/N-CAT. 70
72096-CR
77P.

JPL Publication 87-9

Boundary Value Problem for the Solution of Magnetic Cutoff Rigidities and Some Special Applications

Larry Edmonds

(NASA-CR-180916) BOUNDARY VALUE PROBLEM FOR THE SOLUTION OF MAGNETIC CUTOFF RIGIDITIES AND SOME SPECIAL APPLICATIONS (Jet Propulsion Lab.) 77 p Avail: NTIS HC A05/MF A01 CSCL 20C G3/70
N87-22466
Unclas
0072096

April 15, 1987

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL Publication 87-9

Boundary Value Problem for the Solution of Magnetic Cutoff Rigidities and Some Special Applications

Larry Edmonds

April 15, 1987

NASA

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

ABSTRACT

Since a planet's magnetic field can sometimes provide a spacecraft with some protection against cosmic ray and solar flare particles, it is important to be able to quantify this protection. This is done by calculating cutoff rigidities. The conventional method of calculating cutoff rigidities in a nondipole magnetic field is to trace particle trajectories. This publication introduces an alternate method, which is to treat the problem as a boundary value problem. In this approach, trajectory tracing is only needed to supply boundary conditions. In some special cases, trajectory tracing is not needed at all because the problem can be solved analytically. A differential equation governing cutoff rigidities is derived for static magnetic fields. The presence of solid objects, which can block a trajectory, and other force fields are not included. A few qualitative comments, on existence and uniqueness of solutions, have been made which may be helpful when deciding how the boundary conditions should be set up. The differential equation is first expressed in terms of arbitrary coordinates (using the del operator) and then it is expressed in terms of a specific set of coordinates (two sets of spherical coordinates). This publication also includes topics on axially symmetric fields. If the magnetic field is symmetric about an axis, and if a certain kind of cutoff (which this publication calls the "generalized Störmer cutoff") is being investigated, the equations simplify considerably. It is shown that the vertical cutoff, which is the generalized Störmer cutoff evaluated in a direction perpendicular to magnetic east, is constant on a magnetic field line. A method for obtaining analytic solutions for a large class of problems with axial symmetry is derived and two specific examples are included. A procedure for finding the field that provides the best protection subject to a certain kind of constraint is also derived and an example is given. Each application of this optimization analysis will produce an inequality which can be used as an upper bound estimate of the protection that can be provided by a given axially symmetric field.

CONTENTS

PRECEDING PAGE BLANK NOT FILMED

PAGE IV INTENTIONALLY BLANK

1. Introduction

Cutoff rigidities are the quantitative measure of the protection provided to a region of space by a magnetic field against energetic charged particles. This protection is important to a spacecraft in a planet's magnetosphere and it also has certain laboratory applications.

The conventional method of calculating cutoff rigidities in a nondipole magnetic field is to trace particle trajectories by computer. There are some difficulties (described in Ref. 1) with these methods. This publication introduces an alternate method, which is to treat the problem as a boundary value problem. A differential equation, which will be called the field equation, governing cutoff rigidities for a particle in a static magnetic field will be derived. The presence of solid objects, which can block a trajectory, and other force fields are not included. If the magnetic field has no symmetries, it is not immediately obvious that the field equation derived here is more convenient for numerical work than trajectory tracing (especially since boundary conditions have to be supplied and they will probably come from trajectory tracing) but it does at least provide investigators with a choice of methods and they can use their own preferences to make a selection. The boundary value problem approach may also offer new possibilities for investigating analytic properties of cutoff rigidities.

In a given magnetic field, it typically happens that there are a number of different kinds of cutoff rigidities. For a given location and a given particle arrival direction, it is typical to be able to group rigidities into discrete bands such that for any rigidity inside one of these bands the particle cannot escape to infinity and for any rigidity between the bands the particle can escape to infinity (see Ref. 1). In this publication, a "cutoff rigidity" refers to the least upper bound or greatest lower bound of any of these bands. The field equation will be applicable to any cutoff rigidity providing it is well defined everywhere and a differentiable function of its arguments. Being well defined means that we cannot investigate a cutoff rigidity that "disappears," when position and/or direction are varied, due to several distinct bands merging together into one band. In most applications, the cutoff of interest will be either the maximum or the minimum of the various cutoffs and we expect these cutoffs to be well defined.

For a large class of magnetic fields having axial symmetry, it is possible to analytically solve the equation for the "vertical cutoff" (the cutoff in a direction perpendicular to magnetic east). The solution is first derived in the general case and then it is applied to two specific examples. One example is the dipole field. The solution is already well known but this example is useful for demonstrating the method. The second example is a dipole field superimposed on a uniform field in such a way as to create a spherical magnetopause. Axially symmetric, isotropically protective fields will be defined and these fields will be used when solving for the optimum field (the field that produces the greatest magnetic protection) subject to a certain kind of constraint. A by-product of the optimization analysis is an inequality which can be used as an upper bound estimate of the protection that can be provided by a given axially symmetric field. An example will be given.

This publication assumes that the reader is familiar with the basic physical concepts that are discussed in the first few chapters of Rossi and Olbert (Ref. 2). Other than that, this publication is a self-contained mathematical analysis.

2. The General Field Equation

Let $\vec{P}$ denote the rigidity of a particle, which is the momentum divided by the charge, and let $\vec{v}$ be the particle velocity. The force equation in a pure static magnetic field is

$$\frac{d}{dt}(m\vec{v}) = q\vec{v} \times \vec{B}$$

where $\vec{B}$ is the magnetic field in the MKS system of units and q is the charge of the particle. The relativistic mass, designated by m, is a constant of motion since the speed v of the particle is a constant of motion. It is easy to show that the above equation can be expressed in terms of rigidity according to

$$\frac{d\vec{P}}{dS} = \hat{P} \times \vec{B} \tag{1}$$

where $\hat{P}$ is the unit vector in the direction of $\vec{P}$ and S is arclength with the sense of increasing arclength taken to be in the direction of $\vec{P}$.

Consider a point of observation $\vec{X}$ and an arbitrary direction represented by the unit vector $\hat{P}$. Let $F(\vec{X},\hat{P})$ represent the cutoff rigidity at the point $\vec{X}$ and direction $\hat{P}$. Note that there will typically be several cutoff rigidities, associated with upper and lower bounds of various rigidity bands. The one being investigated is represented by F. Now suppose a particle starts at the point $\vec{X}$ and moves in the direction $\hat{P}$ and travels with a rigidity equal to the cutoff F. The particle is in a state of being on the verge of being able to escape to infinity. An infinitesimal change in the rigidity puts the particle in a bound state or in an escape trajectory depending on the direction of the change. Obviously, as the particle moves to a new point on the trajectory, it is still in a state of being on the verge of being able to escape to infinity. Therefore, the rigidity of the particle at this new point is equal to the cutoff rigidity evaluated at that point and in the direction of the particle motion, if the particle has a positive charge (more generally, the $\hat{P}$ in the argument of F refers to the direction of the particle's rigidity rather than the direction of motion). Since the particle's rigidity is constant on the trajectory we have

$$F(\vec{X},\hat{P}) = \text{constant} \tag{2}$$

when evaluated on the trajectory of a particle that is initially at some point $\vec{X}_0$ and with initial rigidity in some arbitrary direction $\hat{P}_0$ with magnitude equal to $F(\vec{X}_0,\hat{P}_0)$. The above equation gives

$$\frac{d}{dS} F(\vec{X}(S),\hat{P}(S)) = 0 \tag{3}$$

where S represents arclength along the trajectory and this trajectory is represented by $\vec{X} = \vec{X}(S)$. $\hat{P}(S)$ is tangent to the trajectory in the direction of the particle rigidity. The sense of increasing arclength will be taken to be in the direction of $\hat{P}$ so (1) applies and we also have

$$\frac{d\vec{X}}{dS} = \hat{P} \quad . \tag{4}$$

If we let $\vec{\nabla}_x$ and $\vec{\nabla}_p$ be the del operators in position and rigidity space and apply the chain rule to (3) and use (1) and (4) we get

$$\hat{P} \cdot \vec{\nabla}_x F + (\hat{P} \times \vec{B}) \cdot \vec{\nabla}_p F \Big|_{P=F(\vec{X},\hat{P})} = 0 \quad . \tag{5}$$

Equation (5) is an unconventional kind of differential equation because the point of evaluation of $\vec{\nabla}_p F$ has P (the magnitude of the particle rigidity) equal to the unknown cutoff rigidity. This undesirable characteristic can be eliminated the following way. Let P, θ_p, ϕ_p denote spherical coordinates in rigidity space. Since F does not depend on the magnitude P, we can write

$$\vec{\nabla}_p F = \frac{1}{P} \left(\frac{\partial F}{\partial \theta_p} \hat{\theta}_p + \frac{1}{\sin \theta_p} \frac{\partial F}{\partial \phi_p} \hat{\phi}_p \right)$$

where $\hat{\theta}_p$ and $\hat{\phi}_p$ denote the obvious unit vectors. The above equation applies for arbitrary P (note that while F does not depend on P, the above equation shows that $\vec{\nabla}_p F$ does). In particular, the equation applies when the derivative is evaluated at P = F which gives

$$\vec{\nabla}_p F \big|_{P=F} = \frac{1}{F} \left(\frac{\partial F}{\partial \theta_p} \hat{\theta}_p + \frac{1}{\sin \theta_p} \frac{\partial F}{\partial \phi_p} \hat{\phi}_p \right) .$$

The quantity in parentheses was to be evaluated at P = F but since F does not depend on P it can be evaluated at arbitrary P. Combining the two equations gives

$$\vec{\nabla}_p F \big|_{P=F} = \frac{P}{F} \vec{\nabla}_p F$$

so that (5) becomes

$$\hat{P} \cdot \vec{\nabla}_x F + (\hat{P} \times \vec{B}) \cdot \frac{P}{F} \vec{\nabla}_p F = 0 \tag{6}$$

where P is now arbitrary. $\hat{P}$ and $\vec{X}$ are also arbitrary because the initial conditions of the particle trajectory can be chosen arbitrarily. So in (6), each component of $\vec{X}$ and of $\vec{P} = P\hat{P}$ can be treated as independent variables. Equation (6) is one way to write the general field equation. Note that since F does not depend on P, $\vec{\nabla}_p F$ is perpendicular to $\hat{P}$, i.e.,

$$\hat{P} \cdot \vec{\nabla}_p F = 0 \ . \tag{7}$$

Equations (6) and (7) and a little vector manipulating produce an alternate form

$$\hat{P} \cdot [\vec{\nabla}_x F + \vec{B} \times \frac{P}{F} \vec{\nabla}_p F] = 0 \, . \tag{8}$$

3. Boundary Conditions, Uniqueness, and Existence of Solutions

The field equation (equation (6) or (8)) must be supplemented by boundary conditions before a solution can be obtained. Existence and, to a lesser extent, uniqueness of solutions are sometimes taken for granted in physical problems. But existence and uniqueness theorems are still important when dealing with boundary value problems because they indicate what kind of information and how much information must be supplied in the form of boundary conditions before the equation can be solved. This section will not go into any elaborate mathematical detail on the subject but will, instead, only make a few qualitative comments.

If equation (1), the force equation, is taken as a background hypothesis, the steps that produced (6) from (3) are reversable, so the two equations are equivalent. Therefore, existence and uniqueness of solutions to (6) can be investigated by investigating those properties of (3). It will be assumed that the boundary conditions have the following basic structure. Suppose we are seeking a solution that is to apply to some region R of the three-dimensional position space. Let A be the boundary surface of R. It might be that A is a closed surface, so that R is a finite region, but not necessarily. At any given point on A, we assume that the cutoff has been specified in all directions. This is not the only structure that boundary conditions can have and from the point of view of existence and uniqueness theorems it may not be the best structure. But it is conceptually simple and because of this conceptual simplicity, this kind of boundary condition might show up frequently in practical applications (note that if the boundary conditions are obtained by trajectory tracing, the investigator can make the boundary conditions have any structure that he wants).

It is not difficult to see that existence of solutions will place some rather severe restrictions on the boundary conditions that can be used. One restriction is the following. Consider a point $\vec{X}$ on A and an arbitrary direction $\hat{P}$ and let $F(\vec{X},\hat{P})$ be the cutoff that was supplied as a boundary value. Let a particle have initial position $\vec{X}$ and initial rigidity in the direction $\hat{P}$ and with magnitude equal to the cutoff $F(\vec{X},\hat{P})$. According to (3), the cutoff rigidity is constant on the particle trajectory when evaluated in the direction of the particle rigidity (which is tangent to the trajectory). If there is another point where the trajectory intersects the surface A and if the point is represented by $\vec{X}'$ and the direction of the particle rigidity is represented by $\hat{P}'$, the boundary value at $\vec{X}',\hat{P}'$ must be the same as it was at $\vec{X},\hat{P}$, i.e.,

$$F(\vec{X}',\hat{P}') = F(\vec{X},\hat{P}) \ .$$

The above condition is a special case of a slightly stronger restriction on the boundary values, which is as follows. Consider one point $\vec{X}_1$ on A and a direction $\hat{P}_1$ and another point $\vec{X}_2$ on A ($\vec{X}_2 \neq \vec{X}_1$) with another direction $\hat{P}_2$. Let a particle have initial position $\vec{X}_1$ and direction $\hat{P}_1$ and rigidity $P_1 = F(\vec{X}_1,\hat{P}_1)$ and let another particle have initial position $\vec{X}_2$ and direction $\hat{P}_2$ with rigidity $P_2 = F(\vec{X}_2,\hat{P}_2)$. The condition that must be satisfied is that if $P_1 \neq P_2$, then their values must compare in such a way that the trajectories of the two particles do not have a point of tangency. (Note that if $P_1 = P_2$, a point of tangency is possible because $\vec{X}_1$ and $\vec{X}_2$ could be different points on the same trajectory so that the laws of mechanics are not violated.) Stated another way, if the trajectories have a point of tangency, then $P_1 = P_2$. This can be stated in still another way. Let $\vec{X}_0$ be an arbitrary point in R and let $\hat{P}_0$ be an arbitrary direction. Consider a set of particle trajectories associated with different values of rigidity P. Existence of a solution requires that there is not more than one value of P which will produce a trajectory such that the boundary value of the cutoff at the intersection of the trajectory with A (and in the direction of the trajectory) is equal to P. This is a strong requirement on the boundary values and it is not obvious how we can test the boundary values to see if the requirement is satisfied. One way around this difficulty is to work only with boundary values that represent a realistic problem. Boundary values that represent real trajectories in a given field will automatically satisfy this requirement.

Assuming that the conditions required for existence of solutions are satisfied, the next question is uniqueness. No matter how small the surface A is chosen, uniqueness is guaranteed for at least some arguments of F. F is unique at any point $\vec{X}$ and direction $\hat{P}$ such that there exists a point $\vec{X}'$ on A and a direction $\hat{P}'$ such that a trajectory with initial position $\vec{X}'$ and direction $\hat{P}'$ and rigidity $P = F(\vec{X}', \hat{P}')$ will reach the point $\vec{X}$ and direction $\hat{P}$. There are obviously some choices of $\vec{X}$ and $\hat{P}$ that will satisfy this condition. But for some selected $\vec{X}$ and $\hat{P}$, the requirement that this point and direction can be reached by such a trajectory puts some restrictions on where the surface A can be located. At a point $\vec{X}$ in the region R and for a direction $\hat{P}$, let $F(\vec{X}, \hat{P})$ be the correct value for the cutoff. The boundary values on the surface A can only tell us what this value $F(\vec{X}, \hat{P})$ is if the trajectory, characterized by $\vec{X}, \hat{P}$ and $P = F(\vec{X}, \hat{P})$, intersects A. The trajectory is undefined at time equal to infinity (a particle with rigidity exactly equal to the cutoff is as much bound as it is free) but over any finite distance of travel the trajectory is a continuous function of the rigidity. This means that over any fixed but finite distance of travel, the trajectory is approximately that of a particle that is bound but almost able to escape. Therefore, over any finite distance, the particle trajectory is confined to the regions of space where the magnetic field is strong enough to keep it bound. In order for the particle to reach the surface A, the surface will have to extend into these regions of space. We should anticipate problems (i.e., not having enough information to solve the equation) if the boundary surface is taken to be an infinitely large sphere or if the magnetic field is localized and the surface is taken to be entirely outside of the magnetic field.

