LBL-23241

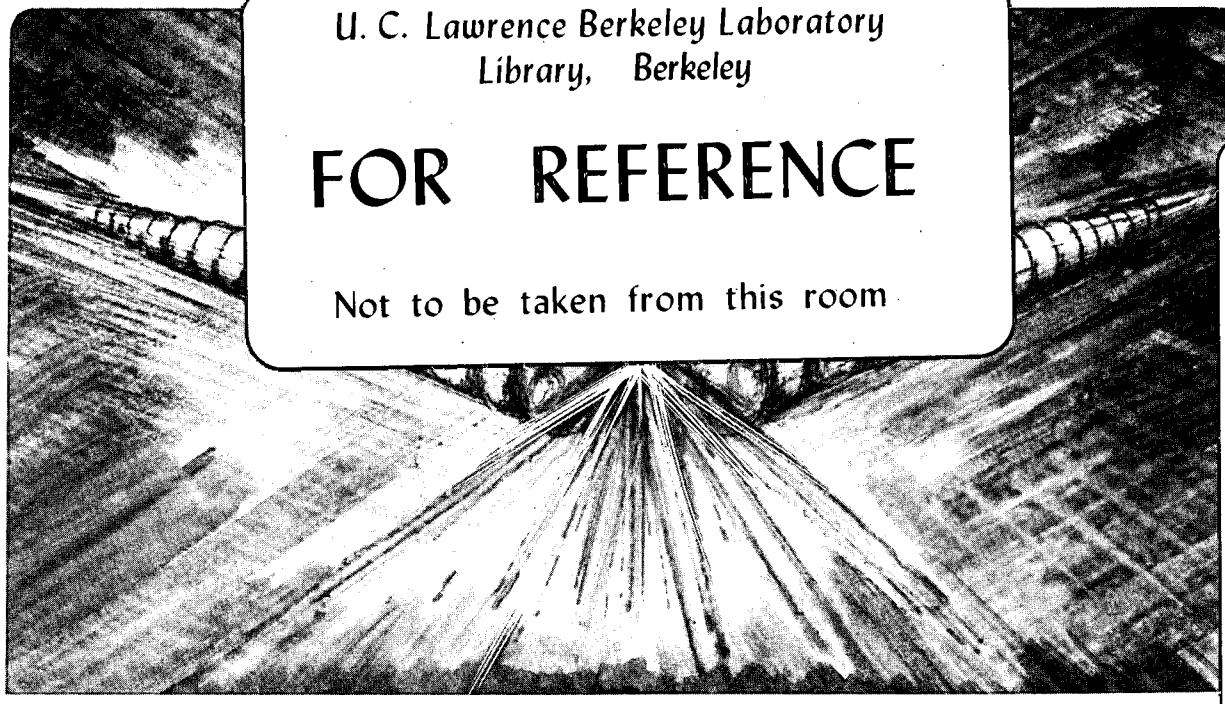
Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Accelerator & Fusion
Research Division

COVARIANT LAGRANGIAN METHODS OF RELATIVISTIC PLASMA THEORY

Bruce M. Boghosian
(Ph.D. Thesis)

February 1987



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

LBL-23241

COVARIANT LAGRANGIAN METHODS OF RELATIVISTIC PLASMA THEORY*

BY

Bruce Michael Boghosian

Lawrence Berkeley Laboratory
and
University of California
Berkeley, CA 94720

February 1987

* This work was supported by the U.S. Department of Energy under Contract Number DE-AC03-76SF00098.

Covariant Lagrangian Methods of Relativistic Plasma Theory

Bruce M. Boghosian
University of California at Davis
Department of Applied Science
Livermore, California 94550

April 27, 1987

ABSTRACT

The relativistic electromagnetic projection operators discovered by Fradkin are used to obtain a covariant decomposition of the motion of a relativistic charged particle into parallel motion and perpendicular gyration. The Lagrangian Lie transform method of Littlejohn is used to achieve a transformation to guiding-center coordinates in which the rapid oscillatory motion is removed. The method parallels the nonrelativistic guiding-center calculation of Littlejohn, and the four-vector notation used throughout facilitates this comparison. The natural guiding-center Poisson bracket structure and Hamiltonian are derived. The guiding-center equations of motion are presented to one order higher than the usual drifts, and the correction to the gyromomentum is given. Correspondence with the usual noncovariant results, as given by Northrop, is demonstrated.

It is possible to add one or more eikonal wave perturbations to the Lagrangian action for a single particle before performing the guiding-center transformation. It is shown that such perturbations can be written in a manifestly gauge-invariant form in guiding-center coordinates; this observation allows us to develop an oscillation-center theory to arbitrarily high order and be guaranteed of manifest gauge invariance at every step of the way. In this way, once again using Lagrangian Lie transforms, we obtain the ponderomotive Hamiltonian.

By summing the guiding-center Lagrangian action over the full distribution of guiding centers present in a plasma and adding the action of the Maxwell field, one obtains the total action of a guiding-center plasma. Upon variation of the total action, we find a self-consistent set of covariant relativistic kinetic and field equations; from these we can identify the guiding-center current density and the guiding-center magnetization. Upon application of Noether's theorem, the total action yields covariant conservation laws for the momentum-energy and the angular momentum of a relativistic guiding-center plasma; from these we can identify the guiding-center stress-energy tensor and the guiding-center spin angular momentum tensor.

If we sum the Lagrangian action for a guiding/oscillation center over the full distribution and add the action of the Maxwell field, then variation yields self-consistent relativistic kinetic and field equations for the plasma in the wave field, including the dispersion relation for the wave; from these we can identify the wave magnetization and the susceptibility, and thereby demonstrate the K-χ theorem. If we then apply Noether's theorem, we get conservation laws for the guiding-center plasma in the presence of a wave field; from these we can identify the wave contribution to the stress-energy and spin angular momentum tensors.

To my family.

Acknowledgements

This thesis is the culmination of a six-year PhD program at the University of California at Davis. During this period, I worked full-time as a plasma physicist in the Magnetic Fusion Energy Division of the Lawrence Livermore National Laboratory. Much of my research was carried out in collaboration with the Plasma Theory group at the Lawrence Berkeley Laboratory. All of this gives rise to a long list of people to thank.

We'll start with the personal one...

First, and foremost, I would like to thank my wife, Laura, for her love, patience and support during the course of my research. Full-time employment plus thesis research is a stressful combination, and I couldn't have done it without her.

Then comes the thesis committee...

More than anyone else, my thesis committee chairman, William A. Newcomb, taught me how to think. I consider myself privileged to have had the benefit of Bill's insights on a variety of topics (from physics, to mathematics, to philosophy, to jogging). The precision of his thinking and the clarity of his explanations will remain a role model for me throughout my career.

Most all of the research in this thesis was carried out with Allan Kaufman and his Plasma Theory group at the Lawrence Berkeley Laboratory. Allan has

been the principal pioneer of the action principle formulation of plasma physics, which is the main underlying theme of this thesis. He has the distinction of being not only the founder of a new school of thought within plasma physics that emphasizes a differential geometric approach to the classical problems of the field, but also of being the founder of a new school of plasma physicists among whom I am proud to be included. Allan's Socratic style of teaching was indispensible in helping me think my way through the maze of concepts encountered in the course of my research. He also taught me the importance of a broadened perspective in physics research. As a result, I became familiar with many areas of physics and mathematics to which I would not have otherwise been exposed, and I was able to apply much of that knowledge in this thesis.

Gary Smith has been a source of learning and guidance for me from the very beginning of my program of study. When one first enters a field as complex as plasma physics, it is indispensible to have someone who is willing to act as a tutor, answering the myriad questions (some stupid, some not-so-stupid) that arise during the learning process. Gary has never failed me in this regard. His uncanny intuition about plasma waves and stability theory has shaped the way I think about the whole field. He always found time to discuss things with me, and was always genuinely concerned about my welfare. Wherever our paths may take us, I know that he will always be my friend.

Next, all the folks at Livermore...

I would like to thank my supervisors at the Lawrence Livermore National Laboratory, Gus Carlson from 1978 to 1982 and L. Donald Pearlstein from 1982 to the present, for generously allowing me the time I needed to complete this thesis.

The entire Mirror Theory and Computations group at Livermore has been a

great source of help, but the following names stand out in particular:

Steven Auerbach for innumerable discussions about everything under the sun (plus some things inside the sun) including lots of stuff in this thesis. It is a pleasure to have a friend who appreciates the love of learning and the joy of figuring things out as much as Steve does.

Bruce Cohen for his concern and guidance during several critical stages of this thesis, and especially for his many excellent suggestions for the application of this work.

Brendan McNamara for teaching me much about the application of Lie transforms to mode-coupling analyses in weak turbulence theory in general, and in the Bénard convection problem in particular.

Dana Richards and Jan Wikkerink for their continued violations of the second law of thermodynamics, without which everything would have fallen into utter disorganization and chaos.

Next, all the folks in Berkeley...

Robert Littlejohn for millions of wonderful discussions on guiding-center theory and differential geometry. Robert's influence permeates this thesis from begining to end. In many ways, this thesis is the relativistic generalization of the revolution in plasma physics that was initiated by him.

Jerrold Marsden for numerous enlightening discussions about reduction, gauge groups, nonlinear stability theory, etc. His course on the theory and applications of symplectic and Poisson reduction has molded the way I think about those topics.

Richard Montgomery for lots of helpful discussions about symplectic geometry, the method of averaging, reduction, and the energy-Casimir method.

Stephen Omohundro for teaching me the importance of geometrical intuition.

Many was the blackboard discussion at which Steve would label something "obvious" that became clear to me only after much calculation. It did not take many such discussions to convince me of the importance of differential geometrical intuition. With Steve's help, I devoured books on the subject, and my entire outlook in mathematical physics has not been the same since. The prevalence of such geometrical techniques in this thesis attests to their influence on my thinking.

Philippe Similon for guidance and helpful suggestions on just about everything in this thesis, especially the Poisson brackets and conservation laws for the guiding-center plasma. Philippe has the rare combination of abilities required to understand and use formalism, while never losing sight of the fact that its purpose is to solve real problems in the real world.

Jonathan Wurtele for many discussions about nonlinear dynamics and differential geometric techniques in physics.

Huanchun Ye for many helpful discussions, especially on the problem of resonances. This problem was largely ignored in my thesis, but Huanchun's thesis should fill in the gap and make the theory more complete. His observations on the relationship between torsion and spin seem to underlie the theory leading to the conservation law for angular momentum in the guiding-center plasma. Once again, it is quite possible that there is much left to be discovered in this area, and it may be in Huanchun's thesis.

Next, the folks at the University of California at Davis Department of Applied Science...

My oral exam committee which consisted of G. Donald Chakerian, Abraham Goldberg, Myron (Mike) A. Hoffman, John Killeen and Frederick Wooten.

Abe Goldberg for teaching me quantum mechanics, and for lots of good advice, discussions and vanpool conversation on everything from Bell's theorem to

quantum field theory.

John Killeen for teaching me much advanced mathematics, and for several enlightening discussions on computational plasma physics.

Fred Wooten for much guidance and advice about how to survive a PhD program.

William Hoover for teaching me statistical mechanics and for several good discussions about nonlinear dynamics.

Robert White for serving as a sounding board for zillions of ideas, some of which appeared in this thesis and some of which did not. As the only other "plasma type" in my class at D.A.S. with an interest in nonlinear dynamics, Robert and I consulted with each other a great deal.

Donna Clifford for making sure I didn't run afoul of the bureaucracy. I'm not even sure I'd be a registered student now were it not for her (not to mention the innumerable late fees that only her reminders saved me from having to pay).

Last, but certainly not least, some folks who don't fit neatly into any of the above categories...

Bedros Afeyan for numerous interesting discussions while he was at Livermore.

John Cary for his insights on ponderomotive theory and the K-χ theorem.

John David Crawford for helping me better understand Van Kampen modes, and for several other enlightening discussions.

Charles Karney and Bruce Nemnich for helping me with LaTeX.

This work was supported by the US DOE under contract numbers W-7405-ENG-48 and DE-AC03-76SF00098.

Contents

Chapter 1

Introduction

There was a time when a thorough working knowledge of geometry was considered an indispensible ingredient in the education of a natural philosopher. From Euclid's first systematization of the subject more than two thousand years ago to well after the end of the Renaissance, the study of the *Elements* was considered a critically important part of mathematical instruction. Indeed, when reading Newton's *Principia* or *Opticks*, one is struck by the prevalence of geometrical arguments and descriptions.

Alas, the introduction of coordinate systems by Descartes and the concommitant analyticization of geometry changed all this. Using coordinates, geometrical problems could be reduced to algebraic problems. The perceived need for good geometrical intuition gradually disappeared. By the time Whittaker's *Treatise on the Analytical Dynamics of Particles and Rigid Bodies* was first published in 1904, this attitude had taken hold to the extent that Whittaker apparently felt no need to include illustrations in his nearly five-hundred-page-long (and otherwise excellent) document. At present, one can obtain an undergraduate degree in

physics or even applied mathematics with little more geometry background than is found in a secondary school textbook.

That this trend is disastrous has been appreciated only for the past couple of decades. This appreciation has been due, in large part, to modern developments in the general theory of relativity. The entire lesson of relativity theory is that physical laws ought not to depend upon the coordinate system chosen to describe them; that is, the meaning of physical laws transcends their coordinate description. Conversely, coordinate descriptions can have a way of masking fundamental physical reality. Thus, a coordinate-free description of physical laws can have the beneficial effect of allowing one more easily to glimpse the underlying fundamental physical reality. Such coordinate-free mathematical language is available, thanks in large part to the works of Cartan and Lie. Modern differential geometry, including the exterior calculus and the theory of Lie groups, is capable of providing a coordinate-free description of physical law. Please note that what is being argued here is that such a coordinate-free description is far more than just an alternative mathematical notation; the contention is that it yields an improved understanding of the *physics* involved. A physicist who takes the time to learn how, say, electromagnetic theory can be described in terms of differential forms will have, as a result, an improved understanding of the electromagnetic field.

There is an additional benefit to the geometrical point of view. Just as Descartes found that algebra can be used as a tool for obtaining geometrical results, likewise geometry can be used as a tool for obtaining analytical results that would be far more difficult to obtain any other way. Several examples of this phenomenon will be pointed out in the course of this thesis.

Since the 1960's it has been known that classical mechanics is describable in terms of symplectic geometry. This observation paved the way for powerful generalizations of some of the traditional methodologies of mechanics. For ex-

ample, whereas Hamiltonian mechanics had been originally formulated in terms of canonically conjugate pairs of coordinates, it was found that noncanonical coordinates could be used instead, oftentimes to great advantage. Powerful new types of perturbation theory, based on Lie transforms, were introduced; this made higher-order perturbative treatments less laborious and more systematic.

Nowhere was the impact of this revolution more profound and beneficial than in the field of plasma physics. Because the motion of charged particles in complicated electromagnetic geometries and in wave fields requires a perturbative treatment, it is not surprising that Lie transform perturbation theory was shown to be a natural tool for systematizing, simplifying and *better understanding* many of the calculations of plasma physics. Furthermore, it was shown that the most natural treatment of the guiding-center problem (i.e. the ubiquitous problem of computing the drifts of a charged particle gyrating in a slowly-varying electromagnetic field) involved the use of noncanonical coordinates and noncanonical coordinate transformations. All of this will become more clear as we proceed.

During the late nineteen seventies, Dewar [1] introduced the idea of canonical oscillation-center transformations. Johnston and Kaufman [2] and Johnston [3] used canonical perturbation theory to perform oscillation-center and mode coupling analyses for the Vlasov plasma. In Cary's PhD thesis [4], Lie transforms were shown to be a useful tool for ponderomotive theory, and the K-χ theorem [5] relating the ponderomotive Hamiltonian with the linear susceptibility was formulated.

The extension of these techniques to magnetized plasma was made possible, or at least greatly facilitated, by Littlejohn's work on the guiding-center problem in his PhD thesis [6]. Littlejohn made the key observation that the transformation from single-particle to guiding-center coordinates was best done using noncanonical methods. This noncanonical transformation was done in his thesis by using the Darboux theorem constructively, and it was followed by a canonical

Lie transformation that averaged over the rapid gyromotion. Subsequently, Littlejohn [7] discovered that the entire transformation could be done by a single Lie transform with a vector generator. This is the approach followed in this thesis.

Ponderomotive theory for a magnetized relativistic plasma was then done by Grebogi and Littlejohn [8], who used canonical Lie transforms. They pointed out that the oscillation-center transformation for a magnetized plasma might best be handled by noncanonical Lie methods, but they did not do it this way. Their result was subsequently simplified by Cary and Newberger [9].

Meanwhile, Dubin, Krommes, Oberman and Lee [10] showed how to use Littlejohn's methods to derive self-consistent gyrokinetic equations for an electrostatic plasma, including the Poisson equation whose source term was written in terms of the guiding-center distribution function. Kaufman and Boghosian [11] showed that this calculation could be done by summing the guiding-center action over the entire distribution and coupling it to the Maxwell action; variation with respect to the coordinate fields (considered to be functions of their initial conditions) then yields the gyrokinetic equation, and variation with respect to the vector potential then yields the self-consistent field equaton. Finally, Similon [12] showed that conservation laws for the guiding-center plasma could be obtained by application of Noether's theorem to this system action.

The above-mentioned work by Grebogi and Littlejohn was done for a relativistic plasma, but was not manifestly covariant in that it was done in "1 + 3" notation. A manifestly covariant treatment is made possible with the help of certain projection operators which were introduced by Fradkin [13] who obtained the drifts for a relativistic guiding center (but did not use Lie methods), and by Dumais [14].

The general plan of this thesis is as follows:

Chapter 2 will cover the mathematical preliminaries necessary to understand the differential geometric arguments used in this thesis. It should be emphasized

that this constitutes no more than a sketchy introduction, and is no substitute for a good text on the subject; nevertheless it is probably sufficient to enable a persistent person with an undergraduate background in physics to read and understand this entire text. Chapter 2 also describes the application of these techniques to Hamiltonian and Lagrangian mechanics; specifically, Lie transform perturbation theory is introduced here and many simple examples of its use are presented.

Chapter 3 will treat the guiding-center problem for a relativistic charged particle. We shall begin by examining the geometry of the electromagnetic field in four-dimensional spacetime, and we shall find that there is a covariant way to isolate the rapidly-gyrating component of the particle's four velocity. Lie transform perturbation theory is then applied to the particle's phase-space Lagrangian in order to remove this rapidly-gyrating component and thus obtain the residual parallel and drift motion. The perturbative calculation is carried out to one order higher than the usual drifts, the natural guiding-center Poisson bracket structure and Hamiltonian are presented, and the correction to the gyromomentum is given. Finally, it is shown how to cast these results in a *manifestly* gyrogauge invariant format.

In Chapter 4 we shall study the effects of eikonal wave perturbations on a guiding center, once again using Lie transform perturbation theory. The result is a complete ponderomotive description of the relativistic guiding center in an eikonal wave field, and we show how to cast this in *manifestly* gauge-invariant form. To achieve manifest gauge-invariance, we shall find it necessary to abandon the usual approach of expanding the eikonal wave perturbation in a series of Bessel functions of $k_\perp \rho$. Instead, we shall first perform a Lagrangian gauge transformation, and then we shall expand in a series of special functions that are related to indefinite integrals of Bessel functions. The required Lagrangian gauge transformation is not obvious, and it would never have been discovered

without the use of differential geometric techniques. Finally, the ponderomotive Hamiltonian is derived using Lie transforms.

In Chapter 5 we shall sum the resulting guiding-center Lagrangian over the entire distribution of particles present in a plasma, and couple with the Maxwell field to obtain the total Lagrangian for a Vlasov plasma of relativistic guiding centers. By varying this it is possible to derive a self-consistent gyrokinetic description of such a plasma, including the magnetic moment tensor, in manifestly-covariant format. Application of Noether's theorem then yields conservation laws for the guiding-center plasma, and these are also cast in manifestly covariant form. Finally, using the results of Chapter 4, the conservation laws are derived for a guiding-center plasma in the presence of a wave field.

In Chapter 6 we discuss some of the unanswered questions raised by this study. These could be topics for future research.

Appendix A is a glossary of the mathematical symbols and notation used in this thesis.

Appendix B is a review of some of the more primitive mathematical concepts used in this thesis, such as *vector spaces*, *dual spaces*, *algebras*, and *modules*.

Appendix C applies vector Lie transforms to the nonrelativistic guiding-center problem in two dimensions, and derives the shift in gyrofrequency due to spatial gradients in the magnetic and (perpendicular) electric fields. This is useful both as a demonstration of the vector Lie transform technique, and as a comparison to the techniques and results of Chapter 3.

