). Note that since the toroidal modes vanish in $\hat{V}$ they are orthogonal over V.

$$-d_\alpha^2\int_{V+\hat{V}}\mathbf{B}_\alpha\cdot\mathbf{B}_\beta\,dV=-\int_V\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{B}_\alpha\cdot\mathbf{B}_\beta\,dV+\int_{\hat{V}}\frac{\partial\mathbf{B}_\alpha}{\partial t}\cdot\mathbf{B}_\beta\,dV\tag{B.26}$$

$$=\int_V\boldsymbol{\nabla}\cdot\Big[\mathbf{B}_\beta\times(\boldsymbol{\nabla}\times\mathbf{B}_\alpha)\Big]-\int_V(\boldsymbol{\nabla}\times\mathbf{B}_\beta)\cdot(\boldsymbol{\nabla}\times\mathbf{B}_\alpha)\,dV-\int_{\hat{V}}\boldsymbol{\nabla}\times\mathbf{E}_\alpha\cdot\mathbf{B}_\beta\,dV\tag{B.27}$$

where we have used the pre-Maxwell equation $\boldsymbol{\nabla}\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}$ and the fact the $\mathbf{B}$ is divergence-free.

It is shown below that the first and third terms of equation (B.27) cancel exactly, thus leaving a symmetric integral in the two decay modes. We may therefore reverse the above

argument to show that

$$-d_\alpha^2 \int_{V+\hat{V}} \mathbf{B}_\alpha \cdot \mathbf{B}_\beta \, dV = -d_\beta^2 \int_{V+\hat{V}} \mathbf{B}_\alpha \cdot \mathbf{B}_\beta \, dV \tag{B.28}$$

so unless $\alpha = \beta$ then the modes must be the orthogonal in the sense defined, i.e.

$$\int_{V+\hat{V}} \mathbf{B}_\alpha \cdot \mathbf{B}_\beta \, dV = 0 \tag{B.29}$$

Some of the poloidal and toroidal decay rates coincide, for example, d_P corresponding to a harmonic of degree l is the same as d_T corresponding a harmonic of degree $l-1$. However, in this case, these modes are trivially orthogonal. There is no degeneracy in either of the poloidal or toroidal decay rates (regarded separately).

It remains therefore to show that the required terms do in fact exactly sum to zero. Writing them as

$$\int_{\partial V} \mathbf{B}_\beta \times (\boldsymbol{\nabla} \times \mathbf{B}_\alpha) \cdot \mathbf{dS} - \int_{\hat{V}} \boldsymbol{\nabla} \cdot \left(\mathbf{E}_\alpha \times \mathbf{B}_\beta \right) dV \tag{B.30}$$

since $\boldsymbol{\nabla} \times \mathbf{B}_\beta = \mathbf{0}$ in $\hat{V}$. We invoke continuity of $\mathbf{B}$ and of the tangential components of $\eta \boldsymbol{\nabla} \times \mathbf{B} = \mathbf{E}$ on the non-slip boundary (by permuting the triple product $\mathbf{B} \times \mathbf{E} \cdot \mathbf{dS} = \mathbf{B} \cdot \mathbf{E} \times \mathbf{dS}$) so that

$$\int_{\partial V} \mathbf{B}_\beta \times \mathbf{E}_\alpha \cdot \mathbf{dS} - \int_{\partial \hat{V}} \left(\mathbf{E}_\alpha \times \mathbf{B}_\beta \right) \cdot \mathbf{dS} = 0 \tag{B.31}$$

since $\mathbf{dS}$ points in opposite directions on either side of $r = 1$.