In an actual application, the decision on what boundary conditions to construct might be made the following way. We might let uniqueness conditions decide what boundary surface to use. We would look for a surface (for example, an infinite plane) that we expect, through physical intuition, to have the property that it will be intersected at least once by the trajectory of any particle that is trapped but almost free. If our physical intuition is wrong, we might find that there is a limited region in the five-dimensional $\vec{X},\hat{P}$ space where the equation can not be solved. Hopefully, this will usually not be a serious problem. The boundary values must comply with the existence conditions which puts strong constraints on them. One obvious way to comply with the existence conditions is to carefully construct the boundary values so that they realistically represent the magnetic field. Arbitrarily chosen boundary values that are intended to represent a "hypothetical problem" are likely to lead to problems.

The solution to (6) depends strongly on the boundary conditions in the sense that it is difficult to extract much information from the equation without being specific about the boundary conditions and actually solving the equation. We can intuitively see why this is so. The equation only "sees" the magnetic field in some limited region of space R. The magnetic field could be anything outside of R and the only way the equation "knows" what the magnetic field looks like outside of R is through the boundary conditions. It would be nice if the region R could be taken to be all space (so that the equation will "see" the magnetic field everywhere) and the boundary conditions taken to be that $F = 0$ at infinity. We could probably get more information out of the equation (without actually solving it) if we could do this. But as already mentioned, there may be some mathematical difficulties with this. However, in some special cases (axially symmetric magnetic field) it is possible to supplement (6) with a little more physical information and the resulting equation can make some useful predictions without the need of being specific about the boundary conditions. Furthermore, less information in the form of boundary conditions is needed to solve the equation. This will be discussed in a later section.

4. Mixed Coordinates

The position and rigidity gradients in (6) treat $\vec{X}$ and $\vec{P}$ as independent quantities. However, even though $\vec{P}$ is not treated as a function of $\vec{X}$, the coordinates that are selected to measure $\vec{P}$ may depend on $\vec{X}$ as well as on $\vec{P}$ (examples are azimuth and zenith angles). Coordinates that are intended to refer to $\vec{P}$ but that actually depend on $\vec{X}$ as well as on $\vec{P}$ will be called mixed coordinates. If we select a set of coordinates with which to express the field equation, we would want the partial derivative with respect to a given coordinate to mean that the remaining coordinates are held fixed while the given coordinate is varied. $\vec{\nabla}_x$ does not have that interpretation if mixed coordinates are used. This operator varies the position coordinate while holding the vector $\vec{P}$ fixed and this will cause the rigidity coordinates to vary. It will be useful to rewrite (6) in terms of different kinds of derivatives which have the desired interpretation.

Let $u_1(\vec{X},\vec{P})$ and $u_2(\vec{X},\vec{P})$ be two coordinates that are intended to represent $\hat{P}$. Let us change the function F in (6) so that it refers to the new arguments u_1 and u_2, i.e., so that the cutoff rigidity can be expressed as $F(\vec{X},u_1(\vec{X},\vec{P}),$ $u_2(\vec{X},\vec{P}))$. Note

$$\frac{\partial}{\partial X_i} F(\vec{X},u_1(\vec{X},\vec{P}),\ u_2(\vec{X},\vec{P})) = F_{x_i} + F_{u_1}\frac{\partial u_1}{\partial X_i} + F_{u_2}\frac{\partial u_2}{\partial X_i}$$

where subscripts to the F indicate derivatives with respect to the indicated arguments with the other arguments held fixed. Subscripts to F should not be confused with components of a vector which are also represented by subscripts. In vector form the equation becomes

$$\vec{\nabla}_x F = \vec{\partial}_x F + F_{u_1}\vec{\nabla}_x u_1 + F_{u_2}\vec{\nabla}_x u_2$$

where $\vec{\partial}_x$ denotes a "partial gradient." This is a gradient that holds u_1 and u_2 constant while varying $\vec{X}$. Equation (6) becomes

$$\hat{P} \cdot \vec{\partial}_x F + [\hat{P} \cdot \vec{\nabla}_x u_1] F_{u_1} + [\hat{P} \cdot \vec{\nabla}_x u_2] F_{u_2} + (\hat{P} \times \vec{B}) \cdot \frac{P}{F} \vec{\nabla}_p F = 0 . \qquad (9)$$

Once the coordinates u_1 and u_2 have been selected, expressions for $\vec{\nabla}_x u_1$, $\vec{\nabla}_x u_2$ and $\vec{\nabla}_p F$ can be explicitly calculated and substituted into (9). The resulting equation will be in a form that is useful for the coordinates selected.

Example: Two Sets of Spherical Coordinates

Consider a coordinate system with rectangular coordinates X,Y,Z and associated unit vectors $\hat{e}_1, \hat{e}_2, \hat{e}_3$. Let r, θ, ϕ be the usual spherical coordinates in that system with unit vectors $\hat{r}, \hat{\theta}, \hat{\phi}$. Now consider a second coordinate system X',Y',Z' having rectangular unit vectors $\hat{e}_1', \hat{e}_2', \hat{e}_3'$ and let this second system be oriented so that $\hat{e}_3'$ is parallel to $\hat{\phi}$ and $\hat{e}_1'$ is parallel to $\hat{e}_z$. Let P, γ, α be the spherical coordinates in this second system with unit vectors $\hat{P}, \hat{\gamma}, \hat{\alpha}$. A diagram of the coordinate systems together with a list of miscellaneous vector identities can be found in Appendix 1. The coordinates r, θ, ϕ will be used to measure $\vec{X}$ and P, γ, α will be used to measure $\vec{P}$ so that u_1 and u_2 in (9) are given by $u_1 = \alpha$, $u_2 = \gamma$.

One of the identities in the appendix is

$$\hat{P} = [\sin\gamma \sin\alpha \cos\phi - \cos\gamma \sin\phi]\, \hat{e}_1 + [\sin\gamma \sin\alpha \sin\phi + \cos\gamma \cos\phi]\, \hat{e}_2 + \sin\gamma \cos\alpha\, \hat{e}_3 \qquad (10)$$

which gives

$$\sin\gamma \sin\alpha \cos\phi - \cos\gamma \sin\phi = \hat{P} \cdot \hat{e}_1 \qquad (11)$$

$$\sin\gamma \sin\alpha \sin\phi + \cos\gamma \cos\phi = \hat{P} \cdot \hat{e}_2 \qquad (12)$$

$$\sin\gamma \cos\alpha = \hat{P} \cdot \hat{e}_3 . \qquad (13)$$

In spherical coordinates,

$$\vec{\nabla}_x u_1 = \vec{\nabla}_x \alpha = \hat{r}\frac{\partial\alpha}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial\alpha}{\partial\theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial\alpha}{\partial\phi} .$$

In these derivatives, $\vec{P}$ is held fixed while the coordinates r,θ,ϕ are given independent variations. Referring to the relative orientation between the coordinate systems makes it obvious that if the direction $\hat{P}$ and the angles θ and ϕ are held fixed as r is varied, α does not change, which implies $\partial\alpha/\partial r = 0$. Therefore,

$$\vec{\nabla}_x u_1 = \frac{\hat{\theta}}{r}\frac{\partial\alpha}{\partial\theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial\alpha}{\partial\phi} . \tag{14}$$

Similarly,

$$\vec{\nabla}_x u_2 = \frac{\hat{\theta}}{r}\frac{\partial\gamma}{\partial\theta} + \frac{\hat{\phi}}{r\sin\theta}\frac{\partial\gamma}{\partial\phi} . \tag{15}$$

We can solve for the derivatives in (14) and (15) by differentiating (11), (12), and (13) with respect to θ and ϕ and solving the simultaneous equations. The equations will not be independent (six equations into four unknowns) so we will work only with (12) and (13). Differentiating with $\hat{P}$ held fixed gives

$$\cos\alpha\cos\gamma\frac{\partial\gamma}{\partial\theta} - \sin\alpha\sin\gamma\frac{\partial\alpha}{\partial\theta} = 0$$

$$\cos\alpha\cos\gamma\frac{\partial\gamma}{\partial\phi} - \sin\alpha\sin\gamma\frac{\partial\alpha}{\partial\phi} = 0$$

$$[\sin\alpha\sin\phi\cos\gamma - \cos\phi\sin\gamma]\frac{\partial\gamma}{\partial\phi} + \sin\gamma\sin\phi\cos\alpha\frac{\partial\alpha}{\partial\phi}$$
$$= \cos\gamma\sin\phi - \sin\gamma\sin\alpha\cos\phi$$

$$[\sin\alpha\sin\phi\cos\gamma - \cos\phi\sin\gamma]\frac{\partial\gamma}{\partial\theta} + \sin\gamma\sin\phi\cos\alpha\frac{\partial\alpha}{\partial\theta} = 0 .$$

The solution is

$$\frac{\partial \gamma}{\partial \theta} = \frac{\partial \alpha}{\partial \theta} = 0 \qquad \frac{\partial \alpha}{\partial \phi} = \cos \alpha \cot \gamma \qquad \frac{\partial \gamma}{\partial \phi} = \sin \alpha$$

which gives

$$\vec{\nabla}_x u_1 = \frac{\hat{\phi}}{r \sin \theta} \cos \alpha \cot \gamma \qquad \vec{\nabla}_x u_2 = \frac{\hat{\phi}}{r \sin \theta} \sin \alpha \ .$$

Another identity from Appendix 1 is

$$\hat{P} = \sin \gamma \cos (\alpha - \theta)\ \hat{r} + \sin \gamma \sin (\alpha - \theta)\ \hat{\theta} + \cos \gamma\ \hat{\phi} \tag{16}$$

so that

$$\hat{P} \bullet \vec{\nabla}_x u_1 = \frac{\cos \alpha \cos^2 \gamma}{r \sin \gamma \sin \theta} \qquad \hat{P} \bullet \vec{\nabla}_x u_2 = \frac{\cos \gamma \sin \alpha}{r \sin \theta} \ . \tag{17}$$

Also note

$$\vec{\partial}_x F = \hat{r} F_r + \frac{\hat{\theta}}{r} F_\theta + \frac{\hat{\phi}}{r \sin \theta} F_\phi$$

which, together with (16), gives

$$\hat{P} \bullet \vec{\partial}_x F = \sin \gamma \cos (\alpha - \theta)\ F_r + \frac{\sin \gamma \sin (\alpha - \theta)}{r} F_\theta + \frac{\cos \gamma}{r \sin \theta} F_\phi \ . \tag{18}$$

The derivatives in $\vec{\nabla}_p F$ hold the position coordinates fixed while varying $\vec{P}$ and this means that the primed coordinate system is fixed in space as $\vec{P}$ is varied. Therefore, $\vec{\nabla}_p$ can be expressed in terms of the spherical coordinates P, γ, α in the usual way, i.e.,

$$\vec{\nabla}_p = \hat{P} \frac{\partial}{\partial P} + \frac{\hat{\gamma}}{P} \frac{\partial}{\partial \gamma} + \frac{\hat{\alpha}}{P \sin \gamma} \frac{\partial}{\partial \alpha} \ .$$

But F does not depend on P so we have

$$\frac{P}{F} \vec{\nabla}_p F = \frac{1}{F} (\hat{\gamma} F_\gamma + \frac{\hat{\alpha}}{\sin \gamma} F_\alpha) \; .$$

Expressing $\vec{B}$ as $B_p \hat{P} + B_\gamma \hat{\gamma} + B_\alpha \hat{\alpha}$ gives

$$\hat{P} \times \vec{B} = B_\gamma \hat{\alpha} - B_\alpha \hat{\gamma}$$

so

$$(\hat{P} \times \vec{B}) \cdot \frac{P}{F} \vec{\nabla}_p F = \frac{1}{F} \{ \frac{B_\gamma}{\sin \gamma} F_\alpha - B_\alpha F_\gamma \} \; .$$

It is more customary to express the magnetic field in terms of the r, θ, ϕ components than the P, γ, α components and this conversion can be done by using a few more identities in Appendix 1 to get

$$\begin{aligned} B_\gamma &= \vec{B} \cdot \hat{\gamma} = \vec{B} \cdot [\cos \gamma \cos (\alpha - \theta) \, \hat{r} + \cos \gamma \sin (\alpha - \theta) \, \hat{\theta} - \sin \gamma \, \hat{\phi}] \\ &= B_r \cos \gamma \cos (\alpha - \theta) + B_\theta \cos \gamma \sin (\alpha - \theta) - B_\phi \sin \gamma \end{aligned}$$

and

$$\begin{aligned} B_\alpha &= \vec{B} \cdot \hat{\alpha} = \vec{B} \cdot [\sin (\theta - \alpha) \, \hat{r} + \cos (\theta - \alpha) \, \hat{\theta}] \\ &= B_r \sin (\theta - \alpha) + B_\theta \cos (\theta - \alpha) \end{aligned}$$

so that

$$(\hat{P} \times \vec{B}) \cdot \frac{P}{F} \vec{\nabla}_p F$$

$$= \frac{1}{F} \{\frac{F_\alpha}{\sin \gamma} [B_r \cos \gamma \cos (\alpha - \theta) + B_\theta \cos \gamma \sin (\alpha - \theta) - B_\phi \sin \gamma]$$

$$- F_\gamma [B_r \sin (\theta - \alpha) + B_\theta \cos (\theta - \alpha)]\} . \qquad (19)$$

Substituting (17), (18), and (19) into (9) finally produces

$$[\sin \gamma \cos (\alpha - \theta)] F_r + [\frac{\sin \gamma \sin (\alpha - \theta)}{r}] F_\theta + [\frac{\cos \gamma}{r \sin \theta}] F_\phi$$

$$+ [\frac{\cos \alpha \cos \gamma \cot \gamma}{r \sin \theta} + \frac{B_r}{F} \cot \gamma \cos (\alpha - \theta) + \frac{B_\theta}{F} \cot \gamma \sin (\alpha - \theta) - \frac{B_\phi}{F}] F_\alpha$$

$$+ [\frac{\cos \gamma \sin \alpha}{r \sin \theta} + \frac{B_r}{F} \sin (\alpha - \theta) - \frac{B_\theta}{F} \cos (\alpha - \theta)] F_\gamma = 0 \qquad (20)$$

which is the form of (6) that is appropriate for this particular choice of coordinates.

5. Magnetic Fields With Axial Symmetry

Other than dipole fields, which are sometimes used to approximate a planet's magnetic field, fields that are symmetric about an axis have limited applications in space physics. But investigating such fields may be useful in some laboratory applications and it may also be useful in spacecraft applications if the idea of using artificial fields to protect small regions of space from charged particles is ever taken seriously. Therefore, investigating axially symmetric fields seems worthwhile, especially because it is possible to go far towards solving the field equation (at least in the case of the vertical cutoff) with only a small amount of effort. As already mentioned in an earlier topic, when the magnetic field is symmetric about an axis, it is possible to supplement the field equation with a little more physics so that more information is built into the equation and less information must be supplied in the form of boundary conditions. The analysis used here starts with concepts that can be found in Ref. 2.

We will work with the same coordinates and coordinate systems that were described in the example under the discussion of mixed coordinates. These coordinates are shown in Appendix 1. The magnetic field is taken to be symmetric about the $\hat{e}_3$ direction.

It is shown in Ref. 2 that on an arbitrary particle trajectory there exists a constant C that satisfies

$$\cos \gamma = \frac{1}{Pr \sin \theta} [C - \frac{\Phi(r,\theta)}{2\pi}] \tag{21}$$

where P is the particle rigidity (not necessarily equal to the cutoff rigidity) and

$$\Phi(r,\theta) = \int \vec{B} \cdot d\vec{A} \tag{22}$$

where the integral is a surface integral on any surface bounded by the circle that has the spherical coordinates r and θ. This makes Φ a function of r and θ. C in (21) is constant on a trajectory but it can be different for different trajectories.