Appendix D derives and discusses the properties of a pair of special functions that were introduced in Chapter 4.

Appendix E is a short tutorial on how to derive Bessel function sum rules, including (but not limited to) those that were useful in Chapter 4.

Chapter 2

Mathematical Preliminaries

2.1 Discussion

This chapter divides naturally into three sections. The first covers the basic results of differential geometry that are necessary to understand the rest of this thesis. This includes the calculus of tensors and the exterior algebra. To reiterate, the exposition here is not intended to replace a good introductory book on the subject (see, for example, the excellent introductory texts by Schutz [15], Edelen [16], Singer and Thorpe [17], or Burke [18]), but it does present enough material to make the thesis self-contained, and to establish notational conventions. The theory of Lie groups has been omitted from this section because it is not absolutely essential to the understanding of what follows, but the reader with background in this area will be at a definite advantage.

Next, these tools are used to reformulate Hamiltonian and Lagrangian mechanics. The generalization to noncanonical coordinates is discussed, including those with singular Poisson structures. Noether's theorem is formulated, and numerous worked examples are given. Mechanical systems with constraints are

examined from this new point of view.

Finally, Lie transform perturbation theory is presented, and its use for non-canonical coordinates is discussed. Because we shall use Lie transforms in a more general context than that in which they are usually presented, I recommend that this section be read even by those already familiar with the subject.

2.2 Differential Geometric Concepts

2.2.1 Manifolds, Vectors, and Covectors

In this subsection, we shall discuss the ideas that are necessary to reformulate tensor calculus in a fashion that more directly illustrates the geometrical foundations of the subject. Appendix B goes one level deeper, and gives set-theoretical definitions for many of the primitive terms that we shall use here (such as *vector space* and *algebra*).

A *manifold* is a space that is locally Euclidean and in which there is a notion of differentiation. This can be made more precise as follows: There must be a differentiable one-to-one map, or *diffeomorphism*, from the neighborhood of any point of a manifold to the points of $\Re^n$, for some n. Such a map is called a *chart*, and the collection of all such maps for a given manifold is called an *atlas*. There is an additional requirement that two maps in the same atlas that overlap must do so smoothly; this means, among other things, that all charts in the same atlas must map to $\Re^n$ with the same n. The number n is thus characteristic of the entire manifold, and is called the *dimension* of the manifold.

A chart is realized by (local) coordinates on the manifold. Since an n-dimensional manifold, M, must map smoothly onto $\Re^n$, it must be possible to label the points of M, at least locally, by n numbers, say $z^1, \ldots, z^n$. Then the map is given by expressing these numbers as functions of the coordinates, $x^1, \ldots, x^n$, on $\Re^n$. Specifically, we write $z^\alpha(x^1, \ldots, x^n)$, for $\alpha = 1, \ldots, n$.

It is generally not possible to cover an entire manifold with one chart. For example the surface of a sphere is a manifold called S^2, and, as is well known, coordinate charts on S^2 must break down somewhere. The chart

$$x = \frac{4\theta}{3\pi}\cos\phi \tag{2.1}$$

$$y = \frac{4\theta}{3\pi}\sin\phi, \tag{2.2}$$

where θ and ϕ are the usual spherical coordinates (colatitude and azimuthal angles, respectively), maps the region $0 \leq \theta < 3\pi/4$ onto the open unit disk in $\Re^2$. The chart

$$x = \frac{4(\pi-\theta)}{3\pi}\cos\phi \tag{2.3}$$

$$y = \frac{4(\pi-\theta)}{3\pi}\sin\phi \tag{2.4}$$

then maps the region $\pi/4 < \theta \leq \pi$ onto the open unit disk in $\Re^2$. These two charts are thus sufficient to cover all of S^2, and therefore constitute an atlas. Any atlas for S^2 must contain at least two charts. In general, the number of charts needed to cover a manifold depends on its global topological properties.

A mapping from an m-dimensional manifold onto an n-dimensional manifold is called an *injection* if $m < n$, a *projection* if $m > n$, and a *bijection* if $m = n$. Consider a map from $\Re$ to an n-dimensional manifold, M. That is, $\Re \mapsto M$. Note that this is an injection if $n > 1$, and a bijection if $n = 1$. This map defines a *path* through the manifold, M. The points in M that are on the path are those in the range of the map. The realization of this mapping is given by expressing each of the coordinates on M as functions of the coordinate, x, on $\Re$. That is, we write $z^\alpha(x)$ for $\alpha = 1,\ldots,n$. As x varies along $\Re$, the coordinates z^α trace out the path in M. Note that although we keep writing down the coordinate realizations of these things, the notion of a map from one manifold to another has an intrinsic geometrical meaning as an association of members of one set of

points with members of another set of points, consistent with local topological properties of nearness, etc.

Now let P be a point on the above-mentioned path through the manifold, M. Denote its coordinates by $z_P^1, \ldots, z_P^n$.Since it lies along the path, there must exist a coordinate, x_0, of a point in $\Re$, such that $z_P^\alpha = z^\alpha(x_0)$ for $\alpha = 1, \ldots, n$. Now consider the derivatives of the functions, $z^\alpha(x)$, with respect to the path parameter, x. Denote these by dz^α/dx. Evaluate these at the point P. This gives the n numbers,

$$V^\alpha \equiv \frac{dz^\alpha}{dx}(x_0), \tag{2.5}$$

associated with the point, P.

It is clear that there are many different curves passing through point P that will yield the same set of n numbers. Indeed, any curve whose coordinates near P are given by

$$z^\alpha = z_P^\alpha + V^\alpha \delta x + \mathcal{O}(\delta x^2) \tag{2.6}$$

where $\delta x \equiv x - x_0$, will do so. The identification of these n numbers thus gives us a way to partition the set of all curves passing through point P into (an infinity of) equivalence classes; two curves are said to be equivalent if they yield the same set of n numbers. That is, two curves are equivalent if they both have the form given in the above equation (with the same V^α's).

Consider the set of equivalence classes of curves thus obtained. We can define addition and scalar multiplication among the elements of this set in the following very natural way: The equivalence class of curves with the n numbers V^α adds to the equivalence class of curves with the n numbers U^α to yield the equivalence class of curves with the n numbers $V^\alpha + U^\alpha$. The scalar a multiplies the equivalence class of curves with the n numbers V^α to yield the equivalence class of curves with the n numbers aV^α. With these operations, we have converted the space of all equivalence classes of curves through the point P into a vector space. This vector space will be called the *tangent space* at point P of the manifold. Its

elements have been introduced as equivalence classes of curves, but it will become clear momentarily that these may be identified with the usual notion of vectors as arrows with a certain magnitude and direction and with certain transformation properties. Note, however, that the base of the arrow is not free to move around, but rather is "pinned down" at the point P. There is a different tangent space at each point of a manifold, and vectors in one tangent space may not be added to vectors in another different tangent space. Note that the dimension of a tangent space is equal to the dimension of the manifold (in the above case, the dimension is n).

It is evident that the above-described n numbers V^α associated with an equivalence class of curves depend on our choice of coordinates for M. If our coordinates on M had been z'^α, then the n numbers would have been

$$V'^\alpha = \frac{dz'^\alpha}{dx}(x_0) = \frac{\partial z'^\alpha}{\partial z^\beta}\frac{dz^\beta}{dx}(x_0) = \frac{\partial z'^\alpha}{\partial z^\beta}V^\beta, \tag{2.7}$$

where we have adopted the convention of summation over repeated indices. Readers familiar with traditional presentations of tensor calculus will recognize this as the transformation law for components of contravariant vectors.

Recall that even though the components of a vector may vary from one coordinate system to another, the vector itself, as an abstract mathematical object, is an invariant geometrical concept. That is, given two sets of basis vectors, $\hat{\mathbf{e}}_\alpha$ and $\hat{\mathbf{e}}'_\alpha$, we can write the components of a vector $\mathbf{V}$ as V^α in the first system and as V'^α in the second. Though these will, in general, be different, the abstract vector $\mathbf{V} = V^\alpha\hat{\mathbf{e}}_\alpha = V'^\alpha_{\cdot}\hat{\mathbf{e}}'_\alpha$ retains its form under the change of basis.

So how can we introduce bases in our tangent spaces that will reflect this idea? Despite the fact that the above-described n numbers are coordinate-dependent, if we form a first-order linear differential operator by using them as coefficients

$$\hat{\mathbf{V}} \equiv V^\alpha\frac{\partial}{\partial z^\alpha} = V'^\alpha\frac{\partial}{\partial z'^\alpha}, \tag{2.8}$$

we see that this operator retains its form under a coordinate transformation. This much is clear from the above equation. By analogy with the argument in the preceeding paragraph, we can thus *identify* the operator $\hat{\mathbf{V}}$ with the vector $\mathbf{V}$, and the n operators $\partial/\partial z^\alpha$ with basis vectors that span the tangent space. Thus the idea of vectors as arrows, as equivalence classes of curves, and as first order linear differential operators are all valid descriptions of the same concept!

A word is in order concerning the basis vectors that we have used above. Note that they were induced by the coordinate system that we used. The choice of a coordinate system z^α on the manifold M gives rise to a natural basis $\partial/\partial z^\alpha$ in each tangent space at each point of the manifold (or, more precisely, at each point of M where the chart z^α is operative). A change in coordinate system thus gives rise to a change of basis; this is in accordance with the usual transformation properties of contravariant vectors. A basis that is thus induced by a coordinate system is called a *coordinate basis.* In the "arrow" picture, the basis vectors lie along the local coordinate axes. In the "equivalence class of curves" picture, they are curves that are locally coincident with the coordinate axes. In the "operator" picture, they are directional derivatives along the coordinate directions.

One might well ask if *all* possible bases are coordinate bases. The answer is "no." If we start from a coordinate basis and make a change of basis by taking various linearly independent combinations of basis vectors in each tangent space, where the combinations may vary from point to point in the manifold, we may arrive at a new basis that is not the coordinate basis for any coordinate system on M. Thus, starting from the coordinate basis, $\partial/\partial z^\alpha$, we may define the new basis

$$\hat{\mathbf{e}}_\beta = \Lambda_\beta{}^\alpha \frac{\partial}{\partial z^\alpha}, \tag{2.9}$$

where $(\Lambda_\beta{}^\alpha)$ is any nonsingular matrix. This new basis is perfectly good for

resolving vectors into coordinates. For example, the vector $\mathbf{V}$ may be written

$$\mathbf{V} = V^\alpha \frac{\partial}{\partial z^\alpha} = (V^\alpha \Lambda^\beta{}_\alpha)\hat{\mathbf{e}}_\beta \tag{2.10}$$

where the matrix $(\Lambda^\beta{}_\alpha)$ is the inverse of the matrix $(\Lambda_\gamma{}^\alpha)$. So the components of $\mathbf{V}$ in the new basis are $V^\alpha \Lambda^\beta{}_\alpha$. The only different thing about this new basis is that there may not be any system of coordinates Z^α such that $\hat{\mathbf{e}}_\alpha = \partial/\partial Z^\alpha$. In this case, such a basis is called a *noncoordinate basis.* This idea will become more clear and examples will be given in Subsection 2.2.5.

Meanwhile, since we have now attached vector spaces to every point of a manifold, we can go on to construct their dual spaces. The dual space to the tangent space of vectors at point P is called the *cotangent space* at point P. Its elements are called *covectors* or *covariant vectors* or *one forms.* Once again, the cotangent space has the same dimension as the manifold.

Once we have a set of basis vectors in the tangent space, say $\hat{\mathbf{e}}_\alpha$, there is induced a preferred set of basis covectors in the cotangent space, call them $\tilde{\omega}^\alpha$, such that $\langle \tilde{\omega}^\alpha, \hat{\mathbf{e}}_\beta \rangle = \delta^\alpha_\beta$. Thus we can represent a covector at point P by n numbers, say a_α, where, as usual, α can range from 1 to n. The abstract covector is then $\mathbf{a} = a_\alpha \tilde{\omega}^\alpha$. The covector a pairs with the vector V to yield

$$\langle \mathbf{a}, \mathbf{V} \rangle = \langle a_\alpha \tilde{\omega}^\alpha, V^\beta \hat{\mathbf{e}}_\beta \rangle = a_\alpha V^\beta \langle \tilde{\omega}^\alpha, \hat{\mathbf{e}}_\beta \rangle = a_\alpha V^\beta \delta^\alpha_\beta = a_\alpha V^\alpha. \tag{2.11}$$

Note that even though there is a naturally induced covector basis corresponding to a given vector basis, there is no natural correspondence between individual vectors and individual covectors. That is, there is no natural map from the tangent space to the cotangent space. Later on, we shall see that if we endow our manifold with a *metric*, such a map is established. The addition of a metric thus gives the manifold much more structure than it would otherwise have. At this point in our discussion, we are not assuming the existence of a metric on our manifold. As we shall see, even without a metric, a manifold has lots of interesting structure to study. The general philosophy of this discussion is to start

simply and slowly add structure; thus a discussion of metrics is deferred to the end of this section.

To make our discussion of covectors more concrete, let us suppose that we have a coordinate system z^α on our manifold, M. This induces the coordinate basis vectors $\partial/\partial z^\alpha$ on each tangent space of M. If we transform coordinates to another system Z^α, the components of the vector $\mathbf{V}$ transform according to Eq. (2.7). Now say the covector $\mathbf{a}$ has components a_α in the first coordinate system. The components of the covector must transform in such a way as to leave the scalar $\langle \mathbf{a}, \mathbf{V} \rangle$ invariant. Thus

$$a_\beta V^\beta = a'_\alpha V'^\alpha = a'_\alpha \frac{\partial z'^\alpha}{\partial z^\beta} V^\beta \tag{2.12}$$

so

$$a'_\alpha = \frac{\partial z^\beta}{\partial z'^\alpha} a_\beta. \tag{2.13}$$

Once again, readers familiar with traditional presentations of tensor calculus will recognize this as the transformation law for components of covariant vectors.

Now, how can we introduce bases in our cotangent spaces that will reflect the above ideas? Despite the fact that the n numbers a_α are coordinate-dependent, if we form the differential that has them as coefficients

$$\tilde{\mathbf{a}} \equiv a_\alpha dz^\alpha = a'_\alpha dz'^\alpha, \tag{2.14}$$

we see that this retains its form under a coordinate transformation. This much is clear from the above equation. We can thus *identify* the differential form $\tilde{\mathbf{a}}$ with the covector $\mathbf{a}$, and the n differentials dz^α with basis covectors that span the cotangent space.

Thus, just as contravariant vectors could be identified with first order linear differential operators, we see that covectors can be identified with differential forms. These descriptions are dual to each other, so

$$\langle dz^\alpha, \frac{\partial}{\partial z^\beta} \rangle = \delta^\alpha_\beta. \tag{2.15}$$

Finally we note that the same distinction between coordinate and noncoordinate bases that applied to our discussion of tangent space bases also applies to cotangent space bases. Up until now, we have restricted our attention to coordinate cotangent bases, but we could define new basis one forms by taking linear combinations of the dz^α where the combinations may vary from point to point in the manifold. In this way, we may arrive at a new basis that is not the coordinate cotangent basis for any coordinate system on M. Thus, starting from the coordinate cotangent space basis, dz^α, we may define the new cotangent space basis

$$\tilde{\omega}^\beta = \Lambda^\beta{}_\alpha dz^\alpha \tag{2.16}$$

where $(\Lambda^\beta{}_\alpha)$ is any nonsingular matrix. This new basis is perfectly good for resolving covectors into coordinates. For example, the covector $\mathbf{a}$ may be written

$$\mathbf{a} = a_\alpha dz^\alpha = (a_\alpha \Lambda_\beta{}^\alpha)\tilde{\omega}^\beta \tag{2.17}$$

where the matrix $(\Lambda_\gamma{}^\alpha)$ is the inverse of the matrix $(\Lambda^\beta{}_\alpha)$. So the components of $\mathbf{a}$ in the new basis are $a_\alpha \Lambda_\beta{}^\alpha$. The only different thing about this new basis is that there may not be any system of coordinates Z^α such that $\tilde{\omega}^\alpha = dZ^\alpha$. Once again, this idea will become more clear and examples will be given in Subsection 2.2.5.

2.2.2 General Tensors and the Tensor Product

Now that we have a tangent space and a cotangent space associated with each and every point of our manifold, we can create still bigger spaces at each point by taking the Cartesian product of some number of tangent spaces and some number of cotangent spaces. Suppose we define the space Π^s_r to be the Cartesian product of s copies of the tangent space and r copies of the cotangent space at point P of a manifold M. Consider a multilinear map $\Pi^s_r \mapsto \Re$. That is, we are considering a map that takes s vectors and r covectors at point P and returns a real number. If the s vectors are denoted $\mathbf{V}_1, \ldots, \mathbf{V}_s$, and the r covectors are denoted

$\mathbf{a}^1, \ldots, \mathbf{a}^r$, then the real number will be denoted by $\mathbf{T}(\mathbf{a}^1, \ldots, \mathbf{a}^r, \mathbf{V}_1, \ldots, \mathbf{V}_s)$. By a "multilinear" map, we mean that $\mathbf{T}$ is linear in all of its arguments. Such a map is said to be a *tensor of type* (r, s). Note that a vector is a tensor of type $(1, 0)$, and a covector is a tensor of type $(0, 1)$; this is because a vector can take a covector and return a real number (by the pairing), and vice versa.

There is an obvious way to define addition among tensors: Given two tensors, $\mathbf{T}_1$ and $\mathbf{T}_2$, we can define a new tensor, $\mathbf{T}_3$, by the prescription

$$\mathbf{T}_3(\mathbf{a}^1, \ldots, \mathbf{a}^r, \mathbf{V}_1, \ldots, \mathbf{V}_s) = \mathbf{T}_1(\mathbf{a}^1, \ldots, \mathbf{a}^r, \mathbf{V}_1, \ldots, \mathbf{V}_s) + \mathbf{T}_2(\mathbf{a}^1, \ldots, \mathbf{a}^r, \mathbf{V}_1, \ldots, \mathbf{V}_s), \quad (2.18)$$

for all possible arguments. In this case, we write $\mathbf{T}_3 = \mathbf{T}_1 + \mathbf{T}_2$. This operation of addition makes the space of all tensors of type (r, s) a vector space.

Suppose we have two vectors, $\mathbf{U}_1$ and $\mathbf{U}_2$, and a covector, $\mathbf{b}^1$, at some point of a manifold. Suppose we are given anew a pair of covectors, $\mathbf{a}^1$ and $\mathbf{a}^2$, and a vector, $\mathbf{V}_1$ (at the same point of the manifold). Consider the following recipe for obtaining a real number: Pair the two covectors with $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively, and pair the vector with $\mathbf{b}^1$. This gives us three real numbers. Multiply them together to get a single real number. In this way, the presence of $\mathbf{U}_1, \mathbf{U}_2$, and $\mathbf{b}^1$ provides us with a map from Π_2^1 to $\Re$. It is easily seen that this map is multilinear. Thus, the presence of $\mathbf{U}_1, \mathbf{U}_2$, and $\mathbf{b}^1$ provides us with the following tensor of type $(2, 1)$:

$$\mathbf{T}(\mathbf{a}^1, \mathbf{a}^2, \mathbf{V}_1) = \langle \mathbf{a}^1, \mathbf{U}_1 \rangle \langle \mathbf{a}^2, \mathbf{U}_2 \rangle \langle \mathbf{b}^1, \mathbf{V}_1 \rangle. \quad (2.19)$$

A tensor formed in this way is said to be the *tensor product* of $\mathbf{U}_1, \mathbf{U}_2$, and $\mathbf{b}^1$. This is denoted

$$\mathbf{T} = \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{b}^1. \quad (2.20)$$

More generally, given r vectors, $\mathbf{U}_1, \ldots, \mathbf{U}_r$, and s covectors, $\mathbf{b}^1, \ldots, \mathbf{b}^s$, we can form a tensor of type (r, s) by taking the tensor product

$$\mathbf{T} = \mathbf{U}_1 \otimes \ldots \otimes \mathbf{U}_r \otimes \mathbf{b}^1 \otimes \ldots \otimes \mathbf{b}^s. \quad (2.21)$$

If we feed this tensor the r covectors, $\mathbf{a}_1, \ldots, \mathbf{a}_r$, and the s vectors, $\mathbf{V}_1, \ldots, \mathbf{V}_s$, then we get the scalar

$$\mathbf{T}(\mathbf{a}_1, \ldots, \mathbf{a}_r, \mathbf{V}_1, \ldots, \mathbf{V}_s) = \langle \mathbf{a}_1, \mathbf{U}_1 \rangle \ldots \langle \mathbf{a}_r, \mathbf{U}_r \rangle \langle \mathbf{b}_1, \mathbf{V}_1 \rangle \ldots \langle \mathbf{b}_s, \mathbf{V}_s \rangle. \quad (2.22)$$

The space of all possible tensors (of any type) at some point in an n-dimensional manifold may be thought of as an infinite dimensional vector space, although it is somewhat strange in that two of its elements can be added if and only if they are tensors of the same type. In any event, the tensor product makes this space an algebra.