Appendix C

An analytic derivation of Proctor's bound using a variational method

C.1 The equations

In section 3.9.6 we derived an equation describing the solution of the minimisation problem:

$$\lambda_P = \min \frac{\int_{V+\hat{V}} \eta |\boldsymbol{\nabla} \times \mathbf{B}|^2 \, dV}{\int_V \mathbf{B}^2 \, dV} \tag{C.1}$$

which was:

$$\int_{V+\hat{V}} \delta\mathbf{B} \cdot \left\{ \chi_V(\mathbf{r}) \left[\nabla^2 \mathbf{B} + \lambda_P \mathbf{B} \right] - \chi_{\hat{V}}(\mathbf{r}) \left[\boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) \right] \right\} dV = 0 \tag{C.2}$$

where $\chi_V(\mathbf{r}) = 1$ if $\mathbf{r} \in V$ and is zero otherwise. The magnetic diffusivity $\eta = \eta(r)$ is 1 in V and large but finite in $\hat{V}$. At the end of the analysis we will let $\eta \to \infty$ in a well defined manner, consistent with $\hat{V}$ being an electrical insulator. We may approach this limit however we choose: using a purely radial dependence facilitates the analysis. Using the Lemma below equation C.2 becomes:

$$\nabla^2 \mathbf{B} + \lambda_P \mathbf{B} + \boldsymbol{\nabla} \xi = \mathbf{0} \qquad \text{in } V \tag{C.3a}$$

$$-\boldsymbol{\nabla} \times (\eta \boldsymbol{\nabla} \times \mathbf{B}) + \boldsymbol{\nabla} \xi = \mathbf{0} \qquad \text{in } \hat{V} \tag{C.3b}$$

where ξ is an unknown scalar function which is continuous at $r = 1$.

C.1.1 Lemma

The following are equivalent for a vector $\mathbf{M}$ continuous and differentiable in V and $\hat{V}$ but not necessarily on ∂V:

(a)

$$\int_{V+\hat{V}} \mathbf{B}\cdot\mathbf{M}\,dV = 0 \tag{C.4}$$

for all continuous vectors $\mathbf{B}$ such that $\boldsymbol{\nabla}\cdot\mathbf{B} = 0$ and $\mathbf{B} = O(r^{-3})$ at infinity.

(b)

$$\mathbf{M} = \boldsymbol{\nabla}\xi \tag{C.5}$$

for some continuous scalar function ξ (differentiable wherever $\mathbf{M}$ is) which is finite at infinity.

The proof is fairly straight-forward. Suppose first that $\mathbf{M} = \boldsymbol{\nabla}\xi$ with both $\mathbf{M}$ and ξ everywhere continuous and differentiable in V and $\hat{V}$.

$$\begin{aligned}\int_{V+\hat{V}} \mathbf{B}\cdot\boldsymbol{\nabla}\xi\,dV &= \int_{V+\hat{V}} \boldsymbol{\nabla}\cdot(\xi\mathbf{B})\,dV \\ &= \oint_{\partial V} \xi\mathbf{B}\cdot\mathbf{dS} + \oint_{\partial\hat{V}} \xi\mathbf{B}\cdot\mathbf{dS} = 0\end{aligned} \tag{C.6}$$

since $\xi\mathbf{B}\cdot\mathbf{dS} = O(r^{-1})$ at infinity and using continuity of $\mathbf{B}$ and ξ at $r = 1$.

For the converse, since $\boldsymbol{\nabla}\cdot\mathbf{B} = 0$ we can write $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{f}$ for some vector potential $\mathbf{f}$. Then:

$$\int_{V+\hat{V}} (\boldsymbol{\nabla}\times\mathbf{f})\cdot\mathbf{M}\,dV = 0 \quad \text{for all vectors } \mathbf{f} \tag{C.7}$$

and using standard identities yields:

$$\oint_{\partial V+\partial\hat{V}} (\mathbf{f}\times\mathbf{M})\cdot\mathbf{dS} + \int_{V+\hat{V}} \mathbf{f}\cdot(\boldsymbol{\nabla}\times\mathbf{M})\,dV = 0 \tag{C.8}$$

Since the above is true for any vector field $\mathbf{f}$, we may choose it such that it is nonzero only in V or $\hat{V}$. In either case, by using standard Calculus of Variations arguments, we may see that the two volume integrals must individually vanish. This means that $\boldsymbol{\nabla}\times\mathbf{M} = \mathbf{0}$ so $\mathbf{M} = \boldsymbol{\nabla}\xi$. The differentiability and continuity of ξ follows that of $\mathbf{M}$. We now substitute this back in to obtain

$$\oint_{\partial V} \xi\mathbf{B}\cdot\mathbf{dS} + \oint_{\partial\hat{V}} \xi\mathbf{B}\cdot\mathbf{dS} = 0 \tag{C.9}$$

as above, so that ξ is continuous also on ∂V, and hence it is everywhere.