Let it be given that a particle was found at a point with coordinates r_o, θ_o, γ_o and rigidity P. Solving (21) for C gives

$$C = Pr_o \sin \theta_o \cos \gamma_o + \frac{\Phi(r_o,\theta_o)}{2\pi} \quad .$$

Substituting back into (21), we see that a necessary condition for a particle to reach a point having coordinates r,θ,γ is that the coordinates satisfy

$$\frac{1}{Pr \sin \theta} [Pr_o \sin \theta_o \cos \gamma_o + \frac{\Phi(r_o,\theta_o) - \Phi(r,\theta)}{2\pi}] = \cos \gamma \tag{23}$$

or

$$-1 \leq \frac{1}{Pr \sin\theta} [Pr_o \sin\theta_o \cos\gamma_o + \frac{\Phi(r_o,\theta_o) - \Phi(r,\theta)}{2\pi}] \leq 1$$

which gives

$$P(r \sin\theta + r_o \sin\theta_o \cos\gamma_o) \geq \frac{\Phi(r,\theta) - \Phi(r_o,\theta_o)}{2\pi}$$

and

$$P(r \sin\theta - r_o \sin\theta_o \cos\gamma_o) \geq \frac{\Phi(r_o,\theta_o) - \Phi(r,\theta)}{2\pi} \quad .$$

These inequalities imply:

If

$$\Phi(r,\theta) \geq \Phi(r_o,\theta_o)$$

then

$$P \geq \frac{\Phi(r,\theta) - \Phi(r_o,\theta_o)}{2\pi\ (r \sin\theta + r_o \sin\theta_o \cos\gamma_o)}$$

and

$$r \sin\theta + r_o \sin\theta_o \cos\gamma_o > 0.$$

If

$$\Phi(r,\theta) > \Phi(r_o,\theta_o)$$

and

$$r_o \sin \theta_o \cos \gamma_o - r \sin \theta > 0$$

then

$$P \le \frac{\Phi(r,\theta) - \Phi(r_o,\theta_o)}{2\pi(r_o \sin \theta_o \cos \gamma_o - r \sin \theta)} .$$

If

$$\Phi(r,\theta) < \Phi(r_o,\theta_o)$$

then

$$P \ge \frac{\Phi(r_o,\theta_o) - \Phi(r,\theta)}{2\pi(r \sin \theta - r_o \sin \theta_o \cos \gamma_o)}$$

and

$$r \sin \theta - r_o \sin \theta_o \cos \gamma_o > 0.$$

If

$$\Phi(r,\theta) < \Phi(r_o,\theta_o)$$

and

$$r \sin \theta + r_o \sin \theta_o \cos \gamma_o < 0$$

then

$$P \le \frac{\Phi(r_o,\theta_o) - \Phi(r,\theta)}{2\pi|r \sin \theta + r_o \sin \theta_o \cos \gamma_o|} .$$

The inequalities that put upper bounds on P indicate the existence of at least one band of allowed rigidities where, in this case, "allowed" rigidities refer to rigidities such that it is possible to connect the coordinates r_o, θ_o, γ_o to the point r, θ. We will derive a generalization of the Störmer cutoff which is a rigidity that must be exceeded in order for a particle to escape to infinity. This kind of cutoff is obtained by working with the implications that place a lower bound on P. Those implications are equivalent to the statement

$$P \geq f(r_o, \theta_o, \gamma_o;\ r, \theta) \geq 0 \tag{24}$$

where f is defined by

$$f(r_o, \theta_o, \gamma_o;\ r, \theta) = \begin{cases} \dfrac{\Phi(r,\theta) - \Phi(r_o,\theta_o)}{2\pi(r \sin\theta + r_o \sin\theta_o \cos\gamma_o)} & \text{if } \Phi(r,\theta) \geq \Phi(r_o,\theta_o) \\[2ex] \dfrac{\Phi(r_o,\theta_o) - \Phi(r,\theta)}{2\pi(r \sin\theta - r_o \sin\theta_o \cos\gamma_o)} & \text{if } \Phi(r_o,\theta_o) > \Phi(r,\theta) \end{cases} \tag{25}$$

The condition (24) is a necessary condition for the particle to reach the point r, θ. (Note that in the arguments that produced (24), it is immaterial whether r_o, θ_o, γ_o are the initial coordinates and r, θ the final coordinates or vice versa. So f is a cutoff rigidity at r_o, θ_o, γ_o for particles going to r, θ or coming from r, θ.) A necessary condition for the particle to reach the radial distance r is that there exists a θ that will satisfy (24). But there will exist a θ satisfying (24) if and only if

$$P \geq \min_{\theta+} f(r_o, \theta_o, \gamma_o;\ r, \theta) \tag{26}$$

where the plus sign in the minimization symbol indicates that the minimization is done over the values of θ that result in f being nonnegative, i.e., the denominator of the appropriate expression on the right side of (25) must be positive. The condition (26) is a necessary condition for the particle to reach the radial distance r but there is another, stronger condition. This condition is that the particle be able to reach every radial distance r' that is between r_o and r, i.e., we must have

$$P \geq \min_{\theta+} f(r_o,\theta_o,\gamma_o;\ r',\theta) \qquad \text{for every } r' \varepsilon (r_o, r)$$

which implies

$$P \geq \max_{r' \varepsilon [r_o, r]} \ \min_{\theta+} f(r_o,\theta_o,\gamma_o;\ r',\theta). \tag{27}$$

The reader might object to the use of minimums and maximums in (27) instead of greatest lower bounds and least upper bounds. This is justified in Appendix 2 which also discusses some other interesting mathematical properties of the cutoff.

Condition (27) is a necessary condition for the particle to reach the radial distance r. A necessary condition for a particle, with initial coordinates r_o,θ_o,γ_o, to escape to infinity is

$$P \geq F(r_o,\theta_o,\gamma_o) \tag{28}$$

where we define

$$F(r_o,\theta_o,\gamma_o) = \max_{r>r_o} \ \min_{\theta+} \ f(r_o,\theta_o,\gamma_o;\ r,\theta). \tag{29}$$

We will call F, defined by (29), the generalized Störmer cutoff. A necessary condition for a particle to reach infinity is, according to (28), that its rigidity exceed the generalized Störmer cutoff.

Equations (25) and (29) could be used to solve for the generalized Störmer cutoff for a given magnetic field. But these equations are not very enlightening in the sense of demonstrating general properties. It is more convenient to work with (20) and use (25) and (29) to obtain boundary conditions. However, it is not obvious that the cutoff defined by (29) must satisfy (20). The derivation of (20) started with (2). The reasoning that produced (2) is intuitively obvious when the cutoff is an upper or lower bound to a rigidity band, i.e., when it has the property that an infinitesimal change in its value makes the difference between a bound trajectory and an escape trajectory. But it wasn't shown that this is the case for the generalized Störmer cutoff. A particle's rigidity must exceed this cutoff in order for the particle to escape but it wasn't shown that if the rigidity does exceed this cutoff by a small amount the particle will escape. It is therefore not obvious that (2) must apply but it is shown in Appendix 2 that (2) does apply and therefore we can use (20).

Note that the generalization of the Störmer cutoff does not depend on α. Assuming that this is the kind of cutoff that is being investigated (which will be the case if we are looking for a necessary but not sufficient condition for particles initially outside the magnetic field to reach a location inside the field), F_α can be set to zero in (20). Obviously F_ϕ is also zero so the equation, after some rearranging, becomes

$$\cos\theta \sin\gamma\, F_r - \frac{\sin\theta\sin\gamma}{r} F_\theta - [\frac{B_r}{F}\sin\theta + \frac{B_\theta}{F}\cos\theta]\, F_\gamma$$

$$= -\tan\alpha\, \{\sin\theta\sin\gamma\, F_r + \frac{\cos\theta\sin\gamma}{r} F_\theta$$

$$+ [\frac{B_r}{F}\cos\theta - \frac{B_\theta}{F}\sin\theta + \frac{\cos\gamma}{r\sin\theta}]\, F_\gamma\}\ .$$

But the left side of the above equation does not depend on α and neither does the coefficient to the $\tan\alpha$ on the right side. So these quantities must individually be zero which produces

$$F_r - \frac{\tan\theta}{r} F_\theta = [\frac{B_r}{F}\frac{\tan\theta}{\sin\gamma} + \frac{B_\theta}{F\sin\gamma}]\, F_\gamma \tag{30}$$

$$F_r + \frac{\cot\theta}{r} F_\theta = -\, [\frac{B_r}{F}\frac{\cot\theta}{\sin\gamma} - \frac{B_\theta}{F\sin\gamma} + \frac{\cot\gamma}{r\sin^2\theta}]\, F_\gamma . \tag{31}$$

The above equations are the final field equations but it is possible to write them in a variety of ways by using one equation to solve for a quantity and substituting into the other. One alternate form is

$$F_r + \left[\frac{\cot \gamma}{r} - \frac{B_\theta}{F \sin \gamma}\right] F_\gamma = 0 \tag{32}$$

$$F_\theta + \left[\cot \theta \cot \gamma + \frac{rB_r}{F \sin \gamma}\right] F_\gamma = 0 \ . \tag{33}$$

Another equation that can be produced from this system is

$$(rB_r + F \cot \theta \cos \gamma)\ F_r + (rB_\theta - F \cos \gamma)\ \frac{1}{r}\ F_\theta = 0 \ . \tag{34}$$

6. Vertical Cutoff

The vertical cutoff for a field with axial symmetry is the cutoff in a direction that is perpendicular to magnetic east, i.e., $\gamma = 90°$. Solving for the vertical cutoff does not immediately tell us the cutoff for other directions but in engineering applications the vertical cutoff is useful to the extent that it can give a rough representation of cutoffs for arbitrary directions. By assuming that the cutoffs in all directions equal the vertical cutoff, we can typically obtain a reasonably accurate approximation of directional average fluxes in the case of a magnetic dipole (see Ref. 3) and also in the case of using a complex magnetic field which accurately represents the Earth's magnetic field (See Ref. 4. Here vertical cutoff means in the direction perpendicular to the surface of the Earth). We can anticipate that this will be the case for many magnetic field patterns and therefore we can anticipate that the vertical cutoff will often have useful engineering applications. Fortunately, it is easy to solve.

Setting γ equal to 90° in (34) produces

$$B_r F_r + \frac{B_\theta}{r} F_\theta = 0 \ .$$

This equation is the spherical coordinate representation of the vector equation

$$\vec{B} \cdot \vec{\nabla} F = 0$$

which states that the vertical cutoff is constant along a magnetic field line. If the vertical cutoff is known on a surface, it will also be known at any point in space that can be connected to that surface by a magnetic field line. Therefore, the only thing that requires any work is the construction of boundary conditions.

The boundary conditions are supplied by equations (25) and (29). These equations can find the cutoff at any point in space but the computations simplify if the equations are used to find the vertical cutoff on the plane $\theta_o = 90°$. At any point in space that can be connected to this plane by a magnetic field line, the vertical cutoff can be solved by following the magnetic field line until it intersects the plane. When applied to the vertical cutoff in the plane, the equations reduce to

$$F(r_o,\theta_o,\gamma_o)\Big|_{\theta_o=90°,\gamma_o=90°} = \max_{r>r_o} \; \min_{\theta} f(r_o,90°,90°;\, r,\theta) \tag{35}$$

where

$$f(r_o,90°,90°;\, r,\theta) = \frac{|\Phi(r,\theta) - \Phi(r_o,90°)|}{2\pi r \sin\theta} \tag{36}$$

Example 1: The Dipole Field

The Störmer cutoff for a dipole is already well known but the dipole provides a simple example to demonstrate the use of the equations in this publication. The magnetic field is given by

$$\vec{B} = \frac{M}{r^3}\,[2\hat{r}\cos\theta + \hat{\theta}\sin\theta] \tag{37}$$

where M is the dipole moment (the dipole moment used here is $\mu_o/4\pi$ times the kind of dipole moment that has the dimensions of current times area). The dipole moment vector points in the $\hat{e}_3$ direction. The flux is given by

$$\Phi(r,\theta) = \int \vec{B} \cdot d\vec{S} = 2\pi r^2 \int B_r \sin\theta d\theta = \frac{4\pi M}{r} \int_o^\theta \cos x \sin x \, dx$$

$$= \frac{\pi M}{r} [1 - \cos 2\theta]$$

and

$$\Phi(r_o, 90°) = \frac{2\pi M}{r_o}$$

so (36) becomes

$$f(r_o, 90°, 90°;\ r, \theta) = \frac{2Mr - Mr_o + Mr_o \cos 2\theta}{2r_o r^2 \sin\theta}$$

where we have used the condition $r > r_o$ to remove the absolute value sign. We can see by inspection that the minimum in θ of F occurs when $\theta = 90°$ and (35) becomes

$$F(r_o, \theta_o, \gamma_o) \Big|_{\theta_o = 90°,\ \gamma_o = 90°} = \max_{r > r_o} \frac{M(r - r_o)}{r_o r^2} \ .$$

Maximizing in r yields

$$F(r_o, \theta_o, \gamma_o) \Big|_{\theta_o = 90°,\ \gamma_o = 90°} = \frac{M}{4r_o^2} \ .$$

Since only one set of coordinates remain, we will drop the subscripts and write

$$F(r,\theta,\gamma)\Big|_{\theta=90°,\ \gamma=90°} = \frac{M}{4r^2} \tag{38}$$

which is the vertical cutoff in the equatorial plane. The vertical cutoff is constant on a magnetic field line so to find its value at an arbitrary point in space, we simply follow the magnetic field line that passes through that point to see where it intersects the equatorial plane and use (38). In this example it is not difficult to express the solution in equation form. A field line satisfies the equation

$$\frac{d\vec{X}}{dS} = \frac{\vec{B}}{B} .$$

But

$$\frac{d\vec{X}}{dS} = \frac{dr}{dS}\,\hat{r} + r\,\frac{d\theta}{dS}\,\hat{\theta} + r\sin\theta\,\frac{d\phi}{dS}\,\hat{\phi} = \frac{dr}{dS}\,\hat{r} + r\,\frac{d\theta}{dS}\,\hat{\theta}$$

so the vector equation produces the two equations

$$\frac{dr}{dS} = \frac{B_r}{B} , \qquad r\,\frac{d\theta}{dS} = \frac{B_\theta}{B}$$

which gives

$$\frac{dr}{d\theta} = r\,\frac{B_r}{B_\theta} . \tag{39}$$

Using (36) gives

$$\frac{dr}{d\theta} = 2r\cot\theta .$$

Separating variables and integrating gives

$$r = K \sin^2 \theta$$

where K is a constant that identifies the magnetic field line that is being investigated. Now consider an arbitrary point r,θ. The field line that passes through this point intersects the equatorial plane at r_o which satisfies

$$r_o = \frac{r_o}{\sin^2 (90°)} = K = \frac{r}{\sin^2 \theta} \quad .$$

The vertical cutoff at r,θ is obtained by evaluating (38) at $r_o = r/\sin^2 \theta$ and the result is

$$F(r,\theta,90°) = F(r_o,90°,90°) = \frac{M \sin^4 \theta}{4r^2} \quad . \tag{40}$$

The well-known solution for the Störmer cutoff for a dipole is (see Ref. 2)

$$F(r,\theta,\gamma) = \frac{M \sin^4 \theta}{r^2[1 + (1 - \cos \gamma \sin^3 \theta)^{1/2}]^2} \tag{41}$$

which obviously reduces to (40) when $\gamma = 90°$. In order to predict the more general form, equation (41), from the field equation, it would be necessary to generalize the boundary values for arbitrary γ and then include the γ dependence by using (32) and (33) or an equivalent system of equations. Direct substitution will show that the more general expression given by (41), with $\vec{B}$ given by (37), is a solution to the field equations (32) and (33). Incidentally, the dipole field is sufficiently simple that it is fairly straightforward to deduce (41) directly from (25) and (29).