It is straightforward to see that the vector space of all tensors of type (r, s) is n^{r+s}-dimensional. That is, a tensor of type (r, s) has n^{r+s} independent components. A moment's thought convinces one that a basis for this space is given by the n^{r+s} basis tensors

$$\hat{\mathbf{e}}_{\alpha_1} \otimes \ldots \otimes \hat{\mathbf{e}}_{\alpha_r} \otimes \tilde{\omega}^{\beta_1} \otimes \ldots \otimes \tilde{\omega}^{\beta_s}, \quad (2.23)$$

where the $\hat{\mathbf{e}}$'s and $\tilde{\omega}$'s are the basis vectors and basis covectors in the tangent and cotangent spaces, respectively, and where the α and β indices all range from 1 to n. Thus, a general tensor may be written

$$\mathbf{T} = T^{\alpha_1 \ldots \alpha_r}_{\beta_1 \ldots \beta_s} \hat{\mathbf{e}}_{\alpha_1} \otimes \ldots \otimes \hat{\mathbf{e}}_{\alpha_r} \otimes \tilde{\omega}^{\beta_1} \otimes \ldots \otimes \tilde{\omega}^{\beta_s}. \quad (2.24)$$

Finally, we consider the transformation properties of the components of these general tensors. We know how vector and covector components transform, and we know that a tensor of type (r, s) takes r covectors and s vectors and returns a scalar invariant. Thus, by an argument identical to that which led to Eq. (2.13), we find that for a transformation from one coordinate basis to another coordinate basis

$$T'^{\alpha_1 \ldots \alpha_r}_{\beta_1 \ldots \beta_s} = \frac{\partial z'^{\alpha_1}}{\partial z^{\mu_1}} \cdots \frac{\partial z'^{\alpha_r}}{\partial z^{\mu_r}} \frac{\partial z^{\nu_1}}{\partial z'^{\beta_1}} \cdots \frac{\partial z^{\nu_s}}{\partial z'^{\beta_s}} T^{\mu_1 \ldots \mu_r}_{\nu_1 \ldots \nu_s}. \quad (2.25)$$

The usual distinction between coordinate and noncoordinate bases applies here as well, so that for a transformation between general bases the above equation generalizes to

$$T'^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_s} = \Lambda^{\alpha_1}{}_{\mu_1}\cdots\Lambda^{\alpha_r}{}_{\mu_r}\Lambda_{\beta_1}{}^{\nu_1}\cdots\Lambda_{\beta_s}{}^{\nu_s}T^{\mu_1\ldots\mu_r}_{\nu_1\ldots\nu_s}. \tag{2.26}$$

2.2.3 The Lie Bracket

Given a vector field, $\mathbf{V}$, the corresponding first-order linear differential operator is:

$$\hat{\mathbf{V}} \equiv V^\alpha \frac{\partial}{\partial z^\alpha}. \tag{2.27}$$

Notice that the αth component of the vector can be recovered by applying the operator to z^α:

$$V^\alpha = \hat{\mathbf{V}} z^\alpha. \tag{2.28}$$

As has been mentioned, it is possible to actually *identify* the vector with its corresponding operator. Many mathematics texts actually do this, and it is perfectly permissible since there is an obvious one to one correspondence between vectors and first-order linear differential operators by the above equations. Indeed, there are numerous advantages to such identification, but we shall continue to use the circumflex to distinguish the operator in order to avoid any ambiguity.

It is important to note that the operators corresponding to two different vector fields do not, in general, commute. Indeed, the commutator of two first-order linear differential operators is another first-order linear differential operator. At first this may seem surprising because it is not obvious that this commutator is a *first* order operator. By writing it in terms of the components of $\mathbf{V}$ and $\mathbf{U}$, however, we see that the second order terms do indeed cancel:

$$\begin{aligned}\left[\hat{\mathbf{V}},\hat{\mathbf{U}}\right] &= V^\beta\frac{\partial}{\partial z^\beta}\left(U^\alpha\frac{\partial}{\partial z^\alpha}\right) - U^\beta\frac{\partial}{\partial z^\beta}\left(V^\alpha\frac{\partial}{\partial z^\alpha}\right)\\ &= \left(V^\beta U^\alpha{}_{,\beta} - U^\beta V^\alpha{}_{,\beta}\right)\frac{\partial}{\partial z^\alpha}.\end{aligned} \tag{2.29}$$

The vector whose operator is the commutator of the operators of two other vectors, $\mathbf{V}$ and $\mathbf{U}$, is said to be the *Lie bracket* of those two vectors, and is denoted by $[\mathbf{V}, \mathbf{U}]$. Note that the Lie bracket operation makes the space of all vector fields into a Lie algebra.

Using the Lie bracket, it is possible to give a simple test that will determine whether or not any given set of basis vectors is a coordinate basis: A set of n linearly independent vectors constitutes a coordinate basis if and only if the Lie bracket of any two elements of the set vanishes. The "only if" part of this theorem is obvious, since coordinate basis vectors are partial derivatives and these always commute with each other. The converse, however, is a special case of something called Frobenius' theorem, and is somewhat harder to see. To prove it algebraically, we must show that it is possible to actually construct a coordinate system (at least locally) given the n linearly independent commuting vectors. We shall not follow this approach here (see Schutz [15] for details on how to prove it this way). Instead, we shall follow a more geometrical line of reasoning that will make the theorem almost obvious. To do this, however, we first need to learn about the Lie derivative.

2.2.4 Lie Derivatives

The *Lie derivative* of a scalar field, $f(z)$, with respect to the vector field, $\mathbf{V}$, is a new scalar field denoted by $\mathcal{L}_V f$, and is given by:

$$\mathcal{L}_V f = \hat{\mathbf{V}} f = V^\alpha \frac{\partial f}{\partial z^\alpha}. \tag{2.30}$$

This is recognized as the directional derivative of f along the vector field, $\mathbf{V}$. Along any given field line of $\mathbf{V}$, it is possible to define a coordinate, λ, such that:

$$V^\alpha = \frac{dz^\alpha}{d\lambda} \tag{2.31}$$

so

$$\hat{\mathbf{V}} = \frac{d}{d\lambda}, \tag{2.32}$$

and so the Lie derivative of f with respect to $\mathbf{V}$ is simply $df/d\lambda$. That is, we evaluate the scalar field at the points $z(\lambda_0)$ and $z(\lambda_0 + \delta\lambda)$ along the field line, subtract the first value from the second, divide the result by $\delta\lambda$, and let $\delta\lambda$ go to zero to get the Lie derivative. In Fig. 2.1, these two points of evaluation are denoted by A and B.

A scalar field, f, whose Lie derivative with respect to $\mathbf{V}$ vanishes is said to be a *Lie dragged* scalar field with respect to the vector field $\mathbf{V}$. Intuitively, this means that the scalar field is constant along the field lines of $\mathbf{V}$. Alternatively stated, it means that the scalar field satisfies the first-order linear differential equation $\hat{\mathbf{V}}f = 0$, whose characteristics are the field lines of $\mathbf{V}$. Thus, if the value of a Lie dragged scalar field is specified at any one point of a field line of $\mathbf{V}$, its value everywhere else on that same field line is determined (it's the same value). Using this concept, we can reword our definition of a Lie derivative: Begin by evaluating the scalar field f at point A. Next, drag the scalar f at point B back to point A to get the scalar f^* at A (note $f^*(A) = f(B)$). Now at the point A we subtract f from f^*, divide the result by $\delta\lambda$, and let $\delta\lambda$ go to zero to get $\mathcal{L}_V f$. This may sound like a fancy way of saying the same thing, but it will aid in our efforts to generalize the Lie derivative to act on other things besides scalars.

Consider the problem of trying to define an analogous derivative that acts on contravariant vectors. We could begin by evaluating a vector field, say $\mathbf{U}$, at the same two points, $z(\lambda_0)$ and $z(\lambda_0 + \delta\lambda)$, along a field line of $\mathbf{V}$. Unfortunately, however, we cannot subtract them because they live in two different vector spaces: The first lives in the space of all vectors at the point $z(\lambda_0)$, while the second lives in the space of all vectors at $z(\lambda_0+\delta\lambda)$. We are dealing with spaces in which there may be no notion of parallel transport, so there is no natural way of comparing vectors located at two different points.

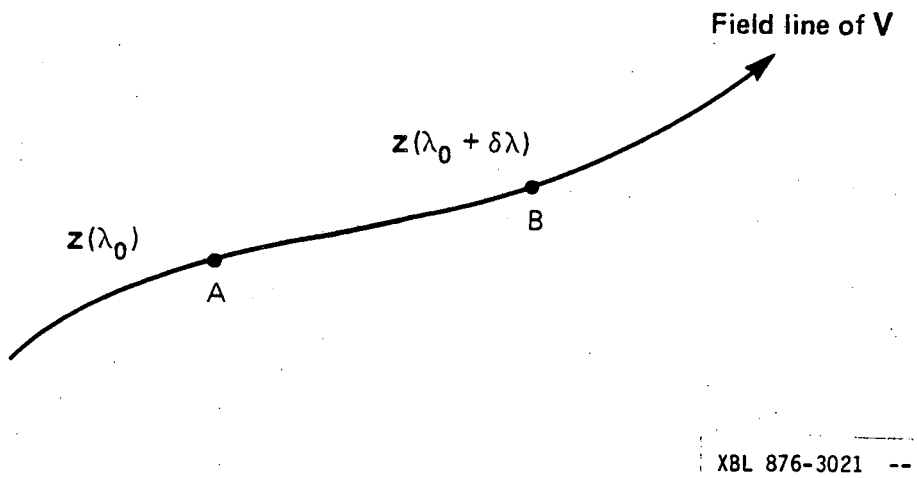


Figure 2.1: Lie Differentiation of a Scalar Field

So we must be a little more clever. Refer to Fig. 2.2. Just as we have coordinatized a given field line of $\mathbf{V}$ by λ, we shall use μ to coordinatize a given field line of $\mathbf{U}$. The operator $\hat{\mathbf{U}}$is then $d/d\mu$. It acts on scalars by evaluating them at $z(\mu_0)$ and $z(\mu_0 + \delta\mu)$, subtracting the first value from the second, dividing by $\delta\mu$, and letting $\delta\mu$ go to zero. In Fig. 2.2, these two points of evaluation are denoted by A and C; note that we have arranged things in this figure so that point A is parametrized by both λ_0 on the $\mathbf{V}$ field line, and μ_0 on the $\mathbf{U}$ field line.

Now we can imagine sliding the points A and C along the $\mathbf{V}$ field lines for an increment $\delta\lambda$, to arrive at the new points B and D, respectively. These new points define a new first-order linear differential operator based at the point B. It acts on scalars by evaluating them at the points B and D, subtracting the first value from the second, dividing by $\delta\mu$ (it is clear that points B and D coincide as $\delta\mu \to 0$), and letting $\delta\mu$ go to zero. This first-order linear differential operator at B corresponds to a vector at point B, and so we see that we have found a natural way to drag the vector field $\mathbf{U}$ along the vector field $\mathbf{V}$. If a vector field $\mathbf{U}$ is unchanged by dragging it along $\mathbf{V}$, then it is said to be a *Lie dragged* vector field with respect to $\mathbf{V}$.

Armed with this insight, we are ready to define the Lie derivative of a vector field, $\mathbf{U}$ with respect to another vector field, $\mathbf{V}$. We begin by evaluating $\mathbf{U}$ at point A. Next, we drag the vector $\mathbf{U}$ at point B back to point A to get the vector $\mathbf{U}^*$ at A. Now we can subtract $\mathbf{U}$ from $\mathbf{U}^*$, divide the result by $\delta\lambda$, and let $\delta\lambda$ go to zero to get $\mathcal{L}_V\mathbf{U}$. It should be clear from this description that the Lie derivative of a Lie dragged vector field vanishes, just as was the case for scalars.

Now that we have the geometrical picture of what is happening, we need to find an analytic expression for $\mathcal{L}_V\mathbf{U}$. Refer to Fig. 2.3. It is clear that we may

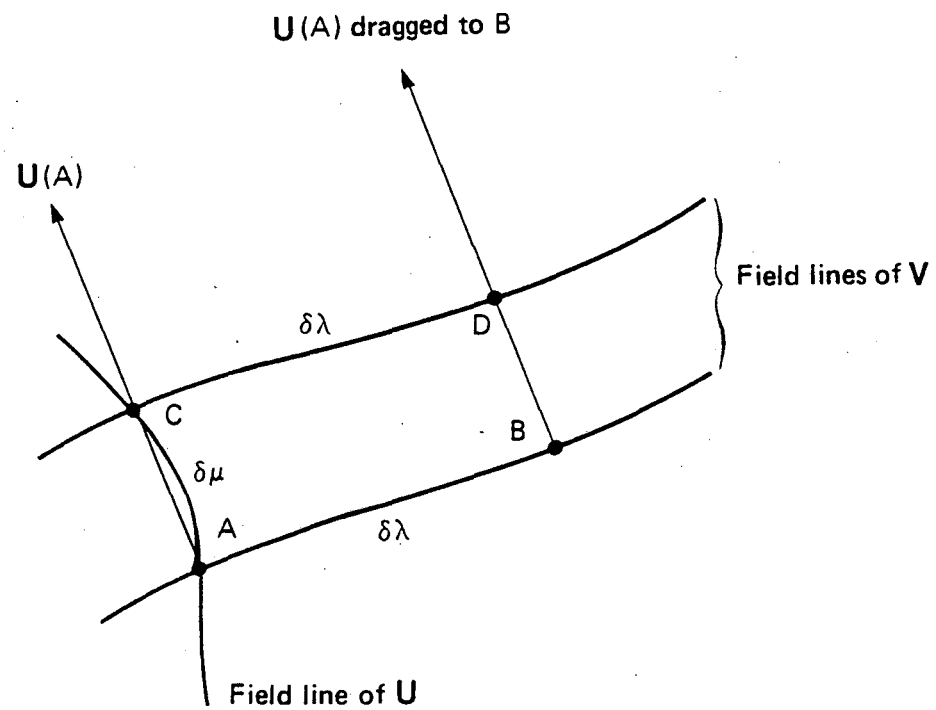


Figure 2.2: Lie Dragging a Vector Field Along Another Vector Field

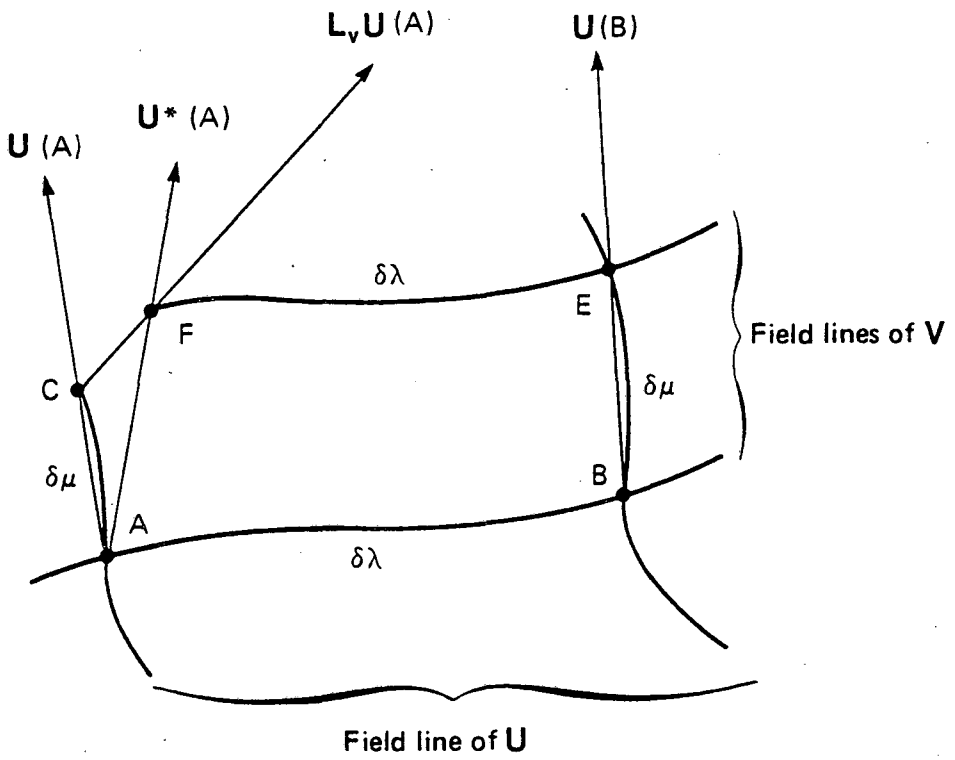


Figure 2.3: Lie Differentiation of a Vector Field

write:

$$U^\alpha(A) = \lim_{\delta\mu\to 0} \frac{z^\alpha_C - z^\alpha_A}{\delta\mu},$$
$$U^\alpha(B) = \lim_{\delta\mu\to 0} \frac{z^\alpha_E - z^\alpha_B}{\delta\mu},$$
$$U^{*\alpha}(A) = \lim_{\delta\mu\to 0} \frac{z^\alpha_F - z^\alpha_A}{\delta\mu},$$

and

$$(\mathcal{L}_V \mathbf{U})^\alpha(A) = \lim_{\delta\lambda\to 0} \frac{U^{*\alpha}(A) - U^\alpha(A)}{\delta\lambda}$$
$$= \lim_{\delta\lambda\to 0} \lim_{\delta\mu\to 0} \frac{z^\alpha_F - z^\alpha_C}{\delta\mu\delta\lambda}, \tag{2.33}$$

where z^α_A through z^α_F are the coordinates at the points A through F, respectively. To find these coordinates, we use Taylor expansion. Thus, to express the coordinates of point B in terms of quantities at point A, we write:

$$z^\alpha_B = z^\alpha(\lambda_0 + \delta\lambda)$$
$$= z^\alpha_A + \left.\frac{dz^\alpha}{d\lambda}\right|_A \delta\lambda + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\lambda^2}\right|_A \delta\lambda^2 + \cdots \tag{2.34}$$

Similarly, the coordinates of point C are given by:

$$z^\alpha_C = z^\alpha(\mu_0 + \delta\mu)$$
$$= z^\alpha_A + \left.\frac{dz^\alpha}{d\mu}\right|_A \delta\mu + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\mu^2}\right|_A \delta\mu^2 + \cdots \tag{2.35}$$

Next, the coordinates of point E can be expressed in terms of quantities at point B:

$$z^\alpha_E = z^\alpha_B + \left.\frac{dz^\alpha}{d\mu}\right|_B \delta\mu + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\mu^2}\right|_B \delta\mu^2 + \cdots, \tag{2.36}$$

and these in turn may be expressed in terms of quantities at point A:

$$z^\alpha_E = z^\alpha_A + \left.\frac{dz^\alpha}{d\lambda}\right|_A \delta\lambda + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\lambda^2}\right|_A \delta\lambda^2 + \cdots$$
$$+ \left.\frac{dz^\alpha}{d\mu}\right|_A \delta\mu + \left.\frac{d^2z^\alpha}{d\lambda d\mu}\right|_A \delta\lambda\delta\mu + \cdots$$
$$+ \frac{1}{2}\left.\frac{d^2z^\alpha}{d\mu^2}\right|_A \delta\mu^2 + \cdots \tag{2.37}$$