□

C.2 Method of solution

To solve these equations, we expand **B** in the usual poloidal-toroidal form and ξ in spherical harmonics

$$\mathbf{B} = \sum_\beta \mathbf{S}_\beta + \mathbf{T}_\beta = \sum_\beta \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \left[S_\beta(r) Y_\beta \hat{\mathbf{r}} \right] + \sum_\beta \boldsymbol{\nabla} \times \left[T_\beta(r) Y_\beta \hat{\mathbf{r}} \right] \tag{C.10}$$

$$\xi = \sum_\beta \xi_\beta(r) Y_\beta \tag{C.11}$$

At $r = 1$ the following four quantities must be continuous: ξ_β, S_β, $\frac{dS_\beta}{dr}$ and $\eta \nabla_\beta^2 S_\beta$ which follow from the usual continuity conditions of section 3.3.1. Recall the notation:

$$\nabla_\beta^2 = \frac{d^2}{dr^2} - \frac{l_\beta(l_\beta+1)}{r^2} \tag{C.12}$$

$$D_\beta^2 = \frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l_\beta(l_\beta+1)}{r^2} \tag{C.13}$$

where

$$\nabla^2 \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \left[Y_\beta S_\beta \hat{\mathbf{r}} \right] = \boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \left[Y_\beta \nabla_\beta^2 S_\beta \hat{\mathbf{r}} \right] \tag{C.14}$$

$$\nabla^2 Y_\beta S_\beta = Y_\beta D_\beta^2 S_\beta \tag{C.15}$$

In addition we have that

(a) at $r = 0$, S_β and ξ_β are non singular,

(b) as $r \to \infty$, we insist that $\mathbf{B} = O(r^{-3})$ and that $\xi = O(1)$

C.2.1 Solving in V

In V, the β harmonic of the $\hat{\mathbf{r}}$ component of (C.3a) is

$$\frac{l_\beta(l_\beta+1)}{r^2} \nabla_\beta^2 S_\beta + \frac{d\xi_\beta}{dr} + \lambda_P \frac{l_\beta(l_\beta+1)}{r^2} S_\beta = 0 \tag{C.16}$$

On taking the divergence of (C.3a) gives

$$D_\beta^2 \xi_\beta = 0 \tag{C.17}$$

so that ξ_β is of the form r^n where $n = -(l_\beta + 1)$ or l_β (as can be easily calculated).

In V we choose the non-singular solution $\xi_\beta = E r^{l_\beta}$, for some constant E. On substituting this into equation (C.16) we get

$$\nabla_\beta^2 S_\beta + \lambda_P S_\beta = -\frac{E}{l_\beta + 1} r^{(l_\beta+1)} \tag{C.18}$$

Fortunately the particular integral is of the form $F r^{(l_\beta+1)}$ and the complementary function is recognisable from the diffusion problem so that we can write the solution as:

$$S_\beta(r) = \frac{A}{\alpha^2}\left[r\, j_{l_\beta}(\alpha r) + C r^{(l_\beta+1)}\right] \tag{C.19}$$

$$\xi_\beta = -(l_\beta + 1) A C r^{l_\beta} \tag{C.20}$$

where j_l is a spherical Bessel function of order l, and $\lambda_P = \alpha^2 > 0$ [equation C.1 has a strictly positive minimum, see section 3.7.4]. The constant E has been rewritten as $E = -AC(l_\beta + 1)$.