Example 2: A Spherical Magnetopause

Let the magnetic field be a superposition of the dipole field in the previous example and a uniform field that is antialigned with the dipole moment. For some r_c, the field can be expressed as

$$\vec{B} = \frac{M}{r^3}\,[2\hat{r}\cos\theta + \hat{\theta}\sin\theta] - \frac{2M}{r_c^3}\,\hat{e}_3 \ .$$

The physical significance of r_c can be seen by using the above equation to obtain

$$\vec{B} \cdot \hat{r} = 0 \text{ at } r = r_c \ .$$

In other words, $\vec{B}$ has no radial component on the sphere of radius r_c. This means that it is permissible to stipulate that the magnetic field is zero for $r > r_c$, i.e., the field has a spherical magnetopause (note that the derivation of (21) made use of $\vec{B}$ having a zero divergence. This means that a magnetopause can't be chosen arbitrarily because it can't cut across magnetic field lines). The magnetic field will be chosen to satisfy

$$\vec{B} = \begin{cases} \dfrac{M}{r^3}\,[2\hat{r}\cos\theta + \hat{\theta}\sin\theta] - \dfrac{2M}{r_c^3}\,\hat{e}_3 & r \le r_c \\[2ex] 0 & r > r_c \ . \end{cases}$$

The flux is given by

$$\Phi(r,\theta) = 2\pi M \sin^2\theta\ [\frac{1}{r} - \frac{r^2}{r_c^3}]$$

for $r \le r_c$ and the flux is zero for $r > r_c$. The same methods that have been used for the dipole can be used for this example and the result is

$$F(r_o,90°,90°) = \begin{cases} 0 & r_o > r_c \\ \dfrac{M}{r_c^2}\left[\dfrac{r_c}{r_o} - \dfrac{r_o^2}{r_c^2}\right] & Ar_c \le r_o \le r_c \\ \dfrac{M}{r_c^2}\left\{\left[\dfrac{r_c}{r_o} - \dfrac{r_o^2}{r_c^2}\right]\dfrac{r_c}{r'} - \dfrac{r_c^2}{r'^2} + \dfrac{r'}{r_c}\right\} & r_o < Ar_c \end{cases} \tag{42}$$

where

$$A = \frac{1}{2^{1/3}}\{(1+\sqrt{5})^{1/3} + (1-\sqrt{5})^{1/3}\}$$

$$\frac{r'}{r_c} = (1+D^2)^{1/6}\{\cos(\tfrac{1}{3}\arctan D) - \sqrt{3}\sin(\tfrac{1}{3}\arctan D)\}$$

$$D = \left|\frac{1}{27}\left[\frac{r_o^2}{r_c^2} - \frac{r_c}{r_o}\right]^3 + 1\right|^{1/2} .$$

Using steps similar to those used in the dipole case, the equation for a magnetic field line can be solved and the result is

$$\frac{r}{r_c^3 - r^3} = K \sin^2 \theta$$

where K is a constant that identifies the field line. Now let r,θ denote an arbitrary point in space. If $r > r_c$, the cutoff is zero (the point is outside the spherical magnetopause). Otherwise, the cutoff at the point r,θ is given by

$$F(r,\theta,90°) = F(r_o,90°,90°)$$

where the expression on the right is given by (42) and r_o is the real solution (which is unique since $r < r_c$) to the equation

$$\frac{r}{(r_c^3 - r^3)\sin^2\theta} = K = \frac{r_o}{r_c^3 - r_o^3} .$$

7. Isotropically Protective Fields

In this section we consider fields that are symmetric about the Z axis and use the vertical cutoff as the quantitative measure of the protection provided by a field.

Consider a point in space $\vec{X}_o$. Translate the coordinate system so that the point lies in the xy plane and let it have the radial coordinate r_o. From condition (24) we see that the cutoff rigidity for the particle to reach the coordinates $r_o, \theta_o = 90°$, $\gamma_o = 90°$ when it has initial coordinates r, θ is

$$f(r_o, 90°, 90°; r, \theta) .$$

This is the protection that the field provides to the coordinates $r_o, \theta_o = 90°$ against particles that are initially at r, θ. We will say that the field is isotropically protective at the point $\vec{X}_o$ if for all $r > r_o$ and $\theta \varepsilon (0, 90°)$, the protection is at least as large as the protection against particles that are initially at r with $\theta = 90°$. In other words, the field is isotropically protective if

$$f(r_o, 90°, 90°; r, \theta) \geq f(r_o, 90°, 90°; r, 90°)$$

for all $r > r_o$ and all $\theta \varepsilon (0, 90°)$. This condition is, of course, equivalent to the statement

$$\min_{\theta} f(r_o, 90°, 90°; r, \theta) = f(r_o, 90°, 90°; r, 90°). \tag{43}$$

It was already seen in one of the examples in the previous section that the dipole field is isotropically protective, i.e., it satisfies (43), with respect to any point in the equatorial plane.

The practical advantage of isotropically protective fields can be seen from equation (35) which implies that the vertical cutoff is bounded according to

$$F(r_o,\theta_o,\gamma_o)\Big|_{\theta_o=90°,\ \gamma_o=90°} \leq \max_{r>r_o} f(r_o,90°,90°;\ r,90°)$$

or, using (36),

$$F(r_o,\theta_o,\gamma_o)\Big|_{\theta_o=90°,\ \gamma_o=90°} \leq \max_{r>r_o} \frac{|\Phi(r,90°) - \Phi(r_o,90°)|}{2\pi r} .$$

From (43) we see that the equality is used for isotropically protective fields which implies that these fields provide the greatest protection that can be obtained for a specified $B_\theta(r,90°)$. In other words, if we are given a functional form for the θ component of the field in the xy plane and consider the set of all fields that have this θ component in the xy plane, the greatest possible protection is provided by the isotropically protective fields. From a computational point of view, these fields have the advantage that the vertical cutoff is calculated from the equation

$$F(r_o,\theta_o,\gamma_o)\Big|_{\theta_o=90°,\ \gamma_o=90°} = \max_{r>r_o} \frac{|\Phi(r,90°) - \Phi(r_o,90°)|}{2\pi r} \tag{44}$$

which means that, for this calculation, the field only has to be known in the xy plane. Of course something has to be known about the field off of the plane in order to conclude that it is an isotropically protective field.

There is a systematic procedure for constructing these fields. The first step in this procedure is to select a function representing the θ component of the field in the xy plane. This step completely determines the protection that will be provided by the field (it also places an upper bound on the protection that can be provided by other kinds of fields having the same B_θ in the xy plane). The field is not yet uniquely determined. The next step in the construction is to select a generating function $h(r,\theta)$ which will be discussed in more detail below. The generating function controls the behavior of the field off of the xy plane. In practice, the choice of what generating function to use is likely to be determined by the geometry of the region of space that is intended to be protected. Since the vertical cutoff is constant on a magnetic field line, the protection provided to points in the xy plane at the radial distance r_o is also provided to all points on the surface defined by the magnetic field lines that intersect the plane at $r = r_o$. The shape of this surface can be controlled by making an appropriate selection (by trial and error or by other methods) of the generating function.

We will now investigate where the generating function comes from and how it is used to complete the construction of the field.

The field is to be constructed to satisfy (43) which can be expressed as

$$\min_{\theta} \frac{|\Phi(r,\theta) - \Phi(r_o,90°)|}{2\pi r \sin\theta} = \frac{|\Phi(r,90°) - \Phi(r_o,90°)|}{2\pi r} .$$

The field will be constructed to have mirror symmetry which means that when minimizing in θ we can confine our attention to the interval $[0,90°]$. A sufficient condition to obtain the above equation is that the function

$$\frac{|\Phi(r,\theta) - \Phi(r_o,90°)|}{2\pi r \sin\theta}$$

be a decreasing function of θ on the interval $(0,90°)$. This condition is obtained if

$$\frac{\partial}{\partial\theta}\;\frac{|\Phi(r,\theta)-\Phi(r_o,90°)|}{2\pi r\sin\theta}\le 0 \qquad \text{for } \theta\epsilon(0,90°). \tag{45}$$

Using

$$\Phi(r,\theta) = 2\pi r^2 \int_o^\theta B_r(r,x)\sin x\,dx \tag{46}$$

and calculating the derivative gives

$$\frac{\partial}{\partial\theta}\;\frac{|\Phi(r,\theta)-\Phi(r_o,90°)|}{2\pi r\sin\theta} = \begin{cases} rB_r(r,\theta) - \dfrac{[\Phi(r,\theta)-\Phi(r_o,90°)]}{2\pi r\sin^2\theta}\cos\theta & \text{if } \Phi(r,\theta) > \Phi(r_o,90°) \\[2ex] -rB_r(r,\theta) + \dfrac{[\Phi(r,\theta)-\Phi(r_o,90°)]}{2\pi r\sin^2\theta}\cos\theta & \text{if } \Phi(r,\theta) < \Phi(r_o,90°) \end{cases}$$

Select a generating function h(r,θ) which is arbitrary except for one condition which will be stated below. Construct the field to satisfy

$$rB_r(r,\theta) = h(r,\theta)\,\frac{[\Phi(r,\theta)-\Phi(r_o,90°)]}{2\pi r\sin^2\theta}\cos\theta\quad . \tag{47}$$

The expression for the derivative becomes

$$\frac{\partial}{\partial\theta}\;\frac{|\Phi(r,\theta)-\Phi(r_o,90°)|}{2\pi r\sin\theta} = \begin{cases} [h(r,\theta)-1]\,\dfrac{[\Phi(r,\theta)-\Phi(r_o,90°)]}{2\pi r\sin^2\theta}\cos\theta & \text{if } \Phi(r,\theta) > \Phi(r_o,90°) \\[2ex] [1-h(r,\theta)]\,\dfrac{[\Phi(r,\theta)-\Phi(r_o,90°)]}{2\pi r\sin^2\theta}\cos\theta & \text{if } \Phi(r,\theta) < \Phi(r_o,90°) \end{cases}$$

It is evident from the above equation that (45) will be satisfied if

$$h(r,\theta) \leq 1 \text{ for all } \theta\varepsilon(0,90°),\ r > r_o. \tag{48}$$

An isotropically protective field can be constructed from a given function $B_\theta(r,90°)$ by selecting any function $h(r,\theta)$ that satisfies (48) and then constructing the field from (47). However, (47) is not in a convenient form for constructing the field. It is more convenient to work with the auxiliary function j which is defined in terms of the generating function by

$$\tan\theta \frac{\partial}{\partial\theta} \ln j(r,\theta) = h(r,\theta) \tag{49}$$

which can be solved for j giving

$$j(r,\theta) = e^{\int h(r,\theta) \cot\theta \, d\theta} . \tag{50}$$

The integral in (50) is indefinite because additive constants to the integral change j by proportionality constants which have no effect on the final results. B_r can be related to j through (47) and (49) which gives

$$\left(\frac{\partial}{\partial\theta} \ln j(r,\theta)\right)^{-1} \sin\theta\, r^2\, B_r(r,\theta) = \frac{\Phi(r,\theta) - \Phi(r_o,90°)}{2\pi} .$$

Using (46), the above equation can be differentiated to give

$$\frac{\partial}{\partial\theta} \left[\left(\frac{\partial}{\partial\theta} \ln j(r,\theta)\right)^{-1} \sin\theta\right] B_r(r,\theta) + \left[\left(\frac{\partial}{\partial\theta} \ln j(r,\theta)\right)^{-1} \sin\theta\right] \frac{\partial}{\partial\theta} B_r(r,\theta)$$
$$= B_r(r,\theta) \sin\theta .$$

This equation can be rearranged into

$$\frac{\partial}{\partial\theta} B_r(r,\theta) - \frac{\partial}{\partial\theta} \left\{\ln \left[\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} j(r,\theta)\right]\right\} B_r(r,\theta) = 0 .$$

It is easy to verify that the solution is

$$B_r(r,\theta) = \frac{A(r)}{\sin\theta}\frac{\partial}{\partial\theta} j(r,\theta) \tag{51}$$

where A(r) is an arbitrary function of r. This arbitrariness can be removed by the requirement that the field have a zero divergence. A necessary condition for a zero divergence is

$$2\pi r^2 \int_0^{90°} B_r(r,\theta) \sin\theta \, d\theta = \Phi(r,90°) \tag{52}$$

where

$$\Phi(r,90°) = -2\pi \int_0^r B_\theta(x,90°)\, x \, dx. \tag{53}$$

$\Phi(r,90°)$ is regarded as a specified function of r since $B_\theta(r,90°)$ is regarded as a specified function of r. It is therefore acceptable to have the field expressed in terms of $\Phi(r,90°)$. Using (51) and (52) gives

$$A(r) = \frac{\Phi(r,90°)}{2\pi r^2 \, [j(r,90°) - j(r,0)]} \tag{54}$$

so (51) becomes

$$B_r(r,\theta) = \frac{\Phi(r,90°)}{2\pi r^2 \sin\theta \, [j(r,90°) - j(r,0)]} \frac{\partial}{\partial\theta} j(r,\theta). \tag{55}$$

Condition (54) is a necessary but not a sufficient condition for the field to have a zero divergence. The remaining conditions will be imposed on $B_\theta(r,\theta)$ for $\theta \neq 90°$. We require that

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 B_r(r,\theta)\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \, B_\theta(r,\theta)\right) = 0$$

or, from (55),

$$\frac{\partial}{\partial\theta}(\sin\theta\, B_\theta(r,\theta)) = -\frac{1}{r}\frac{\partial}{\partial r}\left\{\frac{\Phi(r,90°)}{2\pi[j(r,90°) - j(r,0)]}\frac{\partial}{\partial\theta} j(r,\theta)\right\}.$$

Since integration with respect to θ commutes with differentiation with respect to r, the equation can be integrated to give

$$\sin\theta\, B_\theta(r,\theta) = -\frac{1}{r}\frac{\partial}{\partial r}\left\{\frac{\Phi(r,90°)}{2\pi}\,\frac{j(r,\theta) - j(r,0)}{j(r,90°) - j(r,0)}\right\}. \tag{56}$$

Equations (50), (55) and (56) completely determine the field components in terms of the specified flux function $\Phi(r,90°)$ and the generating function $h(r,\theta)$. If $h(r,\theta)$ is chosen to satisfy (48) for some r_o, the field will be isotropically protective at the point $r = r_o$, $\theta = 90°$. Altogether, the magnetic field is associated with three quantities: The radial coordinate (r_o) of the point to be protected, the function $B_\theta(r,90°)$ (or, equivalently, the function $\Phi(r,90°)$), and the generating function $h(r,\theta)$. In practice it may happen that h is constructed so that it explicitly contains r_o to insure that (48) is satisfied at the r_o that has been selected. Whether h does or does not explicitly contain the coordinate r_o, the field will be isotropically protective at any point in the plane having a radial coordinate that satisfies (48) when substituted for r_o in (48).

Example: Generating function for a dipole

For an arbitrarily selected r_o, (47) can be used to solve for the generating function from a dipole field associated with the point $r = r_o$, $\theta = 90°$ which is to be protected. The result is

$$h(r,\theta) = \frac{2\sin^2\theta}{\sin^2\theta - \frac{r}{r_o}}.$$

It can be seen from inspection that (48) is satisfied, which confirms the already known fact that a dipole field is isotropically protective at an arbitrary point in the equatorial plane. This same generating function can also be used to construct other fields, associated with other functions $B_\theta(r,90°)$, that are isotropically protective at $r = r_o$, $\theta = 90°$. The auxiliary function j is obtained from (50) and is

$$j(r,\theta) = \frac{r - r_o \sin^2 \theta}{r} .$$

For a given flux function in the plane, $\Phi(r,90°)$, the components of the field are found from (55) and (56) to be given by

$$B_r(r,\theta) = \frac{\Phi(r,90°) \cos \theta}{\pi r^2} \tag{57}$$

$$\sin \theta \, B_\theta(r,\theta) = - \frac{\sin^2 \theta}{r} \frac{\partial}{\partial r} \frac{\Phi(r,90°)}{2\pi} = \sin^2 \theta \, B_\theta(r,90°). \tag{58}$$

Note that a point r_o was selected to be protected when constructing the generating function, but, in this case, the field does not depend on r_o (assuming that $B_\theta(r,90°)$ is not chosen to depend on r_o) so the field given by (57) and (58) is isotropically protective everywhere in the equatorial plane. If $B_\theta(r,90°)$ is chosen to be the function that applies to the dipole field, (57) and (58) give the components of the dipole field for $\theta \neq 90°$. For other choices of $B_\theta(r,90°)$, other isotropically protective fields will result from (57) and (58). A word of caution is in order. The field is required to have a zero divergence and this puts a restriction on the choice of functions that can be used for $B_\theta(r,90°)$. The only functions that can be used are the ones that will result in

$$\Phi(r,90°) \to 0 \qquad \text{as } r \to \infty.$$

8. Optimum Fields

This section will investigate the field that provides the greatest magnetic protection subject to a certain (admittedly, an unlikely) class of constraints. This information could be useful when artificial magnetic fields are used for the purpose of providing protection against charged particles. By following an appropriate pattern when winding wire around an electromagnet, it is possible to create almost any desired field pattern and the question arises as to which field pattern is the best. This section answers that question for axially symmetric geometries when the constraint involves only the normal component of the field in a plane perpendicular to the axis of symmetry. Although it is difficult to imagine circumstances where this kind of a constraint would be imposed on an experimenter, this section is still useful because solving for the protection provided by an optimum field provides us with inequalities that can be used to obtain simple upper bound estimates of the protection that can be provided by a given field. This is demonstrated in the example at the end of this section.