Finally, the coordinates of point F can be expressed in terms of quantities at point E, which in turn can be expressed in terms of quantities at point B, which in turn can be expressed in terms of quantities at point A:

$$\begin{aligned}
z_F^\alpha &= z_E^\alpha - \left.\frac{dz^\alpha}{d\lambda}\right|_E \delta\lambda + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\lambda^2}\right|_E \delta\lambda^2 + \cdots \\
&= z_B^\alpha + \left.\frac{dz^\alpha}{d\mu}\right|_B \delta\mu + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\mu^2}\right|_B \delta\mu^2 + \cdots \\
&\quad - \left.\frac{dz^\alpha}{d\lambda}\right|_B \delta\lambda - \left.\frac{d^2z^\alpha}{d\mu d\lambda}\right|_B \delta\mu\delta\lambda - \cdots \\
&\quad + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\lambda^2}\right|_B \delta\lambda^2 + \cdots \\
&= z_A^\alpha + \left.\frac{dz^\alpha}{d\lambda}\right|_A \delta\lambda + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\lambda^2}\right|_A \delta\lambda^2 + \cdots \\
&\quad + \left.\frac{dz^\alpha}{d\mu}\right|_A \delta\mu + \left.\frac{d^2z^\alpha}{d\lambda d\mu}\right|_A \delta\lambda\delta\mu + \cdots \\
&\quad + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\mu^2}\right|_A \delta\mu^2 + \cdots \\
&\quad - \left.\frac{dz^\alpha}{d\lambda}\right|_A \delta\lambda - \left.\frac{d^2z^\alpha}{d\lambda^2}\right|_A \delta\lambda^2 - \cdots \\
&\quad - \left.\frac{d^2z^\alpha}{d\mu d\lambda}\right|_A \delta\mu\delta\lambda - \cdots \\
&\quad + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\lambda^2}\right|_A \delta\lambda^2 + \cdots \\
&= z_A^\alpha + \left.\frac{dz^\alpha}{d\mu}\right|_A \delta\mu + \frac{1}{2}\left.\frac{d^2z^\alpha}{d\mu^2}\right|_A \delta\mu^2 + \cdots \\
&\quad + \left.\frac{d^2z^\alpha}{d\lambda d\mu}\right|_A \delta\lambda\delta\mu - \left.\frac{d^2z^\alpha}{d\mu d\lambda}\right|_A \delta\mu\delta\lambda + \cdots
\end{aligned} \tag{2.38}$$

Thus, using Eqs. (2.33), we find:

$$U^\alpha(A) = \left.\frac{dz^\alpha}{d\mu}\right|_A,$$

and

$$U^{*\alpha}(A) = \left.\frac{dz^\alpha}{d\mu}\right|_A + \left(\left.\frac{d^2z^\alpha}{d\lambda d\mu}\right|_A - \left.\frac{d^2z^\alpha}{d\mu d\lambda}\right|_A\right)\delta\lambda + \cdots$$

and so

$$
\begin{aligned}
(\mathcal{L}_V \mathbf{U})^\alpha(A) &= \left.\frac{d^2 z^\alpha}{d\lambda d\mu}\right|_A - \left.\frac{d^2 z^\alpha}{d\mu d\lambda}\right|_A \\
&= \left(\hat{\mathbf{V}}\hat{\mathbf{U}} - \hat{\mathbf{U}}\hat{\mathbf{V}}\right) z^\alpha \Big|_A \\
&= \left[\hat{\mathbf{V}}, \hat{\mathbf{U}}\right] z^\alpha \Big|_A . \qquad (2.39)
\end{aligned}
$$

We have just demonstrated that the Lie derivative of $\mathbf{U}$ with respect to $\mathbf{V}$ is simply the Lie bracket of $\mathbf{V}$ and $\mathbf{U}$:

$$\mathcal{L}_V \mathbf{U} = [\mathbf{V}, \mathbf{U}] . \qquad (2.40)$$

In a coordinate basis, this result may be written

$$(\mathcal{L}_V \mathbf{U})^\alpha = V^\beta U^\alpha{}_{,\beta} - U^\beta V^\alpha{}_{,\beta}. \qquad (2.41)$$

Note that this way of writing the result may be taken as valid for a noncoordinate basis as well if we reinterpret the commas as meaning "operation by the basis vector." That is, $f_{,\alpha}$ denotes the result of applying to f the operator corresponding to the basis vector $\hat{\mathbf{e}}_\alpha$. For a coordinate basis, the operators corresponding to basis vectors are simply partial derivatives with respect to the coordinates, so this reduces to the usual meaning of the comma. This generalization of what the comma means will be useful in everything that follows.

Now that we know how to take the Lie derivative of a contravariant vector field, we shall try to extend this process to covector fields. Recall that covectors contract with contravariant vectors to give scalars. We define a *Lie dragged* covector field to be one which when contracted with any Lie dragged contravariant vector field yields a Lie dragged scalar field. To take the Lie derivative of a covector field $\mathbf{a}$ with respect to $\mathbf{V}$, we evaluate $\mathbf{a}$ at the points A and B in Fig. 2.1, drag $\mathbf{a}(B)$ back to A to get $\mathbf{a}^*(A)$, subtract $\mathbf{a}(A)$ from $\mathbf{a}^*(A)$, divide by $\delta\lambda$, and let $\delta\lambda$ go to zero. The result is:

$$(\mathcal{L}_V \mathbf{a})_\alpha = V^\beta a_{\alpha,\beta} + V^\beta{}_{,\alpha} a_\beta, \qquad (2.42)$$

Once again, this result is valid for noncoordinate bases if we generalize the meaning of the commas.

Next we consider the Lie derivative of a general tensor. We first define a *Lie dragged* tensor of type (r,s) as one which yields a Lie dragged scalar field when fed r Lie dragged covectors and s Lie dragged vectors. To take the Lie derivative of a tensor $\mathbf{T}$ of type (r,s) with respect to $\mathbf{V}$, we evaluate $\mathbf{T}$ at the points A and B in Fig. 2.1, drag $\mathbf{T}(B)$ back to A to get $\mathbf{T}^*(A)$, subtract $\mathbf{T}(A)$ from $\mathbf{T}^*(A)$, divide by $\delta\lambda$, and let $\delta\lambda$ go to zero. The result is:

$$\begin{aligned}(\mathcal{L}_V\mathbf{T})^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_s} = {} & V^\gamma T^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_s,\gamma} \\ & - V^{\alpha_1}{}_{,\gamma} T^{\gamma\alpha_2\ldots\alpha_r}_{\beta_1\ldots\beta_s} - \cdots \\ & - V^{\alpha_r}{}_{,\gamma} T^{\alpha_1\ldots\alpha_{r-1}\gamma}_{\beta_1\ldots\beta_s} \\ & + V^\gamma{}_{,\beta_1} T^{\alpha_1\ldots\alpha_r}_{\gamma\beta_2\ldots\beta_s} + \cdots \\ & + V^\gamma{}_{,\beta_s} T^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_{s-1}\gamma} \end{aligned} \tag{2.43}$$

Note that the above geometrical picture for Lie derivatives of general tensors is equivalent to the neat coordinate-free algebraic formula

$$\begin{aligned}\mathcal{L}_V\left(\mathbf{T}(\mathbf{a}_1,\ldots,\mathbf{a}_r,\mathbf{U}_1,\ldots,\mathbf{U}_s)\right) = {} & (\mathcal{L}_V\mathbf{T})(\mathbf{a}_1,\ldots,\mathbf{a}_r,\mathbf{U}_1,\ldots,\mathbf{U}_s) \\ & + \mathbf{T}(\mathcal{L}_V\mathbf{a}_1,\ldots,\mathbf{a}_r,\mathbf{U}_1,\ldots,\mathbf{U}_s) + \cdots \\ & + \mathbf{T}(\mathbf{a}_1,\ldots,\mathcal{L}_V\mathbf{a}_r,\mathbf{U}_1,\ldots,\mathbf{U}_s) \\ & + \mathbf{T}(\mathbf{a}_1,\ldots,\mathbf{a}_r,\mathcal{L}_V\mathbf{U}_1,\ldots,\mathbf{U}_s) + \cdots \\ & + \mathbf{T}(\mathbf{a}_1,\ldots,\mathbf{a}_r,\mathbf{U}_1,\ldots,\mathcal{L}_V\mathbf{U}_s). \end{aligned} \tag{2.44}$$

Finally, it is straightforward to show that Lie derivatives obey the Leibniz rule over the tensor product. That is

$$\mathcal{L}_V(\mathbf{T}_1\otimes\mathbf{T}_2) = (\mathcal{L}_V\mathbf{T}_1)\otimes\mathbf{T}_2 + \mathbf{T}_1\otimes(\mathcal{L}_V\mathbf{T}_2). \tag{2.45}$$

Before leaving this subsection, it is important to emphasize that the same geometrical notions that led us to the Lie derivative of a vector field still apply

for arbitrary tensors: The Lie derivative is the natural way to "drag" any tensorial object along the field lines of a vector field. Just as we dragged the vector $\mathbf{U}$ along the field line of the vector $\mathbf{V}$ for a parameter interval $\delta\lambda$ to get the vector

$$\mathbf{U}^* = \mathbf{U} + (\mathcal{L}_V \mathbf{U})\delta\lambda \tag{2.46}$$

(see Eq. (2.33)), so we can drag the tensor $\mathbf{T}$ in exactly the same way to get

$$\mathbf{T}^* = \mathbf{T} + (\mathcal{L}_V \mathbf{T})\delta\lambda. \tag{2.47}$$

This geometrical insight is crucial to the understanding of Lie transforms.

2.2.5 Examples of Coordinate and Noncoordinate Bases

We are now in a position to understand the theorem presented at the end of Section 2.2.3 from a geometrical point of view. Fig. 2.3 and Eq. (2.33) make it clear that the Lie bracket of two vector fields is related to the infinitesimal difference in position resulting from the operation of moving along the first vector field for a certain parameter interval, then along the second vector field, then backwards for the same parameter interval along the first, then backwards along the second. Clearly, if the vector fields involved are basis elements of a coordinate basis, this operation will simply take one around a square right back to the original position. The sides of the square are the contours of constant values of the two coordinates involved. Conversely, if two members of a set of n linearly independent vectors have nonvanishing Lie bracket, then it is impossible to construct a coordinate system that has those vectors as a basis because moving around the above-described infinitesimal loop does not return one to the starting point; the changing parameters do not "hook together" in the manner necessary for them to be coordinates.

Part of the reason that this concept of coordinate and noncoordinate bases is tricky is that there is no need for such a distinction in Cartesian coordinates.

There, the coordinate basis is identical to the usual orthonormal basis

$$\frac{\partial}{\partial x^\alpha} = \hat{\mathbf{e}}_\alpha. \tag{2.48}$$

A good example of a familiar situation for which the distinction *is* important is that of polar coordinates in two dimensions. The usual polar unit vectors, $\hat{\mathbf{r}}$ and $\hat{\theta}$, are not a coordinate basis since

$$\left[\hat{\mathbf{r}}, \hat{\theta}\right] = -\frac{\hat{\theta}}{r}. \tag{2.49}$$

On the other hand, $\partial/\partial r$ and $\partial/\partial\theta$ *do* constitute a valid coordinate basis, and these are related to the above orthonormal basis by

$$\frac{\partial}{\partial r} = \hat{\mathbf{r}}$$

and

$$\frac{\partial}{\partial \theta} = r\hat{\theta}. \tag{2.50}$$

The important point is that there are *no* pair of coordinates, ξ and η, such that $\hat{\mathbf{r}} = \partial/\partial\xi$ *and* $\hat{\theta} = \partial/\partial\eta$. Geometrically, this is because if we traverse an infinitesimal loop following first the $\hat{\mathbf{r}}$ vector field and then the $\hat{\theta}$ vector field (and then returning along them, respectively) we will not arrive at our starting point (see Fig. 2.4). The factor of r on the right hand side of the second of Eqs. (2.50) corrects for this and gives us a coordinate basis.

As mentioned previously, the above distinction also holds for covectors. To pursue the above example, the covector basis consisting of dr and $d\theta$ is dual to the vector basis consisting of $\partial/\partial r$ and $\partial/\partial\theta$. It follows that the covector basis

$$\tilde{\mathbf{r}} = dr$$

and

$$\tilde{\theta} = r\,d\theta \tag{2.51}$$

is dual to the vector basis $\hat{\mathbf{r}}$ and $\hat{\theta}$. Once again, there is no pair of coordinates, ξ and η, such that $\tilde{\mathbf{r}} = d\xi$ *and* $\tilde{\theta} = d\eta$.

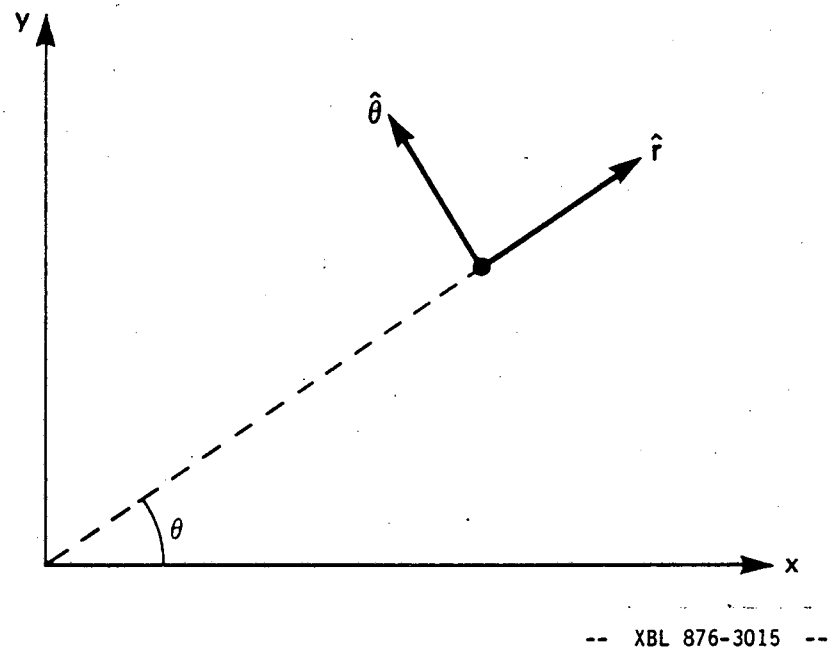


Figure 2.4: Polar Coordinate Unit Vectors

2.2.6 Differential Forms

An *s-form* is defined as a tensor of type $(0, s)$ that is antisymmetric in every pair of its s vector arguments. In particular, a zero form is a scalar and a one form is a covector; a two form, $\tilde{\Omega}$, obeys

$$\tilde{\Omega}(\mathbf{U}, \mathbf{V}) = -\tilde{\Omega}(\mathbf{V}, \mathbf{U}), \tag{2.52}$$

etc. It follows that the components of an s-form are antisymmetric under interchange of any pair of indices. In particular, this means that the s indices must all be different, or else the component will vanish. Hence the requirement of antisymmetry means that there are no longer n^s independent components. Instead, a standard combinatorial argument shows that only

$$\left(\begin{array}{c} n \\ s \end{array} \right) \equiv \frac{n!}{s!(n-s)!} \tag{2.53}$$

of the components are truly independent. This means, among other things, that there are no nontrivial s-forms in an n-dimensional space if $s > n$, that an n-form has only one nontrivial component, etc. The total number of independent components of *all* forms in a space of dimension n is thus

$$\sum_{s=0}^{n} \left(\begin{array}{c} n \\ s \end{array} \right) = (1+1)^n = 2^n. \tag{2.54}$$

Note that s-forms inherit some properties from the fact that they are tensors of type $(0, s)$. In particular, two s-forms may be added to get a third s-form. Thus, the set of all forms at a point in an n-dimensional manifold may be thought of as a 2^n-dimensional vector space, although it is somewhat strange in that its elements may be added if and only if they are both s-forms for some s. Note, however, that this space is *not* an algebra under the tensor product operation because it is not closed under that operation: The tensor product of two forms is not necessarily a form. If we take the tensor product of an s_1-form with an

s_2-form, we get a tensor of type $(0, s_1 + s_2)$ that is clearly antisymmetric under interchange of any two of its first s_1 or last s_2 arguments, but is not necessarily antisymmetric under interchange of one of its first s_1 components with one of its last s_2 components.

2.2.7 The Wedge Product, the Interior Product, Dual Tensors

It would thus be nice to define a product under which the set of all forms becomes a closed algebra. Such a product is called the *wedge product*, and is denoted by the symbol $\wedge$. We motivate its definition as follows: The wedge product of a scalar (zero form) with any s-form is the s-form obtained by simple multiplication by the scalar. The wedge product of two one forms, $\mathbf{a}^1$ and $\mathbf{a}^2$, is the two form given by

$$\mathbf{a}^1 \wedge \mathbf{a}^2 = \mathbf{a}^1 \otimes \mathbf{a}^2 - \mathbf{a}^2 \otimes \mathbf{a}^1. \tag{2.55}$$

It is clear that the two form thus obtained is antisymmetric. For three or more one forms, we demand that the wedge product be associative, so, for example

$$\begin{aligned} \mathbf{a}^1 \wedge \mathbf{a}^2 \wedge \mathbf{a}^3 &= \mathbf{a}^1 \wedge (\mathbf{a}^2 \wedge \mathbf{a}^3) \\ &= (\mathbf{a}^1 \wedge \mathbf{a}^2) \wedge \mathbf{a}^3 \\ &= \mathbf{a}^1 \otimes \mathbf{a}^2 \otimes \mathbf{a}^3 + \mathbf{a}^2 \otimes \mathbf{a}^3 \otimes \mathbf{a}^1 + \mathbf{a}^3 \otimes \mathbf{a}^1 \otimes \mathbf{a}^2 \\ &\quad - \mathbf{a}^3 \otimes \mathbf{a}^2 \otimes \mathbf{a}^1 - \mathbf{a}^1 \otimes \mathbf{a}^3 \otimes \mathbf{a}^2 - \mathbf{a}^2 \otimes \mathbf{a}^1 \otimes \mathbf{a}^3. \end{aligned} \tag{2.56}$$

Recall that the total number of independent components of an s-form in a space of dimension n is given by $\begin{pmatrix} n \\ s \end{pmatrix}$. A moment's thought convinces one that the vector space of all such s-forms is spanned by the $\begin{pmatrix} n \\ s \end{pmatrix}$ independent basis s-forms

$$\tilde{\omega}^{\beta_1} \wedge \ldots \wedge \tilde{\omega}^{\beta_s}, \tag{2.57}$$

where the β indices range from 1 to n, and must all be different (else the above expression will vanish). It is then straightforward to see that an arbitrary s-form, $\mathbf{\Omega}$, is given by

$$\Omega = \frac{1}{s!}\Omega_{\beta_1\ldots\beta_s}\tilde{\omega}^{\beta_1} \wedge \ldots \wedge \tilde{\omega}^{\beta_s}. \tag{2.58}$$

Note that the factor of $s!$ appears here because we did not put it into the definition of the wedge product; we could have done it either way, and authors differ in this convention.

Now that we know how the wedge product operates on scalars and one forms, we can extend its definition to arbitrary forms by writing them in terms of wedge products of basis one forms, as shown above. This makes the 2^n-dimensional vector space of all forms into an algebra, called a *Grassmann algebra*. Note that it is not a commutative algebra: If Ω^1 and Ω^2 are s_1 and s_2-forms, respectively, then

$$\Omega_1 \wedge \Omega_2 = (-1)^{s_1 s_2}\Omega_2 \wedge \Omega_1. \tag{2.59}$$

If we contract the first index of an s-form (where $s \geq 1$), Ω, with a vector, $\mathbf{V}$, then it is straightforward to see that we get an $(s-1)$-form. We call this new form the *interior product* of Ω with $\mathbf{V}$, and we denote it by $i_V\Omega$. Thus

$$i_V\Omega = \frac{1}{(s-1)!}V^{\beta_1}\Omega_{\beta_1\ldots\beta_s}\tilde{\omega}^{\beta_2} \wedge \ldots \wedge \tilde{\omega}^{\beta_s}. \tag{2.60}$$

If Ω^1 and Ω^2 are s_1 and s_2-forms, respectively, then it is straightforward to show

$$i_V(\Omega^1 \wedge \Omega^2) = (i_V\Omega^1) \wedge \Omega^2 + (-1)^{s_1}\Omega^1 \wedge (i_V\Omega^2). \tag{2.61}$$

Also, the antisymmetry of forms makes it clear that

$$i_V i_V \Omega = 0 \tag{2.62}$$

for any n-form Ω with $n \geq 2$.

Recall that we defined an s-form as a completely antisymmetric tensor of type $(0, s)$. Note that we could have done the same thing for completely antisymmetric

tensors of type $(s,0)$. Next note that a completely antisymmetric tensor of type $(0,s)$ has exactly the same number of components as a completely antisymmetric tensor of type $(n-s,0)$ in a space of dimension $n \geq s$. This is because

$$\begin{pmatrix} n \\ s \end{pmatrix} = \begin{pmatrix} n \\ n-s \end{pmatrix}. \tag{2.63}$$

This suggests that there may be a one-to-one correspondence between s-forms and completely antisymmetric tensors of type $(n-s,0)$.

For example, note that there is only one independent component of a completely antisymmetric tensor of type $(n,0)$. This is because the components of such a tensor must be proportional to those of the Levi-Civita symbol, $\epsilon^{\beta_1\ldots\beta_n}$. The proportionality constant is a scalar (zero form). Similarly, we can put any scalar (zero form) in front of the Levi-Civita symbol, and obtain the components of a completely antisymmetric tensor of type $(n,0)$. Thus, there is a one-to-one correspondence between scalars (zero forms) and completely antisymmetric tensors of type $(n,0)$.