C.2.2 Solving in $\hat{V}$

In $\hat{V}$, taking the divergence of (C.3b) gives $D_{l_\beta}^2 \xi_\beta = 0$ as well. Hence

$$\xi_\beta = G r^{-(l_\beta+1)} \tag{C.21}$$

for some constant G. Since $\eta = \eta(r)$, the β harmonic of the $\hat{\mathbf{r}}$ component of equation (C.3b) is

$$\frac{\eta l_\beta (l_\beta + 1)}{r^2} \nabla_\beta^2 S_\beta + \frac{d\xi_\beta}{dr} = 0 \tag{C.22}$$

On substituting for ξ_β we have

$$\nabla_\beta^2 P_\beta = \frac{G}{\eta l_\beta r^{l_\beta}} \tag{C.23}$$

Writing $S_\beta(r) = r^{l_\beta} Q_\beta(r)$ we find that Q_β satisfies

$$\frac{d}{dr}\left(\frac{dQ_\beta}{dr} + \frac{2l_\beta}{r} Q_\beta\right) = \frac{G}{\eta l_\beta r^{2l_\beta}} \tag{C.24}$$

so that

$$\frac{dQ_\beta}{dr}+\frac{2l_\beta Q_\beta}{r}=-\frac{G}{l_\beta}\int_r^\infty \frac{1}{\eta(s)s^{2l_\beta}}ds \tag{C.25}$$

since $\frac{Q_{l_\beta}}{r}$ and $Q_{l_\beta}^{'}$ vanish at ∞ (because of behaviour of S_β).

C.2.3 Matching at $r=1$

Matching ξ across $r=1$ gives $G=-(l_\beta+1)AC$. Applying continuity at $r=1$ of S_β and $\frac{dS_\beta}{dr}$ or equivalently Q_β and $\frac{dQ_\beta}{dr}$ yields:

$$\frac{A}{\alpha^2}\left(\alpha j_{l_\beta}^{'}(\alpha)-(l_\beta-1)j_{l_\beta}(\alpha)+C\right)+\frac{2Al_\beta}{\alpha^2}\left(j_{l_\beta}(\alpha)+C\right)=\frac{(l_\beta+1)}{l_\beta}ACK_{l_\beta}^2 \tag{C.26}$$

where $K_l^2=\int_1^\infty \frac{ds}{\eta(s)s^{2l}}$, and $'$ denotes the derivative. Either $A=0$ or

$$\alpha j_{l_\beta}^{'}(\alpha)+(l_\beta+1)j_{l_\beta}(\alpha)+(2l_\beta+1)C=\frac{(l_\beta+1)}{l_\beta}\alpha^2CK_{l_\beta}^2 \tag{C.27}$$

Using continuity of $\eta\nabla_\beta^2 S_\beta$ which fixes C in terms of α, we get

$$-\frac{(l_\beta+1)}{l_\beta}AC=-A(j_{l_\beta}(\alpha)+C)+AC \tag{C.28}$$

so that $C=\frac{l_\beta j_{l_\beta}(\alpha)}{(l_\beta+1)}$. Using the standard recurrence relation $j_l^{'}(\alpha)+\frac{l+1}{\alpha}j_l(\alpha)=j_{l-1}(\alpha)$ we find that:

$$j_{l_\beta}(\alpha)\left\{l_\beta\left(2l_\beta+1\right)-(l_\beta+1)K_{l_\beta}^2\right\}=-(l_\beta+1)\alpha j_{l_\beta-1}(\alpha) \tag{C.29}$$

In the insulating limit as $K_l^2\to 0$, we must find minimal positive solutions of

$$j_{l_\beta}(\alpha)+\frac{l_\beta+1}{l_\beta(2l_\beta+1)}\alpha j_{l_\beta-1}(\alpha)=0 \tag{C.30}$$

We now have an equation giving stationary values of (C.1) of which $\lambda_P=\alpha^2$ is the minimum. Proctor stated that the lowest mode of $l=1$ gives the required result which leads to

$$j_1(\alpha)\left[3-2\alpha^2K_1^2\right]=-2\alpha j_0(\alpha) \tag{C.31}$$

reducing to

$$j_1(\alpha)+\frac{2}{3}\alpha j_0(\alpha)=0 \tag{C.32}$$

in the insulating limit. The smallest positive root of this is α, the minimum sought being $\alpha^2 = 12.29$ to 2 decimal places.