It is assumed that the set of points where the protection is required is a circle of radius r_o in the xy plane. It is also taken for granted that the vertical Störmer cutoff is an adequate description of the protection and that will be used as the quantitative measure of the magnetic protection. The constraint imposed on the experimenter is assumed to be of the form

$$\int_o^\infty G(r,B_\theta(r,90°))\,dr = C = \text{constant} \qquad (59)$$

for some function G which is arbitrary except for a few restrictions which will be discussed later.

The problem of finding the best field subject to the constraint (59) comes in two separate parts. The first part is to find the function $B_\theta(r,90°)$ which maximizes the right side of (44) which is an upper bound on the protection that can be provided by any field having the same $B_\theta(r,90°)$ and is the protection provided by an isotropically protective field. The second part is to construct, from this $B_\theta(r,90°)$, an isotropically protective field. Such a field is not unique since different generating functions will produce different fields. Since the vertical cutoff is constant on a magnetic field line, the same protection that is provided to the circle in the xy plane is also provided to all points on the surface that is defined by the magnetic field lines that intersect the circle. The shape of this surface depends on the generating function. If the experimenter has a preference concerning the shape of the surface, he might use this as the basis for selecting a generating function.

Methods for constructing isotropically protective fields, when the function $B_\theta(r,90°)$ is given, have already been provided in the previous section. Therefore, in this section we will confine our attention to constructing the function $B_\theta(r, 90°)$. The objective is to find the function that maximizes

$$\max_{r>r_0} \frac{|\Phi(r,90°) - \Phi(r_0,90°)|}{2\pi r} \tag{60}$$

subject to the constraint (59).

There are a few points concerning constraints that must be discussed. Although it is assumed that there is one constraint of the form (59) which can be selected arbitrarily (except for a few restrictions), we will actually be working with three constraints. The first constraint is somewhat artificial and its only purpose is to avoid certain mathematical difficulties. This constraint is the requirement that the field be confined to a finite region of space, i.e., there exists an A such that the field is zero for r > A. This is not a real constraint, in fact it increases the flexibility of the equations to follow, because A is arbitrary and we can, if desired, let A go to infinity. The primary motivation for imposing this condition is that there are cases where the equations to follow will produce mathematically undefined terms if A is taken to be infinite during the early steps of the calculations. This difficulty can be avoided by taking A to be finite until a final result is obtained and then let A go to infinity as the last step.

The second constraint is (59) which is supplied by the experimenter. Note that imposing some constraint of this form on the field is a necessary but not a sufficient condition for the optimum field to exist. To insure that an optimum field exists, it will be assumed that $G(r,B_\theta) > 0$ if $B_\theta \neq 0$ and that G depends on B_θ only through its absolute value, i.e., $G(r,-B_\theta) = G(r,B_\theta)$.

The last constraint results from the fact that the field has a zero divergence. The surface integral of the field on the section of the xy plane that lies inside the circle of radius A must be zero.

For notational brevity, we will write B(r) instead of $B_\theta(r,90°)$. It should be emphasized that in this section, B(r) represents a component of the field rather than the magnitude and it can therefore be negative. The constraints are summarized below for easy reference.

$$B(r) = 0 \qquad \text{for } r > A \tag{61}$$

$$\int_0^A G(r,B(r))dr = C \tag{62}$$

$$\int_0^A rB(r)dr = 0 \tag{63}$$

$$G(r,B) > 0 \qquad \text{if } B \neq 0 \text{ and } G(r,-B) = G(r,B). \tag{64}$$

The cutoff is given by

$$F(r_o) = \max_{r\varepsilon(r_o,A)} \frac{|\Phi(r,90°) - \Phi(r_o,90°)|}{2\pi r}$$

where for notational brevity we write $F(r_o)$ instead of $F(r_o,90°,90°)$. Using (53) gives

$$F(r_o) = \max_{r\varepsilon(r_o,A)} \{\frac{1}{r} \int_{r_o}^{r} \rho B(\rho)d\rho\} . \tag{65}$$

Note that the absolute value sign has been removed in (65). This is because we are looking for a field that maximizes the right side of (65) and the maximum value that can be obtained is the same whether or not the absolute value is retained. A field that produces the maximum in one case can differ, at most, by an overall negative sign from a field that produces the maximum in the other case and this distinction between the fields is of no consequence.

The optimum cutoff is the maximum in the function B of the right side of (65) which is itself a maximum in r. The maximizing can be done in either order and it is convenient to maximize in B first. The optimum cutoff is given by

$$F_{optimum} = \max_{r\epsilon(r_o,A)} \{\frac{1}{r} \max_{B} J_r[B]\} \tag{66}$$

where the functional J_r is defined by

$$J_r[B] = \int_{r_o}^{r} \rho B(\rho) d\rho \ . \tag{67}$$

We first find the function B that maximizes J_r subject to the constraints (62), (63) and (64). Using standard set theory notation, the set of functions that satisfy these constraints can be expressed as

$$\bigcup_{a\epsilon[0,C]} \{B | \int_{r_o}^{r} G(\rho,B)d\rho = a, \int_{o}^{A} G(\rho,B)d\rho = C, \int_{o}^{A} \rho B(\rho)d\rho = 0\}$$

so that

$$\max_{B} J_r[B] = \max \bigcup_{a\epsilon[0,C]} \{J_r[B] | \int_{r_o}^{r} G(\rho,B)d\rho = a, \int_{o}^{A} G(\rho,B)d\rho = C,$$
$$\int_{o}^{A} \rho B(\rho)d\rho = 0\} \ .$$

The maximum of the union can be obtained by replacing each set in the union with the set that contains only the maximum element of the original set. This gives

$$\max_{B} J_r[B] = \max \bigcup_{a \epsilon [0,C]} \{\max\{J_r[B] \mid \int_{r_o}^{r} G(\rho,B)d\rho = a,$$

$$\int_o^A G(\rho,B)d\rho = C,\ \int_o^A \rho B d\rho = 0\}\} \ . \tag{68}$$

We now must find the maximum of the set

$$\{J_r[B] \mid \int_{r_o}^{r} G(\rho,B)d\rho = a,\ \int_o^A G(\rho,B)d\rho = C,\ \int_o^A \rho B d\rho = 0\} \ . \tag{69}$$

From (67) we see that we are looking for a B that maximizes

$$\int_{r_o}^{r} \rho B d\rho \tag{70}$$

subject to the constraint

$$\int_{r_o}^{r} \rho B d\rho = - [\int_o^{r_o} \rho B d\rho + \int_r^A \rho B d\rho] \tag{71}$$

and the constraints

$$\int_{r_o}^{r} G(\rho,B)d\rho = a \tag{72}$$

$$\int_o^{r_o} G(\rho,B)d\rho + \int_r^A G(\rho,B)d\rho = C - a \ . \tag{73}$$

To find the maximum of (70), let B_1 be defined on the interval (r_o, r) and be chosen to produce the maximum of (70) subject to the constraint (72). Let B_2 be defined on $(0, r_o) \cup (r, A)$ and be chosen to produce the maximum of

$$- \left[\int_0^{r_o} \rho B d\rho + \int_r^A \rho B d\rho\right] \tag{74}$$

subject to the constraint (73). There are two possibilities to consider which are:

Case 1:

assume

$$\int_{r_o}^{r} \rho B_1 d\rho \leq - \left[\int_0^{r_o} \rho B_2 d\rho + \int_r^A \rho B_2 d\rho\right] . \tag{75}$$

Then we can find a B_2^* that will satisfy the constraint (73) and produce

$$\int_{r_o}^{r} \rho B_1 d\rho = - \left[\int_0^{r_o} \rho B_2^* d\rho + \int_r^A \rho B_2^* d\rho\right] . \tag{76}$$

If the strict inequality holds in (75), B_2^* will not be unique but it is easy to see that such a function must exist because it is possible to construct such a function out of B_2. This can be done by partitioning the domain $(0, r_o) \cup (r, A)$ into subintervals and reversing the sign of B_2 on selected subintervals so that the various contributions to the integrals in (74) subtract with each other instead of add. By condition (64), the constraint (73) is still satisfied. Using an appropriate partitioning and sign selection, (74) can be given any value between

$$\left[\int_r^{r_o} \rho B_2 d\rho + \int_r^A \rho B_2 d\rho\right]$$

and

$$-\left[\int_0^{r_o} \rho B_2 d\rho + \int_r^A \rho B_2 d\rho\right] .$$

In particular, it is possible to satisfy (76). Now define

$$B^*(\rho) = \begin{cases} B_1(\rho) & \text{if } \rho \in (r_o, r) \\ B_2^*(\rho) & \text{if } \rho \in (0, r_o) \cup (r, A) . \end{cases}$$

Note that any element $J_r[B]$ of the set (69) satisfies (67) subject to the constraint (72) and since B_1 was chosen to maximize (70) subject to (72) we have

$$J_r[B] \leq \int_{r_o}^{r} \rho B_1 d\rho = J_r[B^*]$$

i.e., $J_r[B^*]$ is an upper bound for the set (69). Also, since B_1 satisfies (72) and B_2^* satisfies (73) and (76), it is evident that B^* satisfies the constraints (71), (72) and (73). Therefore $J_r[B^*]$ is an element of the set (69) which implies that it is the maximum of the set. To get an explicit expression for this maximum value, it is necessary to solve for B_1. The maximum of (70) subject to the constraint (72) is a special case of the classic prototype problem, treated in the calculus of variations, where the integrands may depend on ρ, B_1 and the derivative of B_1. The well known equation governing the maximizing function is (see Ref. 5)

$$\left(\frac{\partial}{\partial B_1} - \frac{d}{d\rho}\frac{\partial}{\partial \frac{dB_1}{d\rho}}\right)(\rho B_1 + \lambda_1 G) = 0$$

where λ_1 is a constant which is chosen to satisfy the constraint. In our case, nothing depends on the derivative of B_1 so the equation reduces to

$$\rho + \lambda_{1,r,a} \frac{\partial}{\partial B_{1,r,a}} G(\rho, B_{1,r,a}) = 0 \qquad \text{for } \rho \in (r_o, r) . \tag{77}$$

The subscripts r and a were added to emphasize the fact that both B_1 and λ_1 will in general depend on r and a. This dependence is obtained through (72) which must be combined with (77) to produce two simultaneous equations in the unknown function $B_{1,r,a}$ and unknown constant $\lambda_{1,r,a}$.

The final conclusion for case 1 is that the maximum value of the set (69) is

$$\int_{r_o}^{r} \rho B_{1,r,a}\, d\rho$$

where $B_{1,r,a}$ is obtained from (77) and (72).

Case 2:

Now assume

$$\int_{r_o}^{r} \rho B_1 d\rho > - [\int_{0}^{r_o} \rho B_2 d\rho + \int_{r}^{A} \rho B_2 d\rho] .$$

We parallel the steps used in Case 1. Let B_1^* satisfy (72)

and

$$\int_{r_o}^{r} \rho B_1^*\, d\rho = - [\int_{0}^{r_o} \rho B_2 d\rho + \int_{r}^{A} \rho B_2 d\rho]$$

and define B^* by

$$B^*(\rho) = \begin{cases} B_1^*(\rho) & \text{if } \rho \in (r_o, r) \\ B_2(\rho) & \text{if } \rho \in (0, r_o) \cup (r, A) . \end{cases}$$

Any element $J_r[B]$ in the set (69) satisfies

$$J_r[B] = -\left[\int_0^{r_o} \rho B d\rho + \int_r^A \rho B d\rho\right]$$

subject to the constraint (73) and since B_2 was chosen to maximize (74) subject to (73), we have

$$J_r[B] \leq -\left[\int_0^{r_o} \rho B_2 d\rho + \int_r^A \rho B_2 d\rho\right] = J_r[B^*] ,$$

i.e., $J_r[B^*]$ is an upper bound for the set (69). $J_r[B^*]$ is also an element of the set which implies that it is the maximum of the set. To solve for this maximum we must solve for B_2. Note that the conventional derivation of the method of Lagrange multipliers applies not only to the case where there is one integration interval but also to the case where the region of integration is two disjoint intervals. The result is that B_2, which will now be represented by $B_{2,r,a}$ to emphasize the dependence on r and a, is given by

$$\rho + \lambda_{2,r,a} \frac{\partial}{\partial B_{2,r,a}} G(\rho, B_{2,r,a}) = 0 \qquad \text{for } \rho \in (0, r_o) \cup (r, A) \tag{78}$$

where $\lambda_{2,r,a}$ is a constant. The conclusion for Case 2 is that the maximum value of the set (69) is

$$-\left[\int_0^{r_o} \rho B_{2,r,a}\, d\rho + \int_r^A \rho B_{2,r,a}\, d\rho\right]$$

where $B_{2,r,a}$ is obtained from (78) and (73).

The two above cases can be combined to give

$$\max \left\{J_r[B] \,\Big|\, \int_{r_o}^{r} G(\rho,B)d\rho = a,\ \int_0^A G(\rho,B)d\rho = C,\ \int_0^A \rho B d\rho = 0\right\}$$

$$= \min \left\{\int_{r_o}^{r} \rho B_{1,r,a}\, d\rho,\ -\left[\int_0^{r_o} \rho B_{2,r,a}\, d\rho + \int_r^A \rho B_{2,r,a}\, d\rho\right]\right\} .$$

Using the above equation together with (66) and (68), it is seen that the optimum cutoff compatible with constraints (61), (62), (63) and (64) is given by

$$F_{optimum} = \max_{r\varepsilon(r_o,A)} [\frac{1}{r} \max_{a\varepsilon[0,C]} (\min \{\int_{r_o}^{r} \rho B_{1,r,a} d\rho, - [\int_{o}^{r_o} \rho B_{2,r,a} d\rho + \int_{r}^{A} \rho B_{2,r,a} d\rho]\})] \tag{79}$$

where $B_{1,r,a}$ is solved from the system of equations

$$\rho + \lambda_{1,r,a} \frac{\partial}{\partial B_{1,r,a}} G(\rho,B_{1,r,a}) = 0 \qquad \rho \ \varepsilon \ (r_o,r) \tag{80}$$

$$\int_{r_o}^{r} G(\rho,B_{1,r,a}) \ d\rho = a \tag{81}$$

and $B_{2,r,a}$ is solved from the system of equations

$$\rho + \lambda_{2,r,a} \frac{\partial}{\partial B_{2,r,a}} G(\rho,B_{2,r,a}) = 0 \qquad \rho \ \varepsilon \ (0,r_o) \cup (r,A) \tag{82}$$

$$\int_{o}^{r_o} G(\rho,B_{2,r,a}) \ d\rho + \int_{r}^{A} G(\rho,B_{2,r,a}) \ d\rho = C - a. \tag{83}$$

Equations (80) through (83) will leave $B_{1,r,a}$ and $B_{2,r,a}$ undetermined to within an overall negative sign. The signs are chosen to make the quantities inside the set in (79) positive. This is the sign convention that must be followed in order for (80) through (83) to produce maximum values for (70) and (74). Maximizing the right side of (79) in a and r can be done using elementary methods and is done on a case by case basis.

Example:

In this example the constraint will be that the surface integral of the energy density is specified, i.e.,

$$\int_0^A \rho B^2(\rho)d\rho = C \tag{84}$$

where C is a specified constant. It is difficult to imagine circumstances where this constraint would actually be imposed but this example is still useful because the results will provide a simple upper bound estimate of the protection that can be obtained from a given field.