More generally, we can use the Levi-Civita symbol to obtain a one-to-one correspondence between s-forms and completely antisymmetric tensors of type $(n-s,0)$ as follows:

$$T^{\beta_1\ldots\beta_{n-s}} = \frac{1}{s!}\epsilon^{\beta_1\ldots\beta_n}\Omega_{\beta_{n-s+1}\ldots\beta_n}, \tag{2.64}$$

and

$$\Omega_{\beta_{n-s+1}\ldots\beta_n} = \frac{1}{(n-s)!}\epsilon_{\beta_1\ldots\beta_n}T^{\beta_1\ldots\beta_n}. \tag{2.65}$$

Here we have used the easily verified relation

$$\epsilon^{\beta_1\ldots\beta_n}\epsilon_{\beta_1\ldots\beta_n} = n!. \tag{2.66}$$

Referring to Eq. (2.64), we say that $\mathbf{T}$ is *dual* to $\mathbf{\Omega}$ with respect to ϵ. This is often abbreviated $\mathbf{T} = {}^*\mathbf{\Omega}$. Referring to Eq. (2.65), we say that $\mathbf{\Omega}$ is *dual* to $\mathbf{T}$ with respect to ϵ, or $\mathbf{\Omega} = {}^*\mathbf{T}$. Note that for any form, $\mathbf{\Omega}$, we have ${}^{**}\mathbf{\Omega} = (-1)^{s(n-s)}\mathbf{\Omega}$.

2.2.8 The Exterior Derivative and the Homotopy Formula

We now define a differential operator, d, that converts s-forms into $(s+1)$-forms. This operator is defined as follows: When applied to a scalar (zero form), f, it yields the one form, df, such that

$$df(\mathbf{V}) = \hat{\mathbf{V}} f. \tag{2.67}$$

Thus, in a coordinate basis, z^α, we have

$$df = \frac{\partial f}{\partial z^\alpha} dz^\alpha. \tag{2.68}$$

Next, we demand that the operator be linear, so if $\mathbf{\Omega}$ and $\mathbf{\Lambda}$ are two s-forms then

$$d(\mathbf{\Omega} + \mathbf{\Lambda}) = d\mathbf{\Omega} + d\mathbf{\Lambda}. \tag{2.69}$$

Next, we demand that if $\mathbf{\Omega}_1$ is an s_1-form and $\mathbf{\Omega}_2$ is an s_2-form,

$$d(\mathbf{\Omega}_1 \wedge \mathbf{\Omega}_2) = d\mathbf{\Omega}_1 \wedge \mathbf{\Omega}_2 + (-1)^{s_1} \mathbf{\Omega}_1 \wedge d\mathbf{\Omega}_2. \tag{2.70}$$

Finally, we demand that for any s-form, $\mathbf{\Omega}$, we have

$$dd\mathbf{\Omega} = 0. \tag{2.71}$$

The above demands define the operator d uniquely and unambiguously. We can apply the exterior derivative to an arbitrary form by first expanding it in terms of wedge products of basis one forms, and then applying the above rules.

In terms of components in a coordinate basis, the exterior derivative of a scalar is

$$(df)_\alpha = f_{,\alpha}, \tag{2.72}$$

and the exterior derivative of a one form is

$$(d\mathbf{a})_{\alpha\beta} = a_{\beta,\alpha} - a_{\alpha,\beta}. \tag{2.73}$$

More generally,

$$(d\mathbf{\Omega})_{\alpha_1\ldots\alpha_{s+1}} = \frac{(-1)^s}{s!}\epsilon_{\beta_1\ldots\beta_{s+1}}\Omega_{\alpha_{\beta_1}\ldots\alpha_{\beta_s},\alpha_{\beta_{s+1}}}. \tag{2.74}$$

A form whose exterior derivative vanishes is said to be *closed.* A form that is the exterior derivative of another form is said to be *exact.* Clearly, any exact form is closed. The interesting question is whether or not any closed form is exact. The answer to this depends on the global topology of the manifold on which the closed form lives. Locally, it is always true.

There is a marvelous relationship between Lie derivatives, interior products, and exterior derivatives. It is possible to prove that

$$\mathcal{L}_V\Omega = i_V d\Omega + d i_V\Omega \tag{2.75}$$

for any n-form, Ω, with $n \geq 1$, and any vector field, $\mathbf{V}$. This relationship is called the *homotopy formula.* The proof usually given (see for example Section 4.20 of Schutz [15]) proceeds by induction: It is first proved for a one-form, and then it is shown that it works for an n-form if it works for an $(n-1)$-form.

The *generalized homotopy formula,*

$$\mathcal{L}_V^j\Omega = (i_V d)^j\Omega + (d i_V)^j\Omega \tag{2.76}$$

for $j \geq 1$, is proved by induction as follows: First note that it reduces to the ordinary homotopy formula when $j = 1$. Next, assume that it is true for $j = l$. Then

$$\begin{aligned}\mathcal{L}_V^{l+1}\Omega &= (i_V d + d i_V)[(i_V d)^l\Omega + (d i_V)^l\Omega] \\ &= (i_V d)^{l+1}\Omega + (d i_V)^{l+1}\Omega,\end{aligned} \tag{2.77}$$

where we have used the fact that application of dd or $i_V i_V$ causes any form to vanish. Note that the generalized homotopy formula is not true for $j = 0$.

Finally, we can show that Lie derivatives commute with exterior derivatives. This is done as follows:

$$d\mathcal{L}_V = d(i_V d + d i_V) = d i_V d = (i_V d + d i_V) d = \mathcal{L}_V d, \tag{2.78}$$

where we have used the homotopy formula.

2.2.9 Integration on Manifolds

Differential s-forms can also be introduced as integrands of s-dimensional integrals. See Flanders [19] for more on this approach. Adopting this point of view, it is possible to prove the generalized *Stokes' theorem*

$$\int_U d\mathbf{\Omega} = \int_{\partial U} \mathbf{\Omega}, \tag{2.79}$$

where U is an $(s+1)$-dimensional volume, and ∂U is the s-dimensional surface that bounds it.

We shall not attempt to prove the generalized Stokes' theorem here (see Schutz [15] for a good presentation), but we shall make it plausible by showing how it reduces to the familiar Stokes' theorem and divergence theorem of three dimensional vector calculus. In three dimensional Euclidean space, with Cartesian coordinates, the gradient is given in our notation by

$$(\vec{\nabla} f)_i = (df)_i, \tag{2.80}$$

the divergence is given by

$$\vec{\nabla} \cdot \mathbf{V} = {}^*d{}^*\mathbf{V}, \tag{2.81}$$

and the curl is given by

$$\vec{\nabla} \times \mathbf{V} = {}^*d\mathbf{v}, \tag{2.82}$$

where $\mathbf{v}$ is the one form whose Cartesian components are identical to those of the vector $\mathbf{V}$. Note that

$$\vec{\nabla} \times \vec{\nabla} f = {}^*ddf = 0 \tag{2.83}$$

and

$$\vec{\nabla} \cdot (\vec{\nabla} \times \mathbf{V}) = {}^*d{}^*d\mathbf{v} = {}^*dd\mathbf{v} = 0 \tag{2.84}$$

both follow from $dd = 0$. Then

$$\int_U \vec{\nabla} \cdot \mathbf{V} dx^3 = \int_U d^*\mathbf{V} = \int_{\partial U} {}^*\mathbf{V} = \int_{\partial U} \mathbf{V} \cdot d\sigma, \tag{2.85}$$

and

$$\int_U \vec{\nabla} \times \mathbf{V} \cdot d\sigma = \int_U d\mathbf{v} = \int_{\partial U} \mathbf{v} = \int_{\partial U} \mathbf{V} \cdot d\ell. \tag{2.86}$$

Thus we see that our formalism is the natural generalization of three dimensional vector calculus to manifolds of arbitrary dimension.

2.2.10 Metric Spaces

The usual *dot product* of linear algebra is a rule for taking two vectors, say $\mathbf{U}$ and $\mathbf{V}$, and associating with them a real number, denoted $\mathbf{U} \cdot \mathbf{V}$. The result depends bilinearly on the two vectors involved, so we see that there is a tensor of type $(0, 2)$ at work here. Furthermore, the dot product is required to be commutative, so the tensor must be symmetric. Denoting this tensor by $\mathbf{g}$, we have

$$\mathbf{U} \cdot \mathbf{V} = \mathbf{g}(\mathbf{U}, \mathbf{V}). \tag{2.87}$$

This tensor is called the *metric tensor*. If we also demand that it have an inverse, then we can find a basis for which it has diagonal form with entries equal to ± 1 (if all the diagonal entries can be made equal to $+1$, then we say that the metric is *definite*, otherwise we say that it is *indefinite*). The trace of the metric in this canonical diagonal form is called its *signature*.

A *metric tensor field* is the association of such a type $(0, 2)$ symmetric tensor with every point of a manifold. It must have an inverse at every point. It follows that the signature is the same at every point of the manifold.

A manifold endowed with a metric has all sorts of new structure. For the purposes of our discussion, its most important role is to provide a one-to-one

correspondence between vectors and covectors. For, given any vector, say $\mathbf{V}$, we can form the covector, $\mathbf{g}(\mathbf{V}, \quad)$. The components of this new covector are then $g_{\alpha\beta}V^{\beta}$. Denote the inverse of $g_{\alpha\beta}$ by $g^{\beta\gamma}$, so

$$g_{\alpha\beta}g^{\beta\gamma} = \delta^{\gamma}_{\alpha}. \tag{2.88}$$

Then, given any covector, say $\mathbf{a}$, we can form the vector with components $g^{\alpha\beta}a_{\beta}$. Note that this is a one to one correspondence.

Frequently we shall use the same symbol to denote a vector and its corresponding covector in a metric space. That is, we may write

$$V_{\alpha} = g_{\alpha\beta}V^{\beta}, \tag{2.89}$$

or

$$V^{\alpha} = g^{\alpha\beta}V_{\beta}. \tag{2.90}$$

This process is called *index raising* or *index lowering*, as the case may be. It can be used to raise or lower the indices of any tensor of any type.

We shall frequently abuse notation by using the dot product to denote the interior product of a vector with a covector. That is, we may write

$$\mathbf{a} \cdot \mathbf{V} = \mathbf{a}(\mathbf{V}) = a_{\alpha}V^{\alpha}. \tag{2.91}$$

When this is done, it will be obvious from context, so no confusion should arise. We shall occasionally further abuse notation by using a "double dot" notation for two contracted indices. That is, given two tensors of type $(0,2)$ and $(2,0)$, respectively, we may write

$$\mathbf{F} : \mathbf{G} = F_{\alpha\beta}G^{\alpha\beta}. \tag{2.92}$$

Once again, things should be clear from context.

A metric tensor field does far more than provide an invertible map from vectors to covectors. It also induces on the manifold something called an *affine*

connection. This makes it possible to compare vectors in nearby tangent spaces. Recall that Lie dragging gave us a way to do this, but there had to be a vector field present in the first place along which to drag, and we could drag only in the direction of that field. An affine connection allows us to *parallel transport* vectors from one tangent space to any other one nearby; that is, it gives us a notion of parallelism between vectors in different tangent spaces. Furthermore, it does not require the presence of any vector field there to begin with. One does not need a metric to have an affine connection, but the presence of a metric induces an affine connection in a natural way.

Armed with an affine connection, it is possible to go on to define such things as curvature and torsion. While knowledge of this material is certainly helpful in understanding the material presented in this thesis (especially the curvature and polarization guiding-center drifts and the intimate relationship between torsion and spin angular momentum), it is not essential. Thus we shall not go on to discuss these topics; the interested reader is referred to Schutz [15] for a good introduction, and to Misner, Thorne and Wheeler [20] or Chandrasekhar [21] for a more detailed presentation.

2.3 Noncanonical Hamiltonian and Lagrangian Mechanics

2.3.1 Canonical Versus Noncanonical Coordinates

In elementary classical mechanics courses, Hamiltonian mechanics is derived by application of a Legendre transformation to the system Lagrangian. This process gives rise to canonical coordinates in a very natural way. When it becomes necessary to change coordinates on phase space, the student is taught to restrict attention to the limited class of transformations that will maintain this separa-

tion of the coordinates into canonically conjugate pairs; these are the so-called canonical transformations.

The Poisson bracket of two scalar phase functions, A and B, is then introduced by *defining* it in terms of partial derivatives with respect to the canonically conjugate pairs of coordinates, q^i and p_i (the index i ranges over all the degrees of freedom):

$$\{A, B\} = \frac{\partial A}{\partial q^i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q^i}, \tag{2.93}$$

where we have adopted the convention of summation over repeated indices. It is then shown that this bracket is bilinear:

$$\{xA + yB, C\} = x\{A, C\} + y\{B, C\} \tag{2.94}$$

where x and y are constants, that it is antisymmetric:

$$\{A, B\} = -\{B, A\}, \tag{2.95}$$

that it obeys the Jacobi identity:

$$\{A, \{B, C\}\} + \{C, \{A, B\}\} + \{B, \{C, A\}\} = 0, \tag{2.96}$$

and that it obeys the chain rule:

$$\{f(A), C\} = f'(A)\{A, C\} \tag{2.97}$$

or, equivalently, the Leibniz product rule:

$$\{AB, C\} = A\{B, C\} + B\{A, C\}. \tag{2.98}$$

Mathematicians have a different way of looking at all of this. In mathematics courses on Hamiltonian mechanics, one is more likely to *define* the Poisson bracket as any rule for taking a pair of scalar phase functions and associating with them a third scalar phase function consistent with the properties listed in Eqs. (2.94) through (2.98) above. Now it is manifest that any Poisson bracket given by the

physicists' definition is also a Poisson bracket according to the mathematicians' definition. The converse, however, is *not* true; that is, there exist Poisson brackets that obey all of the above-listed properties, but are not given by Eq. (2.93) for any set of canonical coordinates, q and p. Thus, by adopting the mathematicians' definition, we can generalize what is meant by a Poisson bracket in a very powerful way.

To see how this comes about, let us take the mathematicians' viewpoint and suppose that we have a phase space with coordinates, z^α, where α ranges from 1 to N. For canonical coordinates, N is twice the number of degrees of freedom and the z^α are the q's and p's, but let us not restrict ourselves to this special case in any way; in particular, N could be an odd number, and there need not be any natural pairing amongst the coordinates.

Denote the Poisson bracket of coordinate z^α with coordinate z^β by:

$$J^{\alpha\beta} \equiv \{z^\alpha, z^\beta\}. \tag{2.99}$$

Suppose that we changed our phase space coordinates, $z \mapsto z'$. Then, using the chain rule, Eq. (2.97), we see that the Poisson bracket of two of the new coordinates is given by:

$$J'^{\alpha\beta} \equiv \{z'^\alpha, z'^\beta\} = \frac{\partial z'^\alpha}{\partial z^\xi}\{z^\xi, z^\eta\}\frac{\partial z'^\beta}{\partial z^\eta} \tag{2.100}$$

or

$$J'^{\alpha\beta} = \frac{\partial z'^\alpha}{\partial z^\xi}\frac{\partial z'^\beta}{\partial z^\eta}J^{\xi\eta}. \tag{2.101}$$

This makes it clear that the $J^{\alpha\beta}$ are the components of a second rank contravariant tensor. This tensor will henceforth be called the *Poisson tensor*. Using the chain rule once again, we see that the Poisson bracket of any two phase functions, A and B, may be written in terms of the Poisson tensor as follows:

$$\{A, B\} = \frac{\partial A}{\partial z^\alpha}J^{\alpha\beta}\frac{\partial B}{\partial z^\beta} \tag{2.102}$$

The general form of the bracket given by Eq. (2.102) is clearly bilinear and obeys the chain rule (or, equivalently, the Leibniz product rule). Now, the other two defining properties of the Poisson bracket may be expressed as properties of the Poisson tensor. It is easily seen that antisymmetry of the bracket implies and is implied by antisymmetry of the Poisson tensor:

$$J^{\alpha\beta} = -J^{\beta\alpha}. \tag{2.103}$$

Somewhat more algebra shows that the Jacobi property of the bracket implies and is implied by the following property of the Poisson tensor:

$$J^{\alpha\xi}J^{\beta\gamma}{}_{,\xi} + J^{\gamma\xi}J^{\alpha\beta}{}_{,\xi} + J^{\beta\xi}J^{\gamma\alpha}{}_{,\xi} = 0, \tag{2.104}$$

where the commas denote partial differentiation. Thus, our philosophy shall be that any tensor that has these two properties defines a perfectly legitimate Poisson bracket according to Eq. (2.102).

Let us see how this works for canonical coordinates, q^i and p_i, where i ranges from 1 to the number of degrees of freedom, I. Write $z^\alpha = q^\alpha$ for $\alpha = 1, \ldots, I$, and $z^\alpha = p_{\alpha - I}$ for $\alpha = I+1, \ldots, N$ where $N = 2I$. Now canonical coordinates have the bracket relations, $\{q^i, q^j\} = \{p_i, p_j\} = 0$ and $\{q^i, p_j\} = -\{p_j, q^i\} = \delta^i_j$, so the matrix of components of the Poisson tensor is:

$$\mathbf{J} \equiv \{\mathbf{z}, \mathbf{z}\} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix}, \tag{2.105}$$

where $\mathbf{0}$ and $\mathbf{1}$ are the $I \times I$ null and unit matrices, respectively.

Using this Poisson tensor in Eq. (2.102), we easily recover the usual expression for the canonical bracket, Eq. (2.93). Furthermore, this Poisson tensor is obviously antisymmetric, and it obeys Eq. (2.104) since its components are constants so their derivatives are all zero.

If we start with canonical coordinates, then a *canonical transformation* is any transformation that leaves the Poisson tensor unchanged. If we denote the

Jacobian matrix of the transformation by:

$$\mathbf{M} \equiv \partial \mathbf{z}'/\partial \mathbf{z}, \tag{2.106}$$

then this condition may be written as the matrix equation:

$$\mathbf{J} = \mathbf{M}\mathbf{J}\mathbf{M}^T, \tag{2.107}$$

where the superscript "T" denotes "transpose," and $\mathbf{J}$ is the canonical Poisson tensor given by Eq. (2.105). In what follows, we shall generalize the term *canonical transformation* to mean any bracket-preserving transformation, regardless of whether or not we started from canonical coordinates.

Thus far, we have said nothing about the equations of motion. For canonical coordinates these are well known to be:

$$\dot{q^i} = \frac{\partial H}{\partial p_i},$$

and

$$\dot{p_i} = -\frac{\partial H}{\partial q^i}, \tag{2.108}$$

where H is the Hamiltonian. These may be written in terms of the Poisson bracket as follows:

$$\dot{q^i} = \{q^i, H\},$$

and

$$\dot{p_i} = \{p_i, H\}. \tag{2.109}$$

If we use z to refer to the q's and p's, this becomes even simpler to write:

$$\dot{z^\alpha} = \{z^\alpha, H\}. \tag{2.110}$$

Alternatively, this may be written in terms of the Poisson tensor:

$$\dot{z^\alpha} = J^{\alpha\beta}\frac{\partial H}{\partial z^\beta}. \tag{2.111}$$

Since this last equation is in tensor form, and since it is known to hold for canonical coordinates, it must be the correct generalization of the equation of motion for noncanonical coordinates. Thus, the complete specification of a Hamiltonian system in this new generalized sense requires the specification of both a Poisson tensor and a scalar Hamiltonian.

Any dynamical system on phase space can be expressed in the form $\dot{z}^\alpha = V^\alpha$, where $\mathbf{V}$ is some vector field on the phase space. Eq. (2.111) for a Hamiltonian dynamical system has this form. Note, however, that in order to qualify as "Hamiltonian," the vector field on the right cannot be just any vector field; it must be given by the Poisson tensor contracted with the gradient of some scalar function. A vector field on phase space is called a *Hamiltonian vector field* if there exists some scalar field for which this is true. Thus, if a manifold is endowed with a Poisson tensor, then scalar fields generate Hamiltonian vector fields.

2.3.2 An Example of a Noncanonical Poisson Structure

There are several ways that noncanonical Poisson structures can arise in a problem. The first and most obvious way is to start with canonical coordinates and make a noncanonical transformation. The canonical Poisson tensor is known to obey Eqs. (2.103) and (2.104), and since these are tensorial equations they will hold in all frames if they hold in any one frame. So the result of a noncanonical transformation will be a new bracket that obeys all the required properties.