It is not obvious *a priori* that $l = 1$ gives the minimal root. Figure C.1 shows the graphs of $F_l(\alpha) = j_l(\alpha) + \frac{l+1}{l(2l+1)}\alpha j_{l-1}(\alpha)$ for various l. The smallest non-zero solution corresponds to the lowest mode $l = 1$.

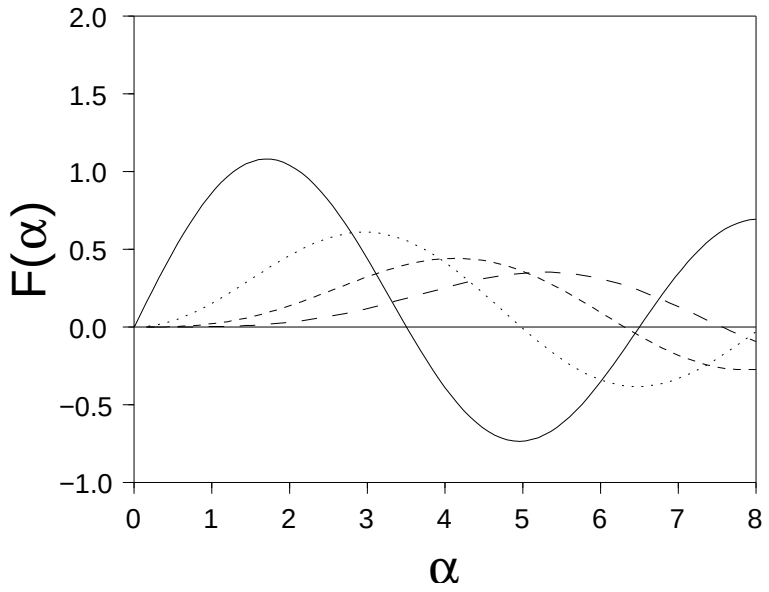


Figure C.1: Graphs of $F_l(\alpha)$ for various l: the solid line is l=1, the dotted is l=2, the dashed is l=3, the long dashed is l=4. The lowest nonzero solution for $F_l(\alpha) = 0$ is $\alpha^2 = 12.29$, $l = 1$.

C.2.4 The conducting mantle

The mantle is not a perfect electrical insulator and this will effect the above analysis slightly. Let us model $\eta(r) = Kr^n$ in $\hat{V}$ where we fit K and n using the estimated non-dimensional values of $\eta(1) = 5000$ and $\eta(1.54) = 50000$ (where $\eta = 1$ in V). These values are taken from the model of *Peyronneau and Poirer* (1989) who calculate values of lower mantle conductivity of 100 Sm^{-1} at the CMB decreasing to $1Sm^{-1}$ at a depth of 1000 km (corresponding to $r = 1.54$). The profile of (non-dimensional) magnetic diffusivity then is $\eta = 5000 \times r^{10.62}$, giving a value of K_1^2 of

$$K_1^2 = \int_1^\infty \frac{ds}{5000\, s^{12.62}} \approx 1.72 \times 10^{-5} \tag{C.33}$$

which in fact leaves the lower bound of 12.29 unchanged (to 2 decimal places). *Buffett* (2002b) suggests a highly conducting layer of 210m thickness at the very base of the mantle, which in non-dimensional terms we model as $\eta = 1$. The corresponding value of K_1^2 also leaves the bound unchanged to 2 decimal places. Thus the inclusion of Earth-like conductivity profiles makes little

difference to Proctor’s bound, in comparison to the electrically insulating mantle analysis.

Appendix D

Computation of Virtual Geomagnetic Poles

We detail below the computation of virtual geomagnetic poles (VGPs), based on *Sarson* (1994), *Merrill et al.* (1998) and *Parkinson* (1983).

Consider three points on the surface of a unit sphere, labelled *A*, *B* and *C*. These will later become the location of the VGP, the geographic north pole and the site location.