Letting

$$G(\rho,B) = \rho B^2,$$

equations (80) and (82) produce

$$B_{1,r,a} = -\frac{1}{2\lambda_{1,r,a}}, \quad B_{2,r,a} = -\frac{1}{2\lambda_{2,r,a}} \tag{85}$$

and (81) and (83) produce

$$\lambda^2_{1,r,a} = \frac{r^2 - r_o^2}{8a}, \quad \lambda^2_{2,r,a} = \frac{r_o^2 + A^2 - r^2}{8\,(C - a)} .$$

The sign of $\lambda_{1,r,a}$ will be chosen to be negative so that

$$\int_{r_o}^{r} \rho B_{1,r,a}\, d\rho$$

will be positive and the sign of $\lambda_{2,r,a}$ will be chosen to be positive so that

$$-\left[\int_0^{r_o} \rho B_{2,r,a}d\rho + \int_r^A \rho B_{2,r,a}d\rho\right]$$

is positive. Replacing the λ's in (85) gives

$$B_{1,r,a} = \left[\frac{2a}{r^2 - r_o^2}\right]^{1/2} \qquad B_{2,r,a} = -\left[\frac{2\,(C - a)}{r_o^2 + A^2 - r^2}\right]^{1/2}$$

so that

$$\int_{r_o}^{r} \rho B_{1,r,a} d\rho = \left[\tfrac{1}{2}\, a\, (r^2 - r_o^2)\right]^{1/2} \tag{86}$$

$$-\left[\int_{o}^{r_o} \rho B_{2,r,a} d\rho + \int_{r}^{A} \rho B_{2,r,a} d\rho\right] = \left[\tfrac{1}{2}\,(C - a)\,(r_o^2 + A^2 - r^2)\right]^{1/2} . \tag{87}$$

Using (86) and (87), it is easy to show that

$$\max_{a\varepsilon[0,C]} \left(\min\left\{\int_{r_o}^{r} \rho B_{1,r,a} d\rho,\ -\left[\int_{o}^{r_o} \rho B_{2,r,a} d\rho + \int_{r}^{A} \rho B_{2,r,a} d\rho\right]\right\}\right)$$

$$= \left[\frac{C(r_o^2 + A^2 - r^2)(r^2 - r_o^2)}{2A^2}\right]^{1/2}$$

where the maximum occurs at

$$a = \frac{C}{A^2}\,(r_o^2 + A^2 - r^2) \ . \tag{88}$$

Equation (79) becomes

$$F_{optimum} = \max_{r\varepsilon(r_o,A)} \left(\frac{1}{r}\left[\frac{C(r_o^2 + A^2 - r^2)(r^2 - r_o^2)}{2A^2}\right]^{1/2}\right).$$

Maximizing in r gives the final result

$$F_{optimum} = \sqrt{\frac{C}{2}}\left(\sqrt{\frac{r_o^2}{A^2}+1} - \frac{r_o}{A}\right) \tag{89}$$

which occurs at

$$r^2 = r_o\sqrt{r_o^2 + A^2}\ . \tag{90}$$

When a and r are given by (88) and (90), the field is given by

$$B(\rho) = \begin{cases} \frac{\sqrt{2C}}{A}\left[1+\frac{A^2}{r_o^2}\right]^{1/4} & \rho\ \varepsilon\left(r_o,\ \sqrt{r_o\sqrt{r_o^2+A^2}}\right) \\ \frac{-\sqrt{2C}}{A}\left[\frac{r_o^2}{r_o^2+A^2}\right]^{1/4} & \rho\ \varepsilon\ (0,r_o)\ \mathrm{U}\left(\sqrt{r_o\sqrt{r_o^2+A^2}},\ A\right)\ . \end{cases} \tag{91}$$

Let us now relax the restriction that the field is confined to a finite region. The largest value (in A) of $F_{optimum}$ occurs, according to (89), when $A \to \infty$ and we have

$$F_{optimum} \to \sqrt{\frac{C}{2}}\ . \tag{92}$$

The field that produces this cutoff can be seen, from (91), to be the limiting case as the field becomes infinitely "spread out", that is, the field becomes infinitely weak but occupies an infinitely large volume of space. This field could never be produced but (92) is still useful as an upper bound on the protection that can be provided. Consider a field that is axially symmetric but otherwise arbitrary. Assuming that it is a real field (e.g., not the ideal dipole), it will have a finite amount of energy stored in it so that

$$\int_0^{\infty} \rho B^2 d\rho \tag{93}$$

is finite. The protection from this field could not exceed the protection from the optimum field subject to the constraint that C is equal to (93) evaluated at this specific B, and in fact the protection from this field must be less than that from the optimum field (since the optimum field is a limiting case that could not exist in reality) so the cutoff for the actual field must satisfy the inequality

$$F < \left[\frac{1}{2} \int_0^{\infty} \rho B^2 d\rho\right]^{1/2} . \tag{94}$$

Since r_o does not show up on the right side of (94), the upper bound applies to all points in the xy plane and also to any point in space that is connected to the xy plane by a magnetic field line.

9. Summary

This publication shows that magnetic cutoff rigidities can be regarded as a solution to a boundary value problem. This method of solving for the cutoff is a possible alternative to trajectory tracing, but in applications, trajectory tracing will probably be needed to supply the boundary conditions unless the magnetic field has special symmetries. A few qualitative comments, on existence and uniqueness of solutions, have been made which may be helpful when deciding how the boundary conditions should be set up. The field equation can be expressed in the form of (6) or (8). These equations can be applied to an arbitrary choice of coordinates providing that the coordinates representing the point of observation $\vec{X}$ and the coordinates representing the direction of observation $\hat{P}$ are separated in the sense that the mapping between the direction $\hat{P}$ and the coordinates that represent that direction is independent of $\vec{X}$. If "mixed" coordinates (which have the property that the coordinates that represent the direction $\hat{P}$, depend on $\vec{X}$ as well as on $\hat{P}$) are used, a more appropriate form for the field equation is given by (9) which applies to a more general class of coordinates. As a special case, (9) was expressed in terms of two sets of spherical coordinates. The result was (20). If the magnetic field has axial symmetry and the cutoff under investigation is the generalized Störmer cutoff, the cutoff is independent of α and satisfies (32) and (33). The vertical cutoff, which is the generalized Störmer cutoff evaluated in a direction perpendicular to magnetic east, is constant on a magnetic field line. This means that the cutoff can be evaluated at an arbitrary point in space by following the magnetic field line that passes through that point until it intersects a surface where the cutoff has been specified as a boundary condition. Analytic expressions for the boundary values are given by (35) and (36). Two examples were given to demonstrate the methods of calculation. Isotropically protective fields were defined and a method for constructing them is given by (50), (55) and (56). These fields were used in an optimization analysis which solves for the best magnetic protection that can be obtained subject to a constraint of the form (59). An example was given and a by-product of this example is (94) which is an inequality that can be used as an upper bound estimate of the protection that can be provided by a given field.

References

1. Shea, M. A., and D. F. Smart, "Tables of Asymptotic Directions and Vertical Cutoff Rigidities for a Five-Degree by Fifteen-Degree World Grid as Calculated Using the International Geomagnetic Reference Field for Epoch 1975.0," Report No. AFCRL-TR-75-0185, Hanscom, AFB, MA, 1975.

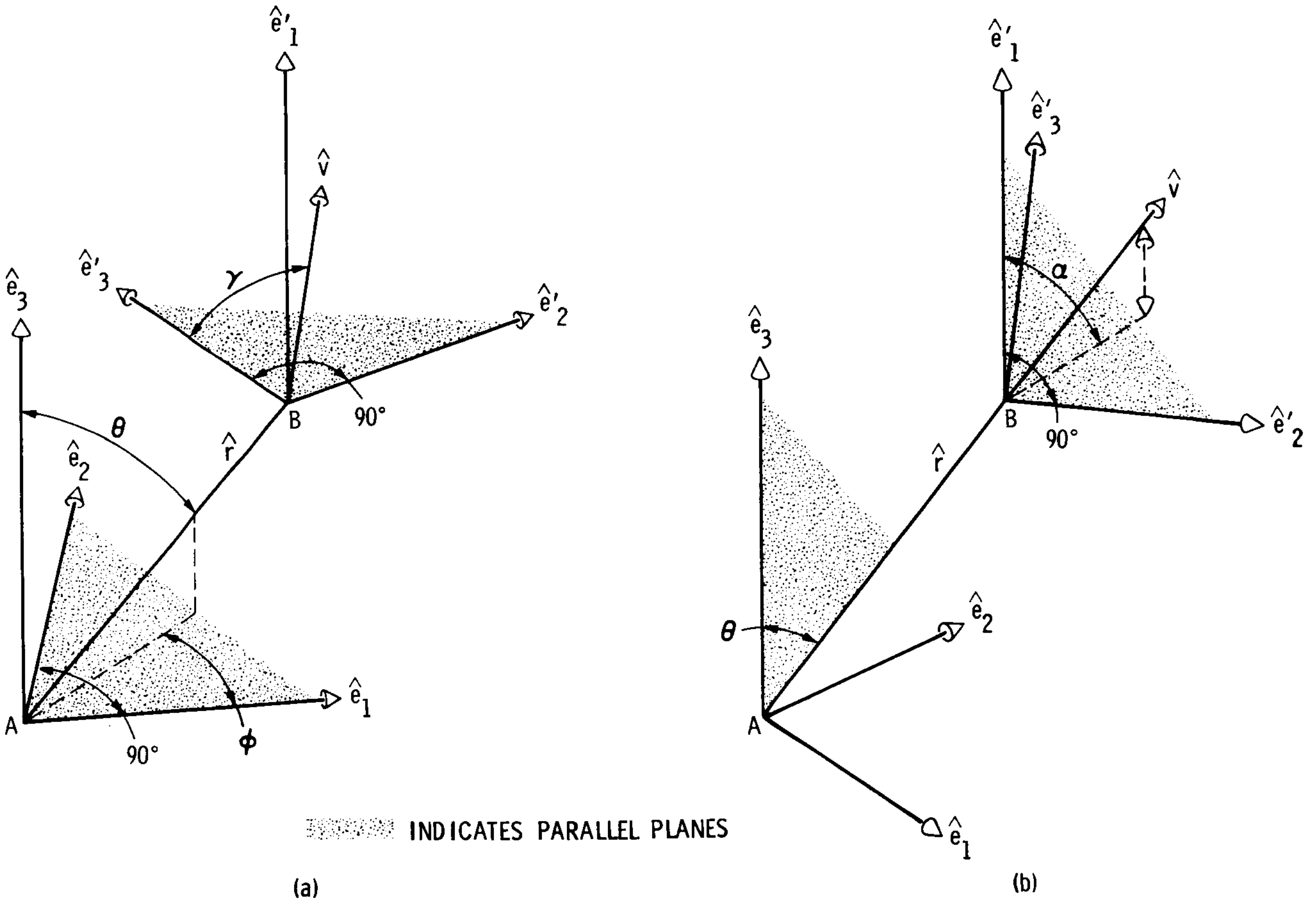
2. Rossi, B., and S. Olbert, Introduction to the Physics of Space, McGraw-Hill Book Company, 1970.

3. Edmonds, L. D., Final Report: Cosmic Ray Environment Model for Earth Orbit, JPL Publication 84-98, Jet Propulsion Laboratory, Pasadena, CA, January 15, 1985.

4. Adams, J. H., Jr., J. R. Letaw, and D. F. Smart, "Cosmic Ray Effects on Microelectronics Part II: The Geomagnetic Cutoff Effects," NRL Memorandum Report 5099, Naval Research Laboratory, May 26, 1983.

5. Gelfand, I. M., and S. V. Fomin, Calculus of Variations, Prentice-Hall, 1963.

Appendix 1: Diagram and Miscellaneous Vector Identities for a Particular Choice of Coordinates



$$\hat{e}'_1 = \hat{e}_3$$

$$\hat{e}'_2 = \hat{e}'_3 \times \hat{e}'_1 = \hat{\phi} \times \hat{e}_3 = \cos\phi\ \hat{e}_1 + \sin\phi\ \hat{e}_2$$

$$\hat{e}'_3 = \hat{\phi} = -\sin\phi\ \hat{e}_1 + \cos\phi\ \hat{e}_2$$

$$\hat{P} = [\sin\gamma \sin\alpha \cos\phi - \cos\gamma \sin\phi]\ \hat{e}_1 + [\sin\gamma \sin\alpha \sin\phi + \cos\gamma \cos\phi]\ \hat{e}_2$$
$$+ \sin\gamma \cos\alpha\ \hat{e}_3.$$

$$\hat{P} = \sin\gamma \cos(\alpha - \theta)\ \hat{r} + \sin\gamma \sin(\alpha - \theta)\ \hat{\theta} + \cos\gamma\ \hat{\phi}$$

$$\hat{\gamma} = [\cos\gamma \sin\alpha \cos\phi + \sin\gamma \sin\phi]\ \hat{e}_1 + [\cos\gamma \sin\alpha \sin\phi - \sin\gamma \cos\phi]\ \hat{e}_2$$
$$+ \cos\gamma \cos\alpha\ \hat{e}_3$$

$$\hat{\gamma} = \cos\gamma \cos(\alpha - \theta)\ \hat{r} + \cos\gamma \sin(\alpha - \theta)\ \hat{\theta} - \sin\gamma\ \hat{\phi}$$

$$\hat{\alpha} = \cos\alpha \cos\phi\ \hat{e}_1 + \cos\alpha \sin\phi\ \hat{e}_2 - \sin\alpha\ \hat{e}_3$$

$$\hat{\alpha} = \sin(\theta - \alpha)\ \hat{r} + \cos(\theta - \alpha)\ \hat{\theta}$$

$$\hat{r} = \sin\gamma \cos(\alpha - \theta)\ \hat{P} + \cos\gamma \cos(\alpha - \theta)\ \hat{\gamma} + \sin(\theta - \alpha)\ \hat{\alpha}$$

$$\hat{\theta} = \sin\gamma \sin(\alpha - \theta)\ \hat{P} + \cos\gamma \sin(\alpha - \theta)\ \hat{\gamma} + \cos(\theta - \alpha)\ \hat{\alpha}$$

$$\hat{\phi} = \cos\gamma\ \hat{P} - \sin\gamma\ \hat{\gamma}$$

Appendix 2: Derivation of Equation (2) for the Generalized Störmer Cutoff

Note that (29) is not a rigorous definition of the Störmer cutoff because the existence of maximums and minimums requires that a function and/or its domain possess certain properties and it was not shown that this is the case. A better definition of the cutoff is

$$F(r_o,\theta_o,\gamma_o) = \sup_{r>r_o} \inf_{\theta^+} f(r_o,\theta_o,\gamma_o;\ r,\theta). \tag{A2.1}$$

It will be shown that the cutoff, as defined by (A2.1), satisfies (2) and it will also be shown that the definition (A2.1) is equivalent to (29).

As a reminder, f is given by

$$f(r_o,\theta_o,\gamma_o;\ r,\theta) = \begin{cases} \dfrac{\Phi(r,\theta) - \Phi(r_o,\theta_o)}{2\pi(r \sin\theta + r_o \sin\theta_o \cos\gamma_o)} & \text{if } \Phi(r,\theta) \geq \Phi(r_o,\theta_o) \\[2ex] \dfrac{\Phi(r_o,\theta_o) - \Phi(r,\theta)}{2\pi(r \sin\theta - r_o \sin\theta_o \cos\gamma_o)} & \text{if } \Phi(r_o,\theta_o) > \Phi(r,\theta) \end{cases} \tag{A2.2}$$

and the "+" in the inf symbol in (A2.1) means to include all values of θ such that the denominator of the appropriate expression on the right side of (A2.2) is positive.

For arbitrary r_o,θ_o,γ_o and $r > r_o$, let $s(r)$ be the set of values of θ satisfying

$$\theta \epsilon s(r) \iff \begin{cases} r \sin\theta + r_o \sin\theta_o \cos\gamma_o > 0 & \text{if } \Phi(r,\theta) \geq \Phi(r_o,\theta_o) \\ r \sin\theta - r_o \sin\theta_o \cos\gamma_o > 0 & \text{if } \Phi(r,\theta) < \Phi(r_o,\theta_o) \end{cases}$$

so that

$$\inf_{\theta^+} f(r_o,\theta_o,\gamma_o;\ r,\theta) = \inf_{\theta\varepsilon s(r)} f(r_o,\theta_o,\gamma_o;\ r,\theta). \tag{A2.3}$$

The set s(r) can be expressed as

$$s(r) = s_1(r) \cup s_2(r) \tag{A2.4a}$$

where

$$s_1(r) = \{\theta \mid \theta\varepsilon[0,\pi],\ \Phi(r,\theta) \geq \Phi(r_o,\theta_o),\ r \sin\theta + r_o \sin\theta_o \cos\gamma_o > 0\} \tag{A2.4b}$$

and

$$s_2(r) = \{\theta \mid \theta\varepsilon[0,\pi],\ \Phi(r,\theta) < \Phi(r_o,\theta_o),\ r \sin\theta - r_o \sin\theta_o \cos\gamma_o > 0\}. \tag{A2.4c}$$

The derivative of f with respect to θ is easily calculated and is listed below for future reference.

$$\frac{\partial f}{\partial\theta}(r_o,\theta_o,\gamma_o;r,\theta) = \frac{r^2 B_r(r,\theta)\sin\theta}{r\sin\theta + r_o\sin\theta_o\cos\gamma_o} - \frac{[\Phi(r,\theta) - \Phi(r_o,\theta_o)]r\cos\theta}{2\pi(r\sin\theta + r_o\sin\theta_o\cos\gamma_o)^2} \quad \text{if } \Phi(r,\theta) > \Phi(r_o,\theta_o) \tag{A2.5a}$$

$$\frac{\partial f}{\partial\theta}(r_o,\theta_o,\gamma_o;r,\theta) = \frac{-r^2 B_r(r,\theta)\sin\theta}{r\sin\theta - r_o\sin\theta_o\cos\gamma_o} - \frac{[\Phi(r_o,\theta_o) - \Phi(r,\theta)]r\cos\theta}{2\pi(r\sin\theta - r_o\sin\theta_o\cos\gamma_o)^2} \quad \text{if } \Phi(r,\theta) < \Phi(r_o,\theta_o) \tag{A2.5b}$$

which can also be expressed as

$$\frac{\partial f}{\partial \theta}(r_o,\theta_o,\gamma_o;4,\theta) = \begin{cases} \dfrac{-r[f(r_o,\theta_o,\gamma_o;r,\theta)\cos\theta - rB_r(r,\theta)\sin\theta]}{r\sin\theta + r_o\sin\theta_o\cos\gamma_o} & \text{if } \Phi(r,\theta) > \Phi(r_o,\theta_o) \\[2ex] \dfrac{-r[f(r_o,\theta_o,\gamma_o;r,\theta)\cos\theta + rB_r(r,\theta)\sin\theta]}{r\sin\theta - r_o\sin\theta_o\cos\gamma_o} & \text{if } \Phi(r,\theta) < \Phi(r_o,\theta_o) \end{cases} \tag{A2.6}$$

We will now prove several lemmas. The first lemma states that if there exists a θ' which is an accumulation point of $s(r)$ (this is a point that can be approached arbitrarily closely by elements of $s(r)$) or is an element of $s(r)$ and has the property that $\Phi(r,\theta') = \Phi(r_o,\theta_o)$, then the greatest lower bound in (A2.3) is zero. This may at first seem obvious from inspection of (A2.2) but it becomes less obvious if we recognize the possibility of indeterminate forms. The second lemma describes the geometric structure of the set $s(r)$ when $r > r_o$ and the greatest lower bound in (A2.3) is not zero. The third and fourth lemmas will show that the greatest lower bound and least upper bound in (A2.1) are a relative minimum and a relative maximum, respectively. In other words, F corresponds to a saddle point in f. This not only verifies that (29) is equivalent to (A2.1), it also shows (after differentiability has been verified) that the extremums are found by setting the obvious derivatives equal to zero. This latter result will be important in the proof that the cutoff satisfies (2).

<u>Lemma 1</u>

Let r_o,θ_o,γ_o be arbitrary coordinates and let $r > r_o$ and let the magnetic field be defined and continuous everywhere. Assume there exists a $\theta' \varepsilon [o,\pi]$ which is either an accumulation point of $s(r)$ or an element of $s(r)$ and that satisfies

$$\Phi(r,\theta') = \Phi(r_o,\theta_o). \tag{A2.7}$$

Then

$$\inf_{\theta\varepsilon s(r)} f(r_o,\theta_o,\gamma_o;\ r,\theta) = 0. \tag{A2.8}$$

Proof

Inspection of (A2.2) makes (A2.8) obvious if neither $r \sin \theta' + r_o \sin \theta_o \cos \gamma_o$ nor $r \sin \theta' - r_o \sin \theta_o \cos \gamma_o$ are zero. This will be the case if θ' is an element of $s(r)$. But since the hypothesis allows for θ' to be an accumulation point of $s(r)$ and not necessarily an element of $s(r)$, it is possible that one or both expressions are zero.

First assume that both expresions are zero, i.e.,

$$r \sin \theta' + r_o \sin \theta_o \cos \gamma_o = 0 \tag{A2.9}$$

and

$$r \sin \theta' - r_o \sin \theta_o \cos \gamma_o = 0. \tag{A2.10}$$

This implies that $r_o \sin \theta_o \cos \gamma_o = 0$ and that $\theta' = 0$ or π. This implies that $\Phi(r,\theta') = 0$ and from (A2.7) we have $\Phi(r_o,\theta_o) = 0$. Equation (A2.2) becomes

$$f(r_o,\theta_o,\gamma_o;r,\theta) = \frac{|\Phi(r,\theta)|}{2\pi r \sin \theta}$$

and (A2.4) gives $s(r) = (0,\pi)$. So in this case we have

$$\inf_{\theta\varepsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) = \inf_{\theta\varepsilon(0,\pi)} \frac{|\Phi(r,\theta)|}{2\pi r \sin \theta} .$$

Using L'Hôpital's Rule together with

$$\frac{\partial\Phi}{\partial\theta} (r,\theta) = 2\pi r^2 B_r(r,\theta) \sin \theta$$

shows that $f(r_o,\theta_o,\gamma_o;r,\theta) \to 0$ as $\theta \to 0$ which proves (A2.8) for the special case that (A2.9) and (A2.10) are both satisfied.

Now assume that (A2.9) is satisfied but (A2.10) is not satisfied. This implies that $r \sin \theta' \neq 0$. Since the sine is nonnegative on $(0,\pi)$ we have the strict inequality

$$r \sin \theta' > 0. \tag{A2.11}$$

But (A2.9) and (A2.11) imply $r_o \sin \theta_o \cos \gamma_o < 0$ which implies

$$r \sin \theta' - r_o \sin \theta_o \cos \gamma_o > 0 . \tag{A2.12}$$

There are two possibilities to consider. The first is that there exists an $\xi > 0$ such that either

$$\theta'' \epsilon (\theta', \theta' + \xi) \Rightarrow \Phi(r,\theta'') < \Phi(r_o,\theta_o)$$

or

$$\theta'' \epsilon (\theta' - \xi, \theta') \Rightarrow \Phi(r,\theta'') < \Phi(r_o,\theta_o).$$

Assuming that such an ξ exists, we can chose it to be sufficiently small such that we also have

$$r \sin \theta'' - r_o \sin \theta_o \cos \gamma_o > 0.$$

Then there exists a sequence, $\{\theta''_n\}$, in $s_2(r)$ (Equation A2.4) that converges to θ'. In this sequence we have (using (A2.7) and (A2.12))

$$\lim_{n\to\infty} \frac{\Phi(r,\theta''_n) - \Phi(r_o,\theta_o)}{r \sin \theta''_n - r_o \sin \theta_o \cos \gamma_o} = \frac{\Phi(r,\theta') - \Phi(r_o,\theta_o)}{r \sin \theta' - r_o \sin \theta_o \cos \gamma_o} = 0.$$

This proves (A2.8) when such an ξ exists. Now suppose that no such ξ exists. Then there exists a sufficiently small $\xi > 0$ such that

$$\theta''\varepsilon(\theta' - \xi, \theta' + \xi) \Rightarrow \Phi(r,\theta'') \geq \Phi(r_o,\theta_o). \tag{A2.13}$$

But we have the equality when $\theta'' = \theta'$ which implies that $\Phi(r_o,\theta_o)$ is a relative minimum of $\Phi(r,\theta)$ on the ξ neighborhood of θ'. This implies

$$\frac{\partial\Phi}{\partial\theta}(r,\theta)\Bigg|_{\theta = \theta'} = 0 \tag{A2.14}$$

Also, from (A2.13), there is a sequence $\{\theta''_n\}$ in s_1 that converges to θ'. On this sequence we have

$$\lim_{n\to\infty} f(r_o,\theta_o,\gamma_o;r,\theta''_n) = \lim_{\theta\to\theta'} \frac{\Phi(r,\theta) - \Phi(r_o,\theta_o)}{2\pi(r \sin\theta + r_o \sin\theta_o \cos\gamma_o)} .$$

To evaluate this limit, note that it is an indeterminate form so use L'Hôpital's Rule to get

$$\lim_{n\to\infty} f(r_o,\theta_o,\gamma_o;r,\theta''_n) = \lim_{\theta\to\theta'} \frac{\frac{\partial\Phi}{\partial\theta}(r,\theta)}{2\pi(r\cos\theta)} .$$

Note that (A2.9) together with the condition $r > r_o$ implies that $\theta' \neq \pi/2$. Therefore, from (A2.14), the limit is zero. This proves (A2.8) when (A2.9) is satisfied.

Now assume that (A2.10) is satisfied but (A2.9) is not satisfied. Then $r \sin\theta' \neq 0$ which implies $r \sin\theta' > 0$ which implies $r_o \sin\theta_o \cos\gamma_o > 0$ and this implies $r \sin\theta' + r_o \sin\theta_o \cos\gamma_o > 0$. Therefore, θ' is an element of $s_1(r)$. Obviously, $f(r_o,\theta_o,\gamma_o;r,\theta') = 0$ which proves (A2.8). This completes the proof.

Lemma 2

Let r_o, θ_o, γ_o be arbitrary coordinates and let $r > r_o$. Let the magnetic field be defined and continuous everywhere. Assume

$$\inf_{\theta \varepsilon s(r)} f(r_o, \theta_o, \gamma_o; r, \theta) > 0 . \tag{A2.15}$$

Then we have the following implications.

If:

$$r_o \sin \theta_o \cos \gamma_o = 0 \tag{A2.16}$$

Then:

$$s(r) = (0, \pi) \tag{A2.17}$$

If:

$$r_o \sin \theta_o \cos \gamma_o > 0 \text{ and there exists a } \theta \varepsilon [0, \pi] \text{ such that } \Phi(r, \theta) \geq \Phi(r_o, \theta_o) \tag{A2.18}$$

Then:

$$\Phi(r, \theta) > \Phi(r_o, \theta_o) \text{ for all } \theta \varepsilon [0, \pi] \tag{A2.19}$$

and

$$s(r) = [0, \pi] \tag{A2.20}$$

If:

$$r_o \sin \theta_o \cos \gamma_o > 0 \text{ and there exists a } \theta \varepsilon [0, \pi] \text{ such that } \Phi(r, \theta) \leq \Phi(r_o, \theta_o) \tag{A2.21}$$

Then:

$$\Phi(r,\theta) < \Phi(r_o,\theta_o) \text{ for all } \theta\varepsilon[0,\pi] \tag{A2.22}$$

and

$$s(r) = (a,\pi-a) \tag{A2.23}$$

where a is the solution to $r \sin a = r_o \sin \theta_o \cos \gamma_o$ (A2.24)

If:

$r_o \sin \theta_o \cos \gamma_o < 0$ and there exists a $\theta\varepsilon[0,\pi]$ such that $\Phi(r,\theta) < \Phi(r_o,\theta_o)$ (A2.25)

Then:

$$\Phi(r,\theta) < \Phi(r_o,\theta_o) \text{ for all } \theta\varepsilon[0,\pi] \tag{A2.26}$$

and

$$s(r) = [0,\pi] \tag{A2.27}$$

If:

$r_o \sin \theta_o \cos \gamma_o < 0$ and there exists a $\theta\varepsilon[0,\pi]$ such that $\Phi(r,\theta) \geq \Phi(r_o,\theta_o)$ (A2.28)

Then:

$$\Phi(r,\theta) \geq \Phi(r_o,\theta_o) \text{ for all } \theta\varepsilon[0,\pi] \tag{A2.29}$$

and

$$s(r) = (a,\pi-a) \tag{A2.30}$$

where a is the solution to $r \sin a = -r_o \sin \theta_o \cos \gamma_o$ (A2.31)

<u>Proof</u>

That (A2.17) follows from (A2.16) is obvious from inspection of (A2.4). Now assume (A2.18). Since $r_o \sin \theta_o \cos \gamma_o > 0$ we have

$$\{\theta \mid \theta \in [0,\pi],\ r \sin \theta + r_o \sin \theta_o \cos \gamma_o > 0\} = [0,\pi]$$

and

$$\{\theta \mid \theta \in [0,\pi],\ r \sin \theta - r_o \sin \theta_o \cos \gamma_o > 0\} = (a,\pi-a)$$

where "a" is given by (A2.24). $s_1(r)$ and $s_2(r)$ (Eq. A2.4) become

$$s_1(r) = \{\theta \mid \theta \in [0,\pi],\ \Phi(r,\theta) \geq \Phi(r_o,\theta_o)\} \tag{A2.32}$$

$$s_2(r) = \{\theta \mid \theta \in (a,\pi-a),\ \Phi(r,\theta) < \Phi(r_o,\theta_o)\} \tag{A2.33}$$

By hypothesis, there exists a $\theta' \in [0,\pi]$ such that $\Phi(r,\theta') \geq \Phi(r_o,r_o)$. We will show that $\Phi(r,\theta) > \Phi(r_o,\theta_o)$ for all $\theta \in [0,\pi]$ by contradiction. Assume there exists a $\theta'' \in [0,\pi]$ such that $\Phi(r,\theta'') \leq \Phi(r_o,\theta_o)$. But $\Phi(r,\theta)$ is continuous so the conditions $\Phi(r,\theta') \geq \Phi(r_o,\theta_o)$ and $\Phi(r,\theta'') \leq \Phi(r_o,\theta_o)$ imply that there exists a $\theta''' \in [0,\pi]$ such that $\Phi(r,\theta''') = \Phi(r_o,\theta_o)$. This implies that $\theta''' \in s_1(r)$ which contradicts (A2.15) and Lemma 1. This proves (A2.19). But (A2.19) implies that $s_1(r) = [0,\pi]$ which proves (A2.20).

Now assume (A2.21). As in the previous case, $s_1(r)$ and $s_2(r)$ are given by (A2.32) and (A2.33) with "a" the solution to (A2.24). By hypothesis there exists a $\theta \in [0,\pi]$ such that $\Phi(r,\theta) \leq \Phi(r_o,\theta_o)$. But this violates the conclusion (A2.19) and therefore the hypothesis (A2.18) must be false, i.e., there is no

$\theta\varepsilon[0,\pi]$ such that $\Phi(r,\theta) \geq \Phi(r_o,\theta_o)$. This proves (A2.22). But (A2.22) implies that $s_1(r)$ is empty and $s_2(r) = (a,\pi-a)$. This proves (A2.23).

Now assume (A2.25). The same steps that produced (A2.32) and (A2.33) will now produce

$$s_1(r) = \{\theta \mid \theta\ \varepsilon(a,\pi-a),\ \Phi(r,\theta) \geq \Phi(r_o,\theta_o)\} \tag{A2.34}$$

$$s_2(r) = \{\theta \mid \theta\ \varepsilon[0,\pi],\ \Phi(r,\theta) < \Phi(r_o,\theta_o)\} \tag{A2.35}$$

where "a" is the solution to (A2.31). By hypothesis there exists a $\theta'\varepsilon[0,\pi]$ such that $\Phi(r,\theta') < \Phi(r_o,\theta_o)$. We will show that $\Phi(r,\theta) < \Phi(r_o,\theta_o)$ for all $\theta\varepsilon[0,\pi]$ by contradiction. Assume there exists a $\theta''\varepsilon[0,\pi]$ such that $\Phi(r,\theta'') \geq \Phi(r_o,\theta_o)$. But $\Phi(r,\theta)$ is continuous so the conditions $\Phi(r,\theta') < \Phi(r_o,\theta_o)$ and $\Phi(r,\theta'') \geq \Phi(r_o,\theta_o)$ implies that there exists a θ''' with the following properties:

$$\Phi(r,\theta''') = \Phi(r_o,\theta_o) \tag{A2.36}$$

and

Every deleted ξ neighborhood of θ''' contains points such that
$$\Phi(r,\theta) < \Phi(r_o,\theta_o) \tag{A2.37}$$

From (A2.37) we have it that θ''' is an accumulation point of $s_2(r)$. Therefore (A2.36) and Lemma 1 produce a contradiction. This proves (A2.26). But (A2.26) implies that $s_2(r) = [0,\pi]$ which proves (A2.27).

Now assume (A2.28). But this violates the conclusion (A2.26) so the hypothesis (A2.25) must be false, i.e., there is no $\theta\varepsilon[0,\pi]$ such that $\Phi(r,\theta) < \Phi(r_o,\theta_o)$. This proves (A2.29). But (A2.29) implies that $s_1(r) = (a,\pi-a)$ and $s_2(r)$ is empty. This proves (A2.30).

Lemma 3

Let r_o, θ_o, γ_o be arbitrary coordinates and let $r > r_o$. Let the magnetic field be defined and continuous everywhere. Assume

$$\inf_{\theta \varepsilon s(r)} f(r_o, \theta_o, \gamma_o; r, \theta) > 0. \tag{A2.38}$$

Then there exists a $\theta^*(r) \varepsilon s(r)$ such that

$$\Phi(r, \theta^*(r)) \neq \Phi(r_o, \theta_o) \tag{A2.39}$$

$$\inf_{\theta \varepsilon s(r)} f(r_o, \theta_o, \gamma_o; r, \theta) = f(r_o, \theta_o, \gamma_o; r, \theta^*(r)) \tag{A2.40}$$

$$\frac{\partial f}{\partial \theta}(r_o, \theta_o, \gamma_o; r, \theta) \Bigg|_{\theta = \theta^*(r)} = 0 \tag{A2.41}$$

Proof

It suffices to show that there exists a $\theta^*(r) \varepsilon s(r)$ that satisfies (A2.40) and (A2.41) since (A2.39) follows from (A2.38) and Lemma 1.

First consider the special case where $r_o \sin \theta_o \cos \gamma_o = 0$. Then $s(r) = (0, \pi)$ and

$$f(r_o, \theta_o, \gamma_o; r, \theta) = \frac{|\Phi(r, \theta) - \Phi(r_o, \theta_o)|}{2\pi r \sin \theta}.$$

From (A2.38) and Lemma 1 we have it that $\Phi(r, 0) \neq \Phi(r_o, \theta_o)$ and $\Phi(r, \pi) \neq \Phi(r_o, \theta_o)$. Therefore f increases without bound as $\theta \to 0$ or $\theta \to \pi$. The greatest lower bound must be a relative minimum at an interior point in $(0, \pi)$ which proves the lemma for this special case.

Now assume

$$r_o \sin \theta_o \cos \gamma_o \neq 0.$$

Lemma 2 states that s(r) is an interval. If it can be shown that the greatest lower bound is not obtained by taking the limit as θ approaches an end point of the interval it will follow that the greatest lower bound is a relative minimum which occurs at some interior point in the interval. This will imply that there exists a $\theta^*(r)\epsilon s(r)$ satisfying (A2.40). It will also imply that $\theta^*(r)$ satisfies (A2.41) providing the derivative exists. Inspection of (A2.2) and the possible intervals listed in Lemma 2 shows that under each possible set of conditions ((A2.18), (A2.21), (A2.25) and (A2.28)), f is differentiable on the interior of the interval that applies to those conditions. Therefore, (A2.40) and (A2.41) can both be proven by showing that

If:

$$r_o \sin \theta_o \cos \gamma_o > 0 \text{ and } \Phi(r,\theta) > \Phi(r_o,\theta_o) \text{ for all } \theta\epsilon[0,\pi] \tag{A2.42}$$

Then:

$$\lim_{\theta\to 0} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\epsilon[0,\pi]} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.43}$$

and

$$\lim_{\theta\to \pi} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\epsilon[0,\pi]} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.44}$$

If:

$$r_o \sin \theta_o \cos \gamma_o > 0 \text{ and } \Phi(r,\theta) < \Phi(r_o,\theta_o) \text{ for all } \theta\epsilon[0,\pi] \tag{A2.45}$$

Then:

$$\lim_{\theta\to a} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\epsilon(a,\pi-a)} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.46}$$

and

$$\lim_{\theta\to\pi-a} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\varepsilon(a,\pi-a)} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.47}$$

where

$$r \sin a = r_o \sin \theta_o \cos \gamma_o \tag{A2.48}$$

If:

$$r_o \sin \theta_o \cos \gamma_o < 0 \text{ and } \Phi(r,\theta) < \Phi(r_o,\theta_o) \text{ for all } \theta\varepsilon[0,\pi] \tag{A2.49}$$

Then:

$$\lim_{\theta\to 0} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\varepsilon[0,\pi]} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.50}$$

and

$$\lim_{\theta\to\pi} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\varepsilon[0,\pi]} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.51}$$

If:

$$r_o \sin \theta_o \cos \gamma_o < 0 \text{ and } \Phi(r,\theta) \geq \Phi(r_o,\theta_o) \text{ for all } \theta\varepsilon[0,\pi] \tag{A2.52}$$

Then:

$$\lim_{\theta\to a} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\varepsilon(a,\pi-a)} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.53}$$

and

$$\lim_{\theta\to\pi-a} f(r_o,\theta_o,\gamma_o;r,\theta) > \inf_{\theta\varepsilon(a,\pi-a)} f(r_o,\theta_o,\gamma_o;r,\theta) \tag{A2.54}$$

where

$$r \sin \theta = -r_o \sin \theta_o \cos \gamma_o \tag{A2.55}$$

Inspection of (A2.2) shows that (A2.46) and (A2.47) follow from (A2.45) because the limits produce singularities. Similarly, (A2.53) and (A2.54) follow from (A2.52) (note that $\Phi(r,a) \neq \Phi(r_o,\theta_o)$ and $\Phi(r,\pi-a) \neq \Phi(r_o,\theta_o)$ because a and π-a are accumulation points of s(r)). Now assume either (A2.42) or (A2.49). Since $r_o \sin \theta_o \cos \gamma_o \neq 0$, then in either case the derivative of f (given by (A2.5)) is negative in the limit as $\theta \to 0$ and positive in the limit as $\theta \to \pi$. Therefore f is decreasing when θ is in a sufficiently small neighborhood of zero and increasing when θ is in a sufficiently small neighborhood of π. The least upper bound cannot occur in either of these limits which proves (A2.43), (A2.44), (A2.50) and (A2.51). This completes the proof.

Lemma 4

Let r_o,θ_o,γ_o be arbitrary coordinates. Let the magnetic field be differentiable in r,θ for all $r > 0$. Assume that $\Phi(r,\theta)$ is bounded as $r \to \infty$. Assume also that

$$\sup_{r>r_o} \inf_{\theta\epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) > 0. \tag{A2.56}$$

Then there exists a finite $r_o^* > r_o$ and a $\theta_o^* \epsilon s(r_o^*)$ that satisfy the following conditions:

$$\sup_{r>r_o} \inf_{\theta\epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) = f(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*) \tag{A2.57}$$

$$\frac{\partial f}{\partial \theta}(r_o,\theta_o,\gamma_o;r,\theta)\Bigg|_{r = r_o^*,\ \theta = \theta_o^*} = 0 \tag{A2.58}$$

$$\frac{\partial f}{\partial r}(r_o,\theta_o,\gamma_o;r,\theta)\Bigg|_{r = r_o^*,\ \theta = \theta_o^*} = 0 \tag{A2.59}$$

Proof

From (A2.56) we have it that there exists a set R such that

$$r \epsilon R \Rightarrow r > r_o \text{ and } \inf_{\theta \epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) > 0 \tag{A2.60}$$

and

$$\sup_{r>r_o} \inf_{\theta \epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) = \sup_{r \epsilon R} \inf_{\theta \epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta).$$

From Lemma 3 we have

$$\sup_{r>r_o} \inf_{\theta \epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) = \sup_{r \epsilon R} f(r_o,\theta_o,\gamma_o;r,\theta^*(r)) \tag{A2.61}$$

for some $\theta^*(r) \epsilon s(r)$ which satisfies (from (A2.41) and (A2.6))

$$f(r_o,\theta_o,\gamma_o;r,\theta^*(r)) \cos \theta^*(r) = rB_r(r,\theta^*(r)) \sin \theta^*(r) \quad \text{if } \Phi(r,\theta^*(r)) > \Phi(r_o,\theta_o) \tag{A2.62}$$

$$f(r_o,\theta_o,\gamma_o;r,\theta^*(r)) \cos \theta^*(r) = -rB_r(r,\theta^*(r)) \sin \theta^*(r) \quad \text{if } \Phi(r,\theta^*(r)) < \Phi(r_o,\theta_o). \tag{A2.63}$$

Note that one expression or the other applies by virtue of (A2.39). First assume that R is bounded above. Then for r sufficiently large we have

$$\inf_{\theta \epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) = f(r_o,\theta_o,\gamma_o;r,\theta^*(r)) = 0$$

so the least upper bounds in (A2.61) must occur at some finite r_o^*. Now assume that R is not bounded above. By hypothesis $\Phi(r,\theta)$ is bounded and this implies that $rB_r(r,\theta) \to 0$ as $r \to \infty$. From (A2.62) and (A2.63) we have it that $f(r_o,\theta_o,\gamma_o;r,\theta^*(r)) \to 0$ as $r \to \infty$ unless $\theta^*(r) \to \pi/2$. But from (A2.2) we see that $f(r_o,\theta_o,\gamma_o;r,\theta^*(r))$ still goes to zero as $r \to \infty$ even if $\theta^*(r)$ does approach $\pi/2$. Therefore, the least upper bound of $f(r_o,\theta_o,\gamma_o;r,\theta^*(r))$ must occur at some finite r_o^* in all cases. Note that we must have $r_o^* > r_o$ because if $r_o^* = r_o$, f can be made equal to zero by letting $\theta = \theta_o$ and this implies that $f(r_o,\theta_o,\gamma_o;r_o,\theta^*(r_o)) = 0$ which is clearly not the least upper bound in (A2.61). This proves that

$$\sup_{r>r_o} \inf_{\theta\epsilon s(r)} f(r_o,\theta_o,\gamma_o;r,\theta) = f(r_o,\theta_o,\gamma_o;r_o^*,\theta^*(r_o^*)) \tag{A2.64}$$

for some finite $r_o^* > r_o$. This proves (A2.57). Note that

$$f(r_o,\theta_o,\gamma_o;r_o^*,\theta^*(r_o^*)) = \inf_{\theta\epsilon s(r_o^*)} f(r_o,\theta_o,\gamma_o;r_o^*,\theta) \tag{A2.65}$$

Conditions (A2.56), (A2.64) and (A2.65) imply that the hypothesis to Lemma 3 is satisfied at $r = r_o^*$ so (A2.39) gives

$$\Phi(r_o^*, \theta^*(r_o^*)) \neq \Phi(r_o,\theta_o).$$

This means that for r in a sufficiently small neighborhood of r_o^*, $\Phi(r,\theta^*(r)) - \Phi(r_o,\theta_o)$ does not change sign, so $f(r_o,\theta_o,\gamma_o;r,\theta^*(r))$ is a well-behaved function of r in that neighborhood, i.e., f is differentiable at r_o^*. Since the least upper bound in (A2.61) is a relative maximum at an interior point r_o^* and f is differentiable there, the derivative is zero there, i.e.,

$$\left.\frac{\partial f}{\partial r}(r_o,\theta_o,\gamma_o;r,\theta^*(r))\right|_{r = r_o^*} = 0.$$

The chain rule gives

$$\frac{\partial f}{\partial r}(r_o,\theta_o,\gamma_o;r,\theta)\Bigg|_{\substack{r = r_o^* \\ \theta = \theta^*(r_o^*)}}$$

$$+ \frac{\partial f}{\partial \theta}(r_o,\theta_o,\gamma_o;r,\theta)\Bigg|_{\substack{r = r_o^* \\ \theta = \theta^*(r_o^*)}} \frac{\partial \theta^*(r)}{\partial r}\Bigg|_{r = r_o^*} = 0.$$

The above equation together with (A2.41) prove (A2.58) and (A2.59) where $\theta_o^* = \theta^*(r_o^*)$. This completes the proof.

We are now in a position to verify (2). Let a particle have initial coordinates r_o,θ_o,γ_o and let the particle move with a rigidity P equal to the cutoff $F(r_o,\theta_o,\gamma_o)$ which is assumed to be greater than zero. Lemma 4 states that the rigidity can be expressed as

$$P = f(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*).$$

Let the particle move a small distance $d\ell$ to the new coordinates $r_1 = r_o + dr$, $\theta_1 = \theta_o + d\theta$, $\gamma_1 = \gamma_0 + d\gamma$. To first order, the cutoff rigidity at the new coordinates is

$$f(r_1,\theta_1,\gamma_1;r_1^*,\theta_1^*) = f(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)$$

$$+ \frac{\partial f}{\partial r_o}(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)\,dr + \frac{\partial f}{\partial \theta_o}(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)\,d\theta$$

$$+ \frac{\partial f}{\partial \gamma_o}(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)\,d\gamma + \frac{\partial f}{\partial r_o^*}(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)\,dr^*$$

$$+ \frac{\partial f}{\partial \theta_o^*}(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)\,d\theta^* . \tag{A2.66}$$

To be definite, suppose that $\Phi(r_o^*,\theta_o^*) > \Phi(r_o,\theta_o)$. Analogous steps will reach the same conclusion if $\Phi(r_o^*,\theta_o^*) < \Phi(r_o,\theta_o)$ (note that the equality is ruled out by (A2.39) together with $\theta_o^* = \theta^*(r_o^*)$). Then we use the upper expression on the right side of (A2.2) when taking the derivatives and we obtain

$$\frac{\partial f}{\partial r_o} = -\frac{\dfrac{\partial\Phi(r_o,\theta_o)}{\partial r_o}}{2\pi(r_o^* \sin\theta_o^* + r_o \sin\theta_o \cos\gamma_o)} - \frac{[\Phi(r_o^*,\theta_o^*) - \Phi(r_o,\theta_o)] \sin\theta_o \cos\gamma_o}{2\pi(r_o^* \sin\theta_o^* + r_o \sin\theta_o \cos\gamma_o)^2}$$

$$\frac{\partial f}{\partial \theta_o} = -\frac{\dfrac{\partial\Phi(r_o,\theta_o)}{\partial \theta_o}}{2\pi(r_o^* \sin\theta_o^* + r_o \sin\theta_o \cos\gamma_o)} - \frac{[\Phi(r_o^*,\theta_o^*) - \Phi(r_o,\theta_o)] r_o \cos\theta_o \cos\gamma_o}{2\pi(r_o^* \sin\theta_o^* + r_o \sin\theta_o \cos\gamma_o)^2}$$

$$\frac{\partial f}{\partial \gamma_o} = \frac{[\Phi(r_o^*,\theta_o^*) - \Phi(r_o,\theta_o)]\, r_o \sin\theta_o \sin\gamma_o}{2\pi(r_o^* \sin\theta_o^* + r_o \sin\theta_o \cos\gamma_o)^2} .$$

The last two derivates on the right side of (A2.66) are zero because of Lemma 4. Expressing the terms that contain $\Phi(r_o^*,\theta_o^*) - \Phi(r_o,\theta_o)$ in terms of $f(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)$, which in turn is equal to P, and substituting into (A2.66) gives

$$f(r_1,\theta_1,\gamma_1;r_1^*,\theta_1^*) - f(r_o,\theta_o,\gamma_o;r_o^*,\theta_o^*)$$

$$= \frac{1}{r_o^* \sin\theta_o^* + r_o \sin\theta_o \cos\gamma_o} \Big\{ - \Big[\frac{1}{2\pi}\frac{\partial\Phi}{\partial r_o}(r_o,\theta_o) + P \sin\theta_o \cos\gamma_o\Big]\, dr$$

$$- \Big[\frac{1}{2\pi}\frac{\partial\Phi}{\partial\theta_o}(r_o,\theta_o) + P r_o \cos\theta_o \cos\gamma_o\Big]\, d\theta + P r_o \sin\theta_o \sin\gamma_o\, d\gamma\Big\}. \quad \text{(A2.67)}$$

Note that (21) gives

$$P r_o \sin\theta_o \cos\gamma_o + \frac{\Phi(r_o,\theta_o)}{2\pi} = \text{constant}.$$

Taking the differential of the above equation shows that the right side of (A2.67) is zero, i.e., the cutoff is constant on the trajectory.