A particularly beautiful example of this has been given by Littlejohn [22] for the problem of a charged particle in a magnetic field. For canonical coordinates, $\mathbf{q}$ and $\mathbf{p}$, the Hamiltonian is well known to be:

$$H = \frac{1}{2m}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{q})\right)^2, \tag{2.112}$$

where $\mathbf{A}(\mathbf{q})$ is the vector potential. Make the noncanonical transformation to

new coordinates,$\mathbf{r}$ and $\mathbf{v}$, where:

$$\mathbf{r} \equiv \mathbf{q}$$

and

$$\mathbf{v} \equiv \frac{1}{m}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{q})\right). \tag{2.113}$$

The bracket relations among the new coordinates are easily calculated:

$$\{\mathbf{r}, \mathbf{r}\} = \mathbf{0},$$
$$\{\mathbf{r}, \mathbf{v}\} = \frac{1}{m}\mathbf{1},$$

and

$$\{\mathbf{v}, \mathbf{v}\} = \frac{1}{m}\mathbf{\Omega}, \tag{2.114}$$

where we have defined the matrix $\mathbf{\Omega}$ with components:

$$\Omega_{ij} \equiv \frac{e}{mc}\left(A_{j,i} - A_{i,j}\right) = \frac{e}{mc}\epsilon_{ijk}B^k, \tag{2.115}$$

and where the B^k are the components of the ordinary magnetic field pseudovector. Thus the bracket of any two scalar phase functions, R and S, is given by:

$$\{R, S\} = \frac{1}{m}\left(\frac{\partial R}{\partial \mathbf{r}} \cdot \frac{\partial S}{\partial \mathbf{v}} - \frac{\partial R}{\partial \mathbf{v}} \cdot \frac{\partial S}{\partial \mathbf{r}}\right) + \frac{e}{m^2 c}\mathbf{B} \cdot \left(\frac{\partial R}{\partial \mathbf{v}} \times \frac{\partial S}{\partial \mathbf{v}}\right). \tag{2.116}$$

This bracket is easily seen to be antisymmetric. That it satisfies the Jacobi identity is less obvious; we know that it must from the arguments given above, but a direct proof involves some tedious algebra. The new Hamiltonian is simply:

$$H(\mathbf{r}, \mathbf{v}) = \frac{m}{2}v^2, \tag{2.117}$$

and it is readily verified that this Hamiltonian, together with the bracket given in Eq. (2.116) yield the correct equations of motion. Note that the vector potential is absent from the new formulation; this is construed as an advantage, since the vector potential is a gauge-dependent quantity. The above Hamiltonian system was the starting point for Littlejohn's work on guiding-center theory [22].

Now that we have seen how noncanonical Poisson structures can arise from noncanonical transformations of a canonical system, it is natural to ask the opposite question: Given a noncanonical Hamiltonian system, is it always possible to find a transformation to canonical coordinates? For noncanonical Hamiltonian systems with a nonsingular Poisson tensor (that is, systems for which the matrix of components of the Poisson tensor is nonsingular), there is an important theorem, called Darboux's theorem, that tells us that the answer is "yes." A proof of Darboux's theorem is given by Littlejohn [22] and is constructive; that is, it gives a prescription for actually finding the transformation to canonical coordinates. For Hamiltonian systems with singular Poisson structures, the situation is more complicated, and will be discussed shortly.

2.3.3 Reduction

Reduction and Noether's Theorem

Noncanonical transformations from canonical coordinates is only one of many ways that interesting Poisson structures can arise naturally. The process of "reduction" of a Hamiltonian system with symmetry is another. Work in this area has been pioneered by Marsden and Weinstein (see, for example, reference [23]).

A detailed discussion of reduction would be out of place in this work, but the general idea is this: Suppose that we have a canonical Hamiltonian system with a configuration space symmetry (e.g. spatial translation, rotation, etc.). Make the configuration space symmetry group parameter one of the generalized coordinates. Noether's theorem then tells us that the corresponding momentum is conserved. It is then possible to eliminate this degree of freedom from the system, thus reducing the dimensionality of the phase space by two. This much is familiar from elementary courses in classical mechanics. Reduction is an important generalization of Noether's theorem that allows us to similarly "mod out" by a

symmetry group that acts on all of phase space rather than just configuration space. After reduction is performed, the resulting Hamiltonian system may very well be noncanonical.

The set of all phase functions together with the Poisson bracket operation constitutes a Lie algebra. From a computational point of view, in order to perform reduction we must find a representation for which this Lie algebra has a closed Lie subalgebra. Furthermore, the Hamiltonian must depend only on the elements of this subalgebra. The elements of the subalgebra then constitute coordinates for a reduced description of the problem. This is best illustrated by example.

The Free Rigid Body

One of the most elementary (but nontrivial) examples of this process is the Hamiltonian system for a free rigid body. The usual generalized coordinates for this problem are the Eulerian angles, θ, ϕ, and ψ, with respect to some fixed space frame. By introducing their canonically conjugate momenta, p_θ, p_ϕ, and p_ψ, it is possible to write the equations of motion in a canonical Hamiltonian format with a six-dimensional phase space. If we choose a body frame for which the inertia tensor is diagonalized, then the Hamiltonian for the free rigid body problem is

$$H = \frac{p_\psi^2}{2I_3} + \frac{1}{2I_2}\left[(p_\phi \csc\theta - p_\psi \cot\theta)\cos\psi - p_\theta \sin\psi\right]^2 + \frac{1}{2I_1}\left[(p_\phi \csc\theta - p_\psi \cot\theta)\sin\psi + p_\theta \cos\psi\right]^2, \tag{2.118}$$

where I_1, I_2, and I_3 are the three diagonal elements of the inertia tensor.

Consider the three components of the angular momentum resolved in the body frame. These can be expressed in terms of our canonical phase space coordinates as follows:

$$m_1 = (p_\phi \csc\theta - p_\psi \cot\theta)\sin\psi + p_\theta \cos\psi,$$

$$m_2 = (p_\phi \csc\theta - p_\psi \cot\theta)\cos\psi - p_\theta \sin\psi,$$

and

$$m_3 = p_\psi. \tag{2.119}$$

(See Goldstein [24] for details. Only the result is needed here.)

By direct calculation with the canonical bracket, we can verify the following relations

$$\begin{aligned} \{m_1, m_2\} &= -m_3 \\ \{m_2, m_3\} &= -m_1 \\ \{m_3, m_1\} &= -m_2. \end{aligned} \tag{2.120}$$

Thus, the three components of the angular momentum in the body frame constitute a closed Lie subalgebra under the operation of the canonical Poisson bracket. This means that the subset of functions on the canonical phase space that are functions of the m's alone (that is, those functions that depend on θ, ϕ, ψ, p_θ, p_ϕ, and p_ψ *only* through their dependence on the m's) constitutes a Lie subalgebra of the Lie algebra of all canonical phase functions.

We thus adopt the m's as generalized coordinates on a *reduced* phase space of three dimensions. The Poisson tensor on this reduced phase space is then given by $J^{\alpha\beta} = -\epsilon^{\alpha\beta\gamma} m_\gamma$, or:

$$\mathbf{J} = \begin{pmatrix} 0 & -m_3 & m_2 \\ m_3 & 0 & -m_1 \\ -m_2 & m_1 & 0 \end{pmatrix}, \tag{2.121}$$

so that the Poisson bracket of any two functions of $\mathbf{m}$, say A and B, is given by:

$$\{A, B\} = -\mathbf{m} \cdot \left(\frac{\partial A}{\partial \mathbf{m}} \times \frac{\partial B}{\partial \mathbf{m}} \right). \tag{2.122}$$

This bracket *must* satisfy all the required properties of a Poisson bracket, since it was derived by specializing the domain of a canonical bracket; nevertheless, it is straightforward and instructive to verify this by direct calculation.

It is possible to perform reduction only if the Hamiltonian is expressible in terms of the reduced coordinate set. For the free rigid body, we have

$$H(\mathbf{m}) = \frac{m_1^2}{2I_1} + \frac{m_2^2}{2I_2} + \frac{m_3^2}{2I_3}. \tag{2.123}$$

As usual, the equations of motion are given by $\dot{\mathbf{m}} = \{\mathbf{m}, H\}$, or:

$$\begin{aligned} \dot{m}_1 &= \left(\frac{1}{I_3} - \frac{1}{I_2}\right) m_2 m_3 \\ \dot{m}_2 &= \left(\frac{1}{I_1} - \frac{1}{I_3}\right) m_3 m_1 \\ \dot{m}_3 &= \left(\frac{1}{I_2} - \frac{1}{I_1}\right) m_1 m_2. \end{aligned} \tag{2.124}$$

As expected, these are indeed Euler's equations for the free rigid body. If the rigid body were not free (say, if it were in a gravitational field), then a potential energy term would have been present in the Hamiltonian, and that term would *not* have been expressible in terms of the m's. Thus, the reduction process would have failed. This is because the gravitational field breaks the SO(3) symmetry that makes the reduction possible.

As we shall see later on in this thesis, the passage from particle coordinates to guiding-center coordinates is another example of reduction. The symmetry involved is the group of rotations by the gyroangle, SO(2), and the reduction eliminates the corresponding degree of freedom from the system. If this gyrosymmetry is somehow broken (say, by a variation in the background field configuration whose length scale is on the order of a gyroradius), then the guiding-center description is invalidated.

Euler's Fluid Equations

Our next example is a Hamiltonian field theory for Euler's equations for the flow of an inviscid, incompressible fluid. Let us adopt a Lagrangian description for such a fluid wherein each fluid particle is labelled by a reference position,

$\mathbf{x}_0$. Then the configuration of the fluid at time t may be specified by giving the particle's current position, $\mathbf{x}$ as a function of $\mathbf{x}_0$ and t. Thus, our dynamical field variable is $\mathbf{x}(\mathbf{x}_0, t)$. The system Lagrangian consists solely of the kinetic energy

$$L = \int d^3x_0 \frac{\rho}{2} \dot{x}^2(\mathbf{x}_0, t), \tag{2.125}$$

where ρ is the constant uniform mass density. The canonical momentum field is then given by

$$\mathbf{p}(\mathbf{x}_0, t) = \frac{\delta L}{\delta \dot{\mathbf{x}}(\mathbf{x}_0, t)} = \rho \dot{\mathbf{x}}(\mathbf{x}_0, t), \tag{2.126}$$

where the δ's denote functional differentiation. Performing the Legendre transformation, we see that the system Hamiltonian is

$$H = \int d^3x_0 \frac{1}{2\rho} \mathbf{p}^2(\mathbf{x}_0, t). \tag{2.127}$$

The canonical bracket of two functionals of $\mathbf{x}$ and $\mathbf{p}$, say A and B, is then

$$\{A, B\} = \int d^3x_0 \left(\frac{\delta A}{\delta \mathbf{x}(\mathbf{x}_0, t)} \cdot \frac{\delta B}{\delta \mathbf{p}(\mathbf{x}_0, t)} - \frac{\delta A}{\delta \mathbf{p}(\mathbf{x}_0, t)} \cdot \frac{\delta B}{\delta \mathbf{x}(\mathbf{x}_0, t)} \right). \tag{2.128}$$

Now suppose that the fluid particles are identical. In that case, specification of $\mathbf{x}(\mathbf{x}_0, t)$ is far more information than is really necessary to determine the configuration of the fluid. This is because $\mathbf{x}(\mathbf{x}_0, t)$ effectively keeps track of particle labels; two configurations that differ only by swapping identical particles will actually have different $\mathbf{x}(\mathbf{x}_0, t)$. For a fluid of identical particles, an Eulerian description, wherein the flow velocity is given as a function of spatial position and time, say $\mathbf{v}(\xi, t)$, suffices to determine the fluid configuration. The Lagrangian description just keeps track of too much information. Thus, in passing from the Lagrangian to the Eulerian description, we are effectively reducing by the group of identical particle interchanges. The Eulerian description is therefore the reduced description. The reduced phase space is the (smaller, though still infinite dimensional) space of all divergenceless vector fields, $\mathbf{v}$, that satisfy the boundary

conditions ($\mathbf{v}$ tangential to the boundary). The requirement that $\mathrm{div}\mathbf{v} = 0$ stems from the fact that we are considering only incompressible flows.

So, from a computational point of view, how do we perform this reduction? Note that the Eulerian velocity field may be written in terms of the Lagrangian fields as follows:

$$\mathbf{v}(\xi, t) = \frac{1}{\rho}\mathbf{p}(\mathbf{x}^{-1}(\xi, t), t). \tag{2.129}$$

This may be interpreted as follows: If we want the Eulerian velocity at spatial point ξ, first take $\mathbf{x}^{-1}(\xi, t)$ to get the reference position of the fluid element currently at ξ, then evaluate the momentum $\mathbf{p}$ of the fluid element with this reference position, then divide the result by ρ to get the desired answer. Now the above equation may be written

$$\mathbf{v}(\xi, t) = \frac{1}{\rho}\int d^3x_0 \mathbf{p}(\mathbf{x}_0, t)\delta(\mathbf{x}(\mathbf{x}_0, t) - \xi), \tag{2.130}$$

where we have used the fact that the Jacobian, $|\partial\mathbf{x}/\partial\mathbf{x}_0|$, is equal to unity because the flow is incompressible. Thus we have succeeded in expressing the reduced field variable, $\mathbf{v}$, in terms of the canonical field variables, $\mathbf{x}$ and $\mathbf{p}$. In this respect, Eq. (2.130) is the exact analog of Eqs. (2.119) for the free rigid body problem.

Thus, we can take the Poisson bracket of the Eulerian field with itself using the canonical bracket. This is straightforward, and the result is

$$\{\mathbf{v}(\xi, t), \mathbf{v}(\xi', t)\} = \frac{1}{\rho}\left(\mathbf{v}(\xi', t)\delta'(\xi' - \xi) - \delta'(\xi - \xi')\mathbf{v}(\xi, t)\right), \tag{2.131}$$

where δ' denotes the gradient of the delta function. Note that we have been able to express this bracket in terms of the Eulerian (reduced) field variables alone. This equation is thus the analog of Eqs. (2.120) for the free rigid body problem.

So we see that the functionals of the Eulerian field variables constitute a closed Lie subalgebra of the Lie algebra of all phase functionals. We thus adopt the Eulerian field variables as coordinates on a reduced phase space. The Poisson bracket of any two functionals of $\mathbf{v}$, say A and B, is then calculated by the Leibniz

rule

$$\{A,B\} = \int d^3\xi \int d^3\xi' \frac{\delta A}{\delta \mathbf{v}(\xi,t)} \cdot \{\mathbf{v}(\xi,t), \mathbf{v}(\xi',t)\} \cdot \frac{\delta B}{\delta \mathbf{v}(\xi',t)}$$
$$= -\frac{1}{\rho} \int d^3\xi \mathbf{v}(\xi,t) \cdot \left[\frac{\delta A}{\delta \mathbf{v}(\xi,t)}, \frac{\delta B}{\delta \mathbf{v}(\xi,t)}\right], \tag{2.132}$$

where the square brackets are Lie brackets, and where $\delta A/\delta \mathbf{v}(\xi,t)$ and $\delta B/\delta \mathbf{v}(\xi,t)$ are regarded as vector fields.

Note that Eq. (2.132) is the analog of Eq. (2.122) for the free rigid body problem.

We must also check that the Hamiltonian may be expressed in terms of the reduced variables. Fortunately, this is not difficult. A change of variables in Eq. (2.127) gives

$$H = \frac{\rho}{2} \int d^3\xi v^2(\xi,t), \tag{2.133}$$

where we have again made use of the fact that the Jacobian, $|\partial \mathbf{x}/\partial \mathbf{x}_0|$, is equal to unity.

It remains to check that the Hamiltonian in Eq. (2.133) together with the bracket in Eq. (2.132) actually yield Euler's fluid equations. This is slightly tricky. Consider a functional $A(\mathbf{v})$. Its equation of motion is

$$\frac{\partial A}{\partial t} = \{A,H\}. \tag{2.134}$$

We insert Eq. (2.133) for the Hamiltonian. After some straightforward manipulation, including an integration by parts where the surface term vanishes due to the boundary condition, we get

$$0 = \int d^3\xi \frac{\delta A}{\delta \mathbf{v}} \cdot \left[\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \vec{\nabla}\, \mathbf{v} + \vec{\nabla}\left(\frac{v^2}{2}\right)\right]. \tag{2.135}$$

At this point, we might be tempted to set the expression in square brackets above equal to zero on the grounds that A is an arbitrary functional. This would, however, be incorrect because $\delta A/\delta \mathbf{v}$ is not really arbitrary. Recall that our phase

space consists only of those vector fields that have zero divergence. This causes an ambiguity in the usual definition of the functional derivative which is such that the equation

$$A(\mathbf{v}+\delta\mathbf{v}) = A(\mathbf{v}) + \int d^3\xi \delta\mathbf{v} \cdot \frac{\delta A}{\delta \mathbf{v}} + \mathcal{O}(\delta v^2) \tag{2.136}$$

is satisfied. If $\mathbf{v}$ and $\mathbf{v}+\delta\mathbf{v}$ are both divergenceless, it follows that $\delta\mathbf{v}$ is divergenceless. This means that the gradient of an arbitrary function, ϕ, may be added to $\delta A/\delta\mathbf{v}$, since

$$\int d^3\xi \delta\mathbf{v} \cdot \vec{\nabla}\, \phi = -\int d^3\xi \phi\, \vec{\nabla} \cdot (\delta\mathbf{v}) = 0. \tag{2.137}$$

We can make the definition of the functional derivative unique by demanding that $\vec{\nabla} \cdot (\delta A/\delta\mathbf{v}) = 0$. This gives a well-posed problem for the determination of ϕ.

Now, in order to incorporate this constraint that $\vec{\nabla} \cdot (\delta A/\delta\mathbf{v}) = 0$, note that if we were to add the gradient of *any* scalar function, ψ, to the expression in square brackets in Eq. (2.135), the equation would still hold because

$$\int d^3\xi \frac{\delta A}{\delta \mathbf{v}} \cdot \vec{\nabla}\, \psi = -\int d^3\xi \psi\, \vec{\nabla} \cdot \left(\frac{\delta A}{\delta \mathbf{v}}\right) = 0. \tag{2.138}$$

So the most that we can write is

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \vec{\nabla}\, \mathbf{v} + \vec{\nabla}\left(\frac{v^2}{2} + \psi\right) = 0. \tag{2.139}$$

We now identify the pressure

$$p \equiv \rho\left(\frac{v^2}{2} + \psi\right), \tag{2.140}$$

so we finally arrive at Euler's fluid equation

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \vec{\nabla}\, \mathbf{v} = -\frac{1}{\rho}\vec{\nabla}\, p. \tag{2.141}$$

Finally, we note that the pressure is not really arbitrary, but is rather determined by taking the divergence of both sides of Eq. (2.141) to get

$$\nabla^2 p = -\vec{\nabla} \cdot (\mathbf{v} \cdot \vec{\nabla}\, \mathbf{v}), \tag{2.142}$$

and by dotting both sides of Eq. (2.141) with the unit normal to the boundary surface, $\hat{\mathbf{n}}$, to get

$$\frac{\partial p}{\partial n} = -\hat{\mathbf{n}} \cdot (\mathbf{v} \cdot \vec{\nabla}\, \mathbf{v}). \tag{2.143}$$

This constitutes a well-posed Neumann problem for p as a functional of $\mathbf{v}$. Thus, Eq. (2.141), coupled with the constraint of incompressibility, determines both $\mathbf{v}$ and p.

It is intriguing that the equations of motion for both examples considered thus far are named after Euler; one wonders if he knew about the beautiful analogy between them. In fact, the first published reference to this analogy seems to be a 1966 paper of Arnold [25].

The Poisson-Vlasov System

Our final example of reduction is also a Hamiltonian field theory, this time for the Poisson-Vlasov equations of plasma physics. For simplicity, we consider a one-dimensional plasma (the methods are trivially generalized to three dimensions). Once again, we label particles by their initial conditions. This time, however, the flow is in phase space, so the initial conditions are r_0 and p_0, and the present phase space position is r and p. The dynamical fields are thus $r(r_0, p_0, t)$ and $p(r_0, p_0, t)$. We shall use z to refer to the set of coordinates, r and p, and z_0 to refer to the set of initial conditions, r_0 and p_0. The fields may thus be abbreviated $z(z_0, t)$.

The Lagrangian for this system that includes the electrostatic potential energy

of interaction was first written down by Low [26]. It is

$$L = \int dz_0 f(z_0) \left[\frac{m}{2} \dot{r}^2(z_0) - \frac{e^2}{2} \int dz_0' f(z_0') g(r(z_0), r(z_0')) \right]. \tag{2.144}$$

Here we have ignored species labels for simplicity. Also, $f(z_0)$ is the distribution of initial conditions on phase space, and $g(r, r')$ is the Coulomb potential kernel. The canonical momentum field is then

$$\pi(z_0) = \frac{\delta L}{\delta \dot{r}(z_0)} = f(z_0) m \dot{r}(z_0). \tag{2.145}$$

The Hamiltonian is obtained by Legendre transformation

$$H = \int dz_0 \frac{\pi^2(z_0)}{2mf(z_0)} + \frac{e^2}{2} \int dz_0 \int dz_0' f(z_0) f(z_0') g(r(z_0), r(z_0')). \tag{2.146}$$

The bracket is canonical, with r and π canonically conjugate.

Now suppose that the particles are identical. Just as with Euler's fluid equations, it turns out that we can reduce to an Eulerian description. This time, the Eulerian field variable is the usual distribution function on phase space, $\mathsf{f}(Z)$. This may be expressed in terms of the Lagrangian field variables as follows:

$$\mathsf{f}(R, P, t) = \int dz_0 f(z_0) \delta\left(R - r(z_0, t)\right) \delta\left(P - \frac{\pi(z_0, t)}{f(z_0)}\right). \tag{2.147}$$

This is the analog of Eqs. (2.119) and (2.130).

Now we can take the canonical bracket of $f(Z)$ with $f(Z')$. We get

$$\{\mathsf{f}(Z), \mathsf{f}(Z')\} = \int dZ'' \mathsf{f}(Z'') \{\delta(Z - Z''), \delta(Z' - Z'')\}_0'', \tag{2.148}$$

where $\{a, b\}_0''$ denotes the single-particle Poisson bracket of $a(R'', P'')$ with $b(R'', P'')$. Note that we have been able to express the bracket of the Eulerian field variables in terms of the canonical field variables; thus we have achieved the desired reduction. The bracket of any two functionals of f is found by application of the Leibniz rule. The result is

$$\{A, B\} = \int dZ \mathsf{f}(Z) \left\{ \frac{\delta A}{\delta \mathsf{f}(Z)}, \frac{\delta B}{\delta \mathsf{f}(Z)} \right\}_0. \tag{2.149}$$

This form for the bracket was first given by Iwinski and Turski [27], by Morrison who credits it to Kaufman [28], and by Gibbons [29]. A derivation similiar to that above can be found in a paper by Kaufman and Dewar [30].

Finally, we see that the Hamiltonian can be expressed in terms of f as follows

$$H = \int dZ \mathsf{f}(Z) \frac{P^2}{2m} + \frac{e^2}{2} \int dZ \int dZ' \mathsf{f}(Z) \mathsf{f}(Z') g(R, R'). \tag{2.150}$$

It is now readily verified that the above brackets and Hamiltonian yield the Poisson-Vlasov equations of motion,

$$\frac{\partial \mathsf{f}}{\partial t} + \frac{P}{m} \frac{\partial \mathsf{f}}{\partial R} - e \frac{\partial \phi}{\partial R} \frac{\partial \mathsf{f}}{\partial P} = 0, \tag{2.151}$$

where

$$\phi(R) = e \int dZ' \mathsf{f}(Z') g(R, R') \tag{2.152}$$

is the electrostatic potential.

Note the similarity in structure of the brackets for all three of the above examples. For example, all three have a Poisson tensor that is linear in the coordinates used. All are examples of what are called *Lie-Poisson brackets*, and there is a rich mathematical literature on brackets of this sort (see, for example, Marsden [23]).

2.3.4 Singular Poisson Structures

There are a few very important observations to be made about the above examples before we go on to talk about perturbation theory. First consider the free rigid body problem. Note that the matrix in Eq. (2.121) is singular with rank two for $\mathbf{m} \neq 0$, and rank zero for $\mathbf{m} = 0$. Indeed, any odd dimensional phase space *must* have a singular Poisson structure, because antisymmetric matrices always have even rank. For these systems, Darboux's theorem does not apply and it is not possible to find a transformation to canonical coordinates; of course, this should

have been obvious because canonical coordinates always come in pairs and you can't pair an odd number of things.

When a system has a singular Poisson structure, the Poisson tensor will have at least one null eigenvector. Let's say it has n of them; note that n is equal to the dimensionality of the phase space, N, minus the rank of the Poisson tensor, r. In this case, it has been shown by Littlejohn [31] that it is *always* possible to find a set of $n = N - r$ scalar phase functions whose gradients are those null eigenvectors. This is not at all obvious and requires an application of the Frobenius theorem of differential geometry, where use is made of the fact that the Poisson tensor satisfies the Jacobi identity.

These n scalar phase functions are very special in that their bracket with *any* other scalar phase function must vanish. This is obvious from Eq. (2.102). Scalar phase functions with this property are called *Casimir functions*. In particular, their bracket with *any* Hamiltonian is zero, so they are always conserved quantities; note that their conservation follows directly from the bracket structure, independent of the particular Hamiltonian under consideration.

For the free rigid body problem presented above, the null eigenvector of the Poisson tensor is any multiple of $\mathbf{m}$ itself. The function:

$$C(\mathbf{m}) = m_1^2 + m_2^2 + m_3^2 \tag{2.153}$$

is then a Casimir function since its gradient is in the direction of $\mathbf{m}$, and we recognize it as the total angular momentum squared. Of course, any other scalar phase function that is functionally dependent upon C could have been used equally well. The pathology at the point $\mathbf{m} = 0$ where the rank of $\mathbf{J}$ changes is called a *symplectic bone*, and is discussed at length by Weinstein [32].

If we were to choose C to be one of our generalized coordinates, say the third coordinate in place of m_3, then it is clear that the third row and column of $\mathbf{J}$ would be zero. The two by two submatrix consisting of rows and columns one

and two would be nonsingular, and Darboux's theorem could be applied to that subsystem. Thus, the correct generalization of Darboux's theorem for singular Poisson structures is to say that it is always possible to find a transformation to a coordinate system for which the matrix of components of the Poisson tensor has an r by r submatrix in canonical form with the rest of the entries vanishing. For the free rigid body problem Poisson structure given above, this has the form:

$$\mathbf{J} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.154}$$

for $\mathbf{m} \neq 0$, and $\mathbf{J} = 0$ for $\mathbf{m} = 0$.

It is worth repeating that Casimir functions are conserved for *any* Hamiltonian. For example, the Hamiltonian:

$$H(\mathbf{m}) = \mu m_3, \tag{2.155}$$

where μ is a constant, together with the same bracket used above for the free rigid body problem, yields the equations of motion for a classical spin gyrating in a uniform magnetic field. That is, m_1 and m_2 undergo simple harmonic oscillations, while m_3 is conserved because it commutes with the Hamiltonian. Note that C is a conserved quantity for this system as well, because the bracket is the same. In general, the Poisson structure is considered to be a more fundamental entity than the Hamiltonian.

The other two examples presented in the last subsection also have singular Poisson structures. It is readily verified that the bracket for Euler's fluid equations has the Casimir functional

$$C = \int d^3\xi \mathbf{v}(\xi, t) \cdot \left[\vec{\nabla} \times \mathbf{v}(\xi, t)\right] \tag{2.156}$$

(the integrand here is called the *helicity*), and that the bracket for the Poisson-

Vlasov equations has the Casimir functionals

$$C_\Phi = \int dZ \Phi(\mathrm{f}(Z)) \tag{2.157}$$

where Φ is an arbitrary function of its argument.

2.3.5 Phase-Space Lagrangian Techniques

In this section, we review the phase space Lagrangian formalism; for more details on this subject see Littlejohn [33] and Littlejohn and Cary [34]. For a system with canonical coordinates, q and p, and time-independent Hamiltonian, $H(q,p)$, the phase space Lagrangian is given by

$$L(q,p,\dot{q},\dot{p}) = p \cdot \dot{q} - H(q,p), \tag{2.158}$$

where a dot denotes differentiation with respect to time, t. Note that L may depend upon *all* the phase space coordinates and their time derivatives, unlike ordinary configuration space Lagrangians, $L(q,\dot{q})$. The associated action is

$$A = \int dt L(q,p,\dot{q},\dot{p}), \tag{2.159}$$

the variation of which yields the Euler-Lagrange equations

$$0 = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = \dot{p} - \left(-\frac{\partial H}{\partial q}\right) = \dot{p} + \frac{\partial H}{\partial q}, \tag{2.160}$$

and

$$0 = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{p}}\right) - \frac{\partial L}{\partial p} = 0 - \left(\dot{q} - \frac{\partial H}{\partial p}\right) = -\dot{q} + \frac{\partial H}{\partial p}; \tag{2.161}$$

these are recognized as the canonical equations of motion.

We denote by z^μ, where $\mu = 1,\ldots,N$, (where $N = 2I$) the coordinates of phase space. The phase space Lagrangian may then be written

$$L(z,\dot{z}) = \gamma_\mu \dot{z}^\mu - H(z), \tag{2.162}$$

where the covector whose components are γ_μ will be called the *action one-form*. For the canonical coordinate system used above, these components are

$$\gamma_\mu = \begin{cases} p_\mu & \text{if } \mu = 1, \ldots, I \\ 0 & \text{if } \mu = I+1, \ldots, N. \end{cases} \tag{2.163}$$

The fact that I of these components are zero is a manifestation of the fact that the coordinate system is canonical. For more general coordinate systems this will not be true, as we shall see shortly. Note that phase space Lagrangians are always linear in $\dot{z}$. Also note that knowledge of the action one-form and the Hamiltonian is completely equivalent to knowledge of the phase space Lagrangian by Eq. (2.162).

The equations of motion may be written in this notation as follows:

$$\begin{aligned} 0 &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{z}^\mu}\right) - \frac{\partial L}{\partial z^\mu} \\ &= \frac{d\gamma_\mu}{dt} - \gamma_{\nu,\mu}\dot{z}^\nu + \frac{\partial H}{\partial z^\mu} \\ &= \left(\gamma_{\mu,\nu} - \gamma_{\nu,\mu}\right)\dot{z}^\nu + \frac{\partial H}{\partial z^\mu} \end{aligned} \tag{2.164}$$

or

$$\omega_{\mu\nu}\dot{z}^\nu = \frac{\partial H}{\partial z^\mu}, \tag{2.165}$$

where we have defined the Lagrangian two-form

$$\omega_{\mu\nu} \equiv \gamma_{\nu,\mu} - \gamma_{\mu,\nu}, \tag{2.166}$$

or

$$\omega \equiv d\gamma. \tag{2.167}$$

For $z = (q, p)$, where q and p are canonically conjugate, it is easily verified that Eq. (2.165) is equivalent to Eqs. (2.160) and (2.161).

We can recover the more familiar Hamiltonian formalism in the following manner: Assuming that $[\omega_{\mu\nu}]$ is a nonsingular matrix, we denote its inverse by

$J^{\mu\nu}$, so

$$J^{\mu\rho}\omega_{\rho\nu} = \delta^{\mu}_{\nu}. \tag{2.168}$$

Then Eq. (2.165) becomes

$$\dot{z}^{\mu} = J^{\mu\nu}\frac{\partial H}{\partial z^{\nu}}. \tag{2.169}$$

These are recognized as Hamilton's equations if we identify $J^{\mu\nu}$ as the Poisson tensor. That the Poisson tensor is antisymmetric and obeys the Jacobi identity is easily verified. In particular, the Jacobi identity follows directly from $d\omega = dd\gamma = 0$.

Under a (possibly noncanonical) transformation of phase space coordinates, $z \mapsto Z$, the action one-form transforms in the usual fashion of a covariant vector to give

$$\Gamma_{\mu} = \frac{\partial z^{\xi}}{\partial Z^{\mu}}\gamma_{\xi}. \tag{2.170}$$

Similarly, the Lagrangian two-form transforms like a second rank covariant tensor

$$\Omega_{\mu\nu} = \frac{\partial z^{\xi}}{\partial Z^{\mu}}\frac{\partial z^{\eta}}{\partial Z^{\nu}}\omega_{\xi\eta} = \Gamma_{\nu,\mu} - \Gamma_{\mu,\nu}, \tag{2.171}$$

where the commas in Eq. (2.171) denote partial differentiation with respect to Z. The Hamiltonian, of course, transforms as a scalar, $K(Z) = H(z)$. The new equation of motion is then

$$\Omega_{\mu\nu}\dot{Z}^{\nu} = \frac{\partial K}{\partial Z^{\mu}}, \tag{2.172}$$

which may be compared to Eq. (2.165).

Note that all of the above considerations assume a time-independent Hamiltonian. This restriction is not important for two reasons: First, we could always work in extended phase space to treat a time-dependent system; this is the approach taken by Littlejohn and Cary [34]. Second, all of our relativistic equations of motion will have the single-particle proper time as the independent variable, and nothing depends explicitly on this.

The transformation

$$\gamma_\xi \mapsto \gamma_\xi + \frac{\partial S}{\partial z^\xi}, \tag{2.173}$$

where S is an arbitrary scalar field on extended phase space, is called a *Lagrangian gauge transformation.* Though it alters the action one-form, it is easily seen to have no effect on the Lagrangian two-form, and so it does not change the equation of motion, Eq. (2.165).

It is clear that if L is independent of one of the extended phase space coordinates, say z^μ, then the associated momentum, $\partial L/\partial \dot{z}^\mu$, is conserved by Noether's theorem. Note, however, that a gauge transformation, like Eq. (2.173), using a scalar field, S, that depends upon the ignorable coordinate, could destroy the Noether symmetry, even though the associated momentum would still be conserved. The same is true for coordinate transformations like Eq. (2.170). Conversely, we see that it may be necessary to perform gauge or coordinate transformations in order to uncover Noether symmetries and, hence, to discover conserved quantities.

The strategy for our treatment of the guiding-center problem will be to start with the phase space Lagrangian for a single relativistic charged particle in an electromagnetic field, and, via a sequence of gauge and coordinate transformations, find a representation in which the gyroangle, θ, is ignorable. This is the Noether symmetry for the gyromomentum. When this is achieved, the gyroangle will no longer appear in the equations of motion for the other variables, and the magnetic moment will appear only as a constant parameter like the rest mass. Thus, in this system of "gyrocoordinates," the rapid oscillatory motion is effectively decoupled from the slower guiding-center motion, and the dimensionality of our phase space is reduced by two.

2.3.6 Constrained Systems

Eqs. (2.162) and (2.159) may be interpreted as follows: The variation of the action one form must vanish, subject to the constraint that the Hamiltonian is constant. By including other constraints, besides the fact that the Hamiltonian is constant, we can discover new and interesting Poisson structures that have those other constraints "built in."

For example, consider a particle that is constrained to move on the surface of a sphere of radius r. To model this system, we take the canonical action one form,

$$\gamma = \mathbf{p} \cdot d\mathbf{r} = p_x dx + p_y dy + p_z dz, \tag{2.174}$$

and vary it subject to the constraints that the Hamiltonian, H, be constant, that the particle position be on the sphere

$$|\mathbf{r}|^2 = x^2 + y^2 + z^2 = r^2, \tag{2.175}$$

and that the particle momentum be tangent to the sphere

$$\mathbf{r} \cdot \mathbf{p} = x p_x + y p_y + z p_z = 0. \tag{2.176}$$

The constrained variation may be done in any one of a number of ways; e.g. by use of Lagrange multipliers. Thus we write

$$L = \mathbf{p} \cdot \dot{\mathbf{r}} - \frac{1}{2}\lambda_1 |\mathbf{r}|^2 - \lambda_2 \mathbf{r} \cdot \mathbf{p} - H, \tag{2.177}$$

and form the Euler-Lagrange equations

$$\dot{\mathbf{p}} = -\lambda_1 \mathbf{r} - \lambda_2 \mathbf{p} - \frac{\partial H}{\partial \mathbf{r}} \tag{2.178}$$

$$0 = \dot{\mathbf{r}} - \lambda_2 \mathbf{r} - \frac{\partial H}{\partial \mathbf{p}}. \tag{2.179}$$

Dot the first of these equations with $\mathbf{r}$ to get

$$\mathbf{r} \cdot \dot{\mathbf{p}} = -\lambda_1 |\mathbf{r}|^2 - \mathbf{r} \cdot \frac{\partial H}{\partial \mathbf{r}}, \tag{2.180}$$

from which it follows that

$$\lambda_1 = -\frac{1}{r^2}\mathbf{r}\cdot\left(\dot{\mathbf{p}} + \frac{\partial H}{\partial \mathbf{r}}\right). \tag{2.181}$$

Then dot the second with $\mathbf{r}$ to get

$$0 = \mathbf{r}\cdot\dot{\mathbf{r}} - \lambda_2|\mathbf{r}|^2 - \mathbf{r}\cdot\frac{\partial H}{\partial \mathbf{p}}, \tag{2.182}$$

from which it follows that

$$\lambda_2 = \frac{1}{r^2}\mathbf{r}\cdot\left(\dot{\mathbf{r}} - \frac{\partial H}{\partial \mathbf{p}}\right). \tag{2.183}$$

Note that Eqs. (2.181) and (2.183) may be written in the form

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \frac{1}{r^2}\begin{pmatrix} -\mathbf{r}\cdot(\{\mathbf{p},\mathbf{r}\}+\mathbf{1}) & -\mathbf{r}\cdot\{\mathbf{p},\mathbf{p}\} \\ \mathbf{r}\cdot\{\mathbf{r},\mathbf{r}\} & -\mathbf{r}\cdot(\{\mathbf{r},\mathbf{p}\}-\mathbf{1}) \end{pmatrix}\cdot\begin{pmatrix} \partial H/\partial \mathbf{r} \\ \partial H/\partial \mathbf{p} \end{pmatrix}. \tag{2.184}$$

To get the Poisson brackets, first substitute the Lagrange multipliers, (2.181) and (2.183), back into the equations of motion, (2.178) and (2.179). We get

$$\left(\mathbf{1} - \frac{\mathbf{rr}}{r^2}\right)\cdot\dot{\mathbf{r}} = \left(\mathbf{1} - \frac{\mathbf{rr}}{r^2}\right)\cdot\frac{\partial H}{\partial \mathbf{p}} \tag{2.185}$$

and

$$\left(\mathbf{1} - \frac{\mathbf{rr}}{r^2}\right)\cdot\dot{\mathbf{p}} = -\left(\mathbf{1} - \frac{\mathbf{rr}}{r^2}\right)\cdot\frac{\partial H}{\partial \mathbf{r}} - \frac{1}{r^2}\mathbf{pr}\cdot\left(\dot{\mathbf{r}} - \frac{\partial H}{\partial \mathbf{p}}\right). \tag{2.186}$$

Note that these two equations do not determine the motion completely; they give only the projection of the motion on the sphere. To fully determine $\dot{\mathbf{r}}$ and $\dot{\mathbf{p}}$, we need to employ the derivatives of the constraints,

$$\mathbf{r}\cdot\dot{\mathbf{r}} = 0 \tag{2.187}$$

and

$$\mathbf{r}\cdot\dot{\mathbf{p}} + \dot{\mathbf{r}}\cdot\mathbf{p} = 0. \tag{2.188}$$

Using these, we finally get

$$\dot{\mathbf{r}} = \left(\mathbf{1} - \frac{\mathbf{rr}}{r^2}\right)\cdot\frac{\partial H}{\partial \mathbf{p}} \tag{2.189}$$

and

$$\dot{\mathbf{p}} = -\left(\mathbf{1} - \frac{\mathbf{rr}}{r^2}\right) \cdot \frac{\partial H}{\partial \mathbf{r}} + \frac{1}{r^2}(\mathbf{pr} - \mathbf{rp}) \cdot \frac{\partial H}{\partial \mathbf{p}}. \tag{2.190}$$

These equations of motion are Hamiltonian with the quadratic Poisson structure

$$\begin{aligned} \{r^i, r^j\} &= 0 \\ \{r^i, p_j\} &= \delta^i_j - \frac{r^i r_j}{r^2} \\ \{p_i, p_j\} &= \frac{r_j p_i - r_i p_j}{r^2}. \end{aligned} \tag{2.191}$$

Note that the constraints, Eqs. (2.175) and (2.176), are Casimir functions of this Poisson structure. This means that the Hamiltonian equations of motion will yield dynamics that respect these constraints for *any* Hamiltonian whatsoever.

There is another approach to deriving the above set of brackets. We could have adopted the spherical coordinates,

$$r = \sqrt{x^2 + y^2 + z^2} \tag{2.192}$$

$$\theta = \arctan(\sqrt{x^2 + y^2}/z) \tag{2.193}$$

$$\phi = \arctan(y/x), \tag{2.194}$$

on $\Re^3$. These have the canonically conjugate momenta

$$p_r = (xp_x + yp_y + zp_z)/\sqrt{x^2 + y^2 + z^2} \tag{2.195}$$

$$p_\theta = z(xp_x + yp_y)/\sqrt{x^2 + y^2} \tag{2.196}$$

$$p_\phi = xp_y - yp_x, \tag{2.197}$$

as is easily verified. The advantage to using these spherical coordinates is that the constraint surface in phase space is simply described by setting p_r equal to zero, and r equal to a constant.

Now we can write

$$\begin{aligned}\mathbf{r} &= x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \\ &= \cos\theta\hat{\mathbf{z}} + \sin\theta\cos\phi\hat{\mathbf{x}} + \sin\theta\sin\phi\hat{\mathbf{y}} \end{aligned} \tag{2.198}$$

and

$$\begin{aligned}\mathbf{p} &= p_x\hat{\mathbf{x}} + p_y\hat{\mathbf{y}} + p_z\hat{\mathbf{z}} \\ &= (p_r\cos\theta - \frac{p_\theta}{r}\sin\theta)\hat{\mathbf{z}} \\ &\quad + (p_r\sin\theta\cos\phi + \frac{p_\theta}{r}\cos\theta\cos\phi - \frac{p_\phi}{r}\csc\theta\sin\phi)\hat{\mathbf{x}} \\ &\quad + (p_r\sin\theta\sin\phi + \frac{p_\theta}{r}\cos\theta\sin\phi + \frac{p_\phi}{r}\csc\theta\cos\phi)\hat{\mathbf{y}}. \end{aligned} \tag{2.199}$$

Eqs. (2.198) and (2.199) and the Leibniz rule allow us to compute the brackets for the system of coordinates $(\mathbf{r}, \mathbf{p})$ in terms of the brackets for the system of coordinates $(r, \theta, \phi, p_r, p_\theta, p_\phi)$. If we ignore the constraint, then the latter system is canonical, and it follows that the former system is also canonical. If, on the other hand, we incorporate the constraint by *dictating* that r and p_r are Casimir functions and that $p_r = 0$, then the brackets (2.191) follow immediately.

It is interesting to contrast these two methods for obtaining the brackets (2.191). We shall use these methods when we cast our guiding-center equations of motion in gyrogauge and boostgauge invariant format, towards the end of the Chapter 3. Our guiding-center Poisson brackets will also have a quadratic Poisson structure, similar to that of the above set of brackets. Such quadratic Poisson structures seem to arise naturally from this type of manipulation. The reader who is interested in pursuing this topic further is encouraged to read about Dirac's theory of constraints [35].

2.4 Lie Transform Perturbation Theory

2.4.1 General Discussion of Lie Transforms

Recall that we first introduced coordinates on manifolds using the concepts of charts and atlases. A chart is a one-to-one map from a region of $\Re^n$ to a region of an n-dimensional manifold. Each coordinate, z^α, may thus be thought of as a function on the manifold. When we change coordinates, we are effectively transforming these functions.

Consider an *infinitesimal* transformation of coordinates given by

$$Z^\alpha = z^\alpha + h(\mathcal{L}_g z)^\alpha = z^\alpha + hg^\alpha(z), \tag{2.200}$$

where h is an infinitesimal, g is a vector field, and the Lie derivative acts on the coordinates as though they were scalar functions. From our geometrical interpretation of the Lie derivative, we see that we are effectively taking the functions that define the coordinates, and sliding them an infinitesimal parameter interval, h, along the field lines of g. The inverse transformation is

$$z^\alpha = Z^\alpha - h(\mathcal{L}_g Z)^\alpha = Z^\alpha - hg^\alpha(Z). \tag{2.201}$$

Of course, since h is an infinitesimal, we are scrupulously ignoring anything of order h^2.

Now we ask how basis vector components behave under the above transformation. Assume a coordinate basis for simplicity. We have

$$\frac{\partial}{\partial Z^\alpha} = \frac{\partial z^\beta}{\partial Z^\alpha}\frac{\partial}{\partial z^\beta} = \frac{\partial}{\partial z^\alpha} - h\frac{\partial g^\beta}{\partial z^\alpha}\frac{\partial}{\partial z^\beta}. \tag{2.202}$$

Similarly, basis covector components transform as follows:

$$dZ^\alpha = \frac{\partial Z^\alpha}{\partial z^\beta}dz^\beta = dz^\alpha + h\frac{\partial g^\alpha}{\partial z^\beta}dz^\beta. \tag{2.203}$$

Now suppose that $\mathbf{t}$ is some tensor field on the manifold. We can ask how the components of $\mathbf{t}$ behave under the above transformation. Use a prime to

distinguish the components of $\mathbf{t}$ in the new coordinate system. We demand

$$t'^{\alpha_1\cdots\alpha_r}_{\beta_1\cdots\beta_s}(Z)\frac{\partial}{\partial Z^{\alpha_1}}\otimes\cdots\otimes\frac{\partial}{\partial Z^{\alpha_r}}\otimes dZ^{\beta_1}\otimes\cdots\otimes dZ^{\beta_s}$$
$$= t^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s}(z)\frac{\partial}{\partial z^{\mu_1}}\otimes\cdots\otimes\frac{\partial}{\partial z^{\mu_r}}\otimes dz^{\nu_1}\otimes\cdots\otimes dz^{\nu_s}. \quad (2.204)$$

Now expand in h, retaining only first order terms. We find

$$t'^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s} = t^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s} - h\big(t^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s,\alpha}g^{\alpha}$$
$$- t^{\alpha\mu_2\cdots\mu_r}_{\nu_1\cdots\nu_s}g^{\mu_1}{}_{,\alpha} - \cdots - t^{\mu_1\cdots\mu_{r-1}\alpha}_{\nu_1\cdots\nu_s}g^{\mu_r}{}_{,\alpha}$$
$$+ t^{\mu_1\cdots\mu_r}_{\alpha\nu_2\cdots\nu_s}g^{\alpha}{}_{,\nu_1} + \cdots + t^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_{s-1}\alpha}g^{\alpha}{}_{,\nu_s}\big). \quad (2.205)$$

Suppose that we define a new tensor field, $\mathbf{T}$, whose components in the old system are the same as those of $\mathbf{t}$ in the new system. Then, by comparison with Eq. (2.43), we may write

$$\mathbf{T} = \mathbf{t} - h(\mathcal{L}_g\mathbf{t}), \quad (2.206)$$

where comparison with Eq. (2.43) is helpful. Furthermore, since this last equation is in coordinate-free form, it is true for coordinate bases and noncoordinate bases alike.

Compare the signs of the second terms on the right-hand sides of Eqs. (2.206) and (2.200). Despite the algebra that went into proving the above result, it has a marvelously simple geometric interpretation. If we slide the values of the coordinates one way along a field line of g, then we must slide the tensor field in the other direction. In case this is not obvious, a trivial example is afforded by a scalar field on $\Re$, call it $f(x)$. If we transform coordinates to $X = x + h$, then $F(X) = f'(X) = f(x) = f(X - h) = f(X) - h(df/dX)(X) = (f - h\mathcal{L}f)(X)$.

Suppose that our tensor field is the tensor product of two tensor fields, say $\mathbf{t} = \mathbf{t}_1 \otimes \mathbf{t}_2$. Then, since Lie derivatives obey the Leibniz rule over the tensor product, we have

$$\mathbf{T} = \mathbf{t} - h(\mathcal{L}_g\mathbf{t})$$

$$
\begin{aligned}
&= \mathbf{t}_1 \otimes \mathbf{t}_2 - h\mathcal{L}_g(\mathbf{t}_1 \otimes \mathbf{t}_2) \\
&= \mathbf{t}_1 \otimes \mathbf{t}_2 - h(\mathcal{L}_g \mathbf{t}_1) \otimes \mathbf{t}_2 - h\mathbf{t}_1 \otimes (\mathcal{L}_g \mathbf{t}_2) \\
&= [\mathbf{t}_1 - h(\mathcal{L}_g \mathbf{t}_1)] \otimes [\mathbf{t}_2 - h(\mathcal{L}_g \mathbf{t}_2)], \qquad (2.207)
\end{aligned}
$$

where, as always, we neglect $\mathcal{O}(h^2)$. This result indicates that the infinitesimal transformation commutes with the tensor product.

Next suppose that the tensor field is obtained by starting with a tensor of higher rank and applying to it some number of vectors and/or covectors. For example, say $\mathbf{t} = \mathbf{s}(\mathbf{a}, \mathbf{U})$ where $\mathbf{a}$ is a covector field and $\mathbf{U}$ is a vector field; we could have let $\mathbf{s}$ have more than one of each type of argument or other unfilled slots without affecting the following reasoning in any way. Apply the transformation, and use Eq. (2.44) to write

$$
\begin{aligned}
\mathbf{T} &= \mathbf{t} - h(\mathcal{L}_g \mathbf{t}) \\
&= \mathbf{s}(\mathbf{a}, \mathbf{U}) - h\mathcal{L}_g[\mathbf{s}(\mathbf{a}, \mathbf{U})] \\
&= \mathbf{s}(\mathbf{a}, \mathbf{U}) - h(\mathcal{L}_g \mathbf{s})(\mathbf{a}, \mathbf{U}) - h\mathbf{s}(\mathcal{L}_g \mathbf{a}, \mathbf{U}) - h\mathbf{s}(\mathbf{a}, \mathcal{L}_g \mathbf{U}) \\
&= (\mathbf{s} - h\mathcal{L}_g \mathbf{s})(\mathbf{a} - h\mathcal{L}_g \mathbf{a}, \mathbf{U} - h\mathcal{L}_g \mathbf{U}). \qquad (2.208)
\end{aligned}
$$

This result indicates that the transformation commutes with the application of the vectors and/or covectors.

Next suppose that the tensor field is an exact form. That is, say $\mathbf{t} = d\mathbf{\Omega}$. Since Lie derivatives commute with exterior derivatives, it follows that the transformation commutes with the application of the exterior derivative.

The above results indicate that *any* tensorial relationship, including those with differential operators, retains its form under a transformation of the form given in Eq. (2.200). This crucial point makes the Lie transform method possible.

Now suppose that we wish to consider finite (rather than infinitesimal) changes of coordinates. That is, suppose we wish to slide the coordinate values a finite parameter interval, ϵ, along the field lines of g. The easiest approach

is to divide the finite interval into a large number of infinitesimal intervals by writing

$$Z = \lim_{N\to\infty}(1+\frac{\epsilon}{N}\mathcal{L}_g)^N z = \exp(\epsilon\mathcal{L}_g)z. \tag{2.209}$$

The finite transformation of the tensor, $\mathbf{t}$, is then

$$\mathbf{T} = \lim_{N\to\infty}(1-\frac{\epsilon}{N}\mathcal{L}_g)^N \mathbf{t} = \exp(-\epsilon\mathcal{L}_g)\mathbf{t}. \tag{2.210}$$

The transformation given by the above equations is called a *Lie transform* generated by the vector field, g.

Because the infinitesimal transformations of the form given in Eq. (2.200) are known to preserve tensorial relationships, and because a Lie transform is composed of nothing more than a large number of these infinitesimal transformations, it follows that Lie transforms preserve tensorial relationships. That is

$$\exp(-\epsilon\mathcal{L}_g)(\mathbf{t}_1\otimes\mathbf{t}_2) = (\exp(-\epsilon\mathcal{L}_g)\mathbf{t}_1)\otimes(\exp(-\epsilon\mathcal{L}_g)\mathbf{t}_2), \tag{2.211}$$

and

$$\exp(-\epsilon\mathcal{L}_g)\big[\mathbf{s}(\mathbf{a},\mathbf{U})\big] = \big[\exp(-\epsilon\mathcal{L}_g)\mathbf{s}\big]\big(\exp(-\epsilon\mathcal{L}_g)\mathbf{a},\exp(-\epsilon\mathcal{L}_g)\mathbf{U}\big), \tag{2.212}$$

and

$$\exp(-\epsilon\mathcal{L}_g)(d\boldsymbol{\Omega}) = d\big(\exp(-\epsilon\mathcal{L}_g)\boldsymbol{\Omega}\big). \tag{2.213}$$

We now have a way of making finite coordinate transformations of any tensorial equation that is guaranteed to preserve its tensorial form.

By Taylor expanding the exponential in Eq. (2.209) and using Eq. (2.30) for the Lie derivative, it is possible to develop the transformation to arbitrarily high order in ϵ. In practice, we want to be able to control the transformation order by order in ϵ. There are two ways to do this. The first, due to Deprit [36], is to order the generator, g, in ϵ. The second, due to Dragt and Finn [37], is to make a succession of transformations like Eq. (2.209), as follows:

$$Z = \exp(\epsilon\mathcal{L}_{g_1})\exp(\epsilon^2\mathcal{L}_{g_2})\exp(\epsilon^3\mathcal{L}_{g_3})\cdots z. \tag{2.214}$$

In this work, we adopt the second procedure, as it was shown by Cary [38] to involve fewer terms in the perturbation series at each order. Expanding the above equation in ϵ and using Eq. (2.30), we get

$$Z = z + \epsilon\mathcal{L}_1 z + \epsilon^2(\mathcal{L}_2 + \frac{1}{2}\mathcal{L}_1^2)z + \epsilon^3(\mathcal{L}_3 + \mathcal{L}_1\mathcal{L}_2 + \frac{1}{6}\mathcal{L}_1^3)z + \cdots, \tag{2.215}$$

Here we have used $\mathcal{L}_n$ to abbreviate $\mathcal{L}_{g_n}$. The inverse transformation is then

$$z = \cdots \exp(-\epsilon^3\mathcal{L}_3)\exp(-\epsilon^2\mathcal{L}_2)\exp(-\epsilon\mathcal{L}_1)Z. \tag{2.216}$$

Developing this order by order, we get

$$z = Z - \epsilon\mathcal{L}_1 Z - \epsilon^2(\mathcal{L}_2 - \frac{1}{2}\mathcal{L}_1^2)Z - \epsilon^3(\mathcal{L}_3 - \mathcal{L}_2\mathcal{L}_1 + \frac{1}{6}\mathcal{L}_1^3)Z - \cdots. \tag{2.217}$$

The transformation of the tensor $\mathbf{t}$ is then

$$\mathbf{T} = \cdots \exp(-\epsilon^3\mathcal{L}_3)\exp(-\epsilon^2\mathcal{L}_2)\exp(-\epsilon\mathcal{L}_1)\mathbf{t}. \tag{2.218}$$

Let us suppose that $\mathbf{t}$ is given as a power series in the expansion parameter, ϵ, so

$$\mathbf{t} = \mathbf{t}_0 + \epsilon\mathbf{t}_1 + \epsilon^2\mathbf{t}_2 + \epsilon^3\mathbf{t}_3 + \cdots. \tag{2.219}$$

Then Eq. (2.218) yields

$$\mathbf{T} = \mathbf{T}_0 + \epsilon\mathbf{T}_1 + \epsilon^2\mathbf{T}_2 + \epsilon^3\mathbf{T}_3 + \cdots, \tag{2.220}$$

where

$$\mathbf{T}_0 = \mathbf{t}_0, \tag{2.221}$$

$$\mathbf{T}_1 = \mathbf{t}_1 - \mathcal{L}_1\mathbf{t}_0, \tag{2.222}$$

$$\mathbf{T}_2 = \mathbf{t}_2 - \mathcal{L}_2\mathbf{t}_0 - \mathcal{L}_1\mathbf{t}_1 + \frac{1}{2}\mathcal{L}_1^2\mathbf{t}_0, \tag{2.223}$$

$$\mathbf{T}_3 = \mathbf{t}_3 - \mathcal{L}_3\mathbf{t}_0 - \mathcal{L}_2\mathbf{t}_1 + \mathcal{L}_2\mathcal{L}_1\mathbf{t}_0 - \mathcal{L}_1\mathbf{t}_2 + \frac{1}{2}\mathcal{L}_1^2\mathbf{t}_1 - \frac{1}{6}\mathcal{L}_1^3\mathbf{t}_0, \tag{2.224}$$

etc.

Given any equation written in tensor form, we can now make near-identity coordinate transformations to perform perturbation analyses. That is, if the equation has the form of a solvable equation plus a small perturbation, we can make a Lie transform to coordinates for which the perturbation is removed or at least simplified. The form of the generator, g, required to achieve this simplification depends on the specific problem, and is chosen order by order in the perturbation series.

Once this process has been carried out to first order, we could continue on to second and higher order, or we could regard the first-order problem as a new solvable problem and renormalize the perturbation series accordingly before proceeding to higher order. The latter strategy is called the *superconvergent* Lie transform procedure; superconvergent perturbation series were first investigated by Kolmogorov [39]. All this will be made clear by selected examples in the next few subsections.

2.4.2 Lie Transforming a Scalar Field

Consider the scalar equation

$$f(x) = \epsilon x^2 + 2x - 2c = 0, \tag{2.225}$$

where c is a constant and ϵ is our expansion parameter. Let's pretend for a moment that we do not know how to solve a quadratic equation. The scalar field, f, is ordered in ϵ as follows:

$$f_0(x) = 2x - 2c, \tag{2.226}$$

$$f_1(x) = x^2, \tag{2.227}$$

and $f_n(x) = 0$ for $n \geq 2$.

We wish to perform a Lie transform to a new coordinate, X, for which the transformed scalar will be denoted by F. Since we are working in $\mathfrak{R}$, the generator,

g, has only one component. At order zero, use Eq. (2.221),

$$F_0 = f_0. \tag{2.228}$$

At order one, use Eq. (2.222),

$$F_1 = f_1 - g_1 f_0' = x^2 - 2g_1. \tag{2.229}$$

Thus, we see that we can make F_1 vanish by choosing $g_1 = x^2/2$. Moving on to second order, we use Eq. (2.223),

$$F_2 = -2g_2 - \frac{x^3}{2}. \tag{2.230}$$

So we can make F_2 vanish by choosing $g_2 = -x^3/4$. Thus, to order ϵ^2, we have the Lie transformed scalar equation

$$F(X) = 2X - 2c = 0. \tag{2.231}$$

This has solution, $X = c$. Now x is given in terms of X by Eq. (2.217) which becomes

$$\begin{aligned} x &= X - \epsilon g_1 - \epsilon^2 (g_2 - \frac{1}{2} g_1 g_1') - \cdots \\ &= X - \frac{\epsilon}{2} X^2 + \frac{\epsilon^2}{2} X^3 - \cdots \\ &= c - \frac{\epsilon}{2} c^2 + \frac{\epsilon^2}{2} c^3 - \cdots . \end{aligned} \tag{2.232}$$

This matches the Taylor expansion of the exact solution to the quadratic equation

$$x = \frac{1}{\epsilon} \left(-1 + \sqrt{1 + 2\epsilon c} \right), \tag{2.233}$$

to $\mathcal{O}(\epsilon^2)$, as is easily verified.

Note that there is another solution to the quadratic equation

$$x = \frac{1}{\epsilon} \left(-1 - \sqrt{1 + 2\epsilon c} \right), \tag{2.234}$$

of leading order ϵ^{-1} that our technique does not give. This is because it is not continuously connected to the solution of the unperturbed problem as ϵ goes to zero. Lie transforms are useful only for near-identity coordinate transformations.

2.4.3 Lie Transforming a Vector Field

Now consider the following dynamical system:

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= -x - \epsilon x^2. \end{aligned} \tag{2.235}$$

If we use $\mathbf{z}$ to denote (x, y), then this may be written

$$\dot{\mathbf{z}} = \mathbf{v}_0 + \epsilon \mathbf{v}_1, \tag{2.236}$$

where we have defined the vectors, $\mathbf{v}_0 \equiv (y, -x)$ and $\mathbf{v}_1 \equiv (0, -x^2)$. We now try to Lie transform to new coordinates, $Z = (X, Y)$, in an attempt to get rid of the order ϵ term. The transformed vector field is $\mathbf{V} = \mathbf{V}_0 + \epsilon \mathbf{V}_1$, where $\mathbf{V}_0 = \mathbf{v}_0$, and $\mathbf{V}_1$ is given from Eq. (2.222),

$$\mathbf{V}_1 = \mathbf{v}_1 - \mathcal{L}_g \mathbf{v}_0. \tag{2.237}$$

Using the formula for the Lie derivative of a vector, the demand that $\mathbf{V}_1 = 0$ is seen to be equivalent to the following pair of equations:

$$\left(y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right) g_1^x = g_1^y$$

and

$$\left(y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right) g_1^y = -g_1^x + x^2. \tag{2.238}$$

These may be solved by the method of characteristics to yield

$$g_1^x = \frac{1}{3}(x^2 + 2y^2)$$

and

$$g_1^y = -\frac{2}{3} xy. \tag{2.239}$$

Note that the characteristic equations for this system are the unperturbed equations of motion. This "integration along unperturbed orbits" is a generic feature of problems of this sort.

Now then, the new coordinates are given in terms of the old by

$$X = x + \frac{\epsilon}{3}(x^2 + 2y^2)$$
$$Y = y - \frac{2\epsilon}{3}xy. \tag{2.240}$$

The inverse transformation is then

$$x = X - \frac{\epsilon}{3}(X^2 + 2Y^2)$$
$$y = Y + \frac{2\epsilon}{3}XY. \tag{2.241}$$

Note that we are ignoring terms of order ϵ