It is not obvious how to measure an angle, or indeed even how to construct the triangle *ABC* on a spherical surface; we proceed in the following way. To make the sides of the spherical triangle, we draw an arc which lies on the great circle passing through the two end points. A great circle is defined by the intersection of the spherical surface and a plane passing through 2 points and the origin. How do we measure the length of the arc? Since it is a unit sphere, the arc length is simply the angle subtended at the origin between lines connecting the two end points.

To measure angles i.e. that between 2 arcs, we define it to be the angle between two tangents of the great circles (on which the arcs lie) at that point.

Figure D.1 shows the geometry; each vertex also marks the angle at that point. The length of the associated opposite arc is denoted in small case. We identify two important points:

- The angle *A* is merely the difference between the azimuthal angles of points *B* and *C*. To see this, consider *A* to lie on the *OZ* axis. Points *B* and *C* then have an azimuth defined from *A*.
- If *B* lies on the *OZ* axis, then the arc length *BC* or *BA* is simply the colatitude of points *C* or *A*.

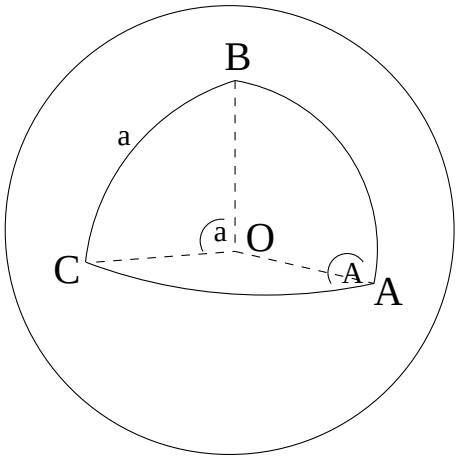


Figure D.1: Sphere layout

Analogues of the cosine and sine formula for planar triangles may be found (e.g. *Parkinson*, 1983) to be

$$\cos c = \cos a \cos b + \sin a \sin b \cos C \tag{D.1}$$

$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c} \tag{D.2}$$

One last result is that of the computation of the geomagnetic colatitude at a point, assuming a dipole field. If the field is dipolar with a magnetic north pole at $\theta = 0$ (in some coordinate system, not necessarily representing the geographic north pole), then

$$\mathbf{B} = \nabla A r^{-2} \cos\theta \tag{D.3}$$

for some constant A. The Z and H (local downwards and horizontal) field components at some geomagnetic colatitude θ_m, at $r = 1$ (the surface of the sphere) are

$$\begin{aligned} Z &= -B_r = 2A\cos\theta_m \\ H &= |B_\theta| = A\sin\theta_m \end{aligned} \tag{D.4}$$

The inclination I is given by $\tan I = Z/H = 2\cot\theta_m$. Suppose then we take measurements of I at some geographic location. This is enough to determine θ_m, the arc length between the sampling site and the VGP.

All the building blocks are now in position to compute the location of the VGP. Suppose then that point B is the geographic north pole ($\theta = 0, \phi = 0$), A is the VGP (θ_p, ϕ_p) and C is the site sampled (θ_s, ϕ_s). We take measurements of I and $D = \tan^{-1}(Y/X)$, the declination.

The spherical cosine rule gives, with $c = \theta_p$ and noting that the angle C is the declination:

$$\cos\theta_p = \cos\theta_s \cos\theta_m + \sin\theta_s \sin\theta_m \cos D \tag{D.5}$$

All of the above quantities are known. To determine ϕ_p, we use the spherical sine formula with angles and sides C and B, noting that angle B is merely the difference in azimuths of locations A and C.

$$\frac{\sin D}{\sin\theta_p} = \frac{\sin(\phi_p - \phi_s)}{\sin\theta_m} \tag{D.6}$$

This leads to

$$\phi_p = \phi_s + \sin^{-1}\left\{\sin\theta_m \sin D (\sin\theta_p)^{-1}\right\} \tag{D.7}$$

At certain locations these formulae become singular, e.g. if $\tan I = 0$ then θ_m is undetermined. However, this represents the geomagnetic equator, so that $\theta_m = \frac{\pi}{2}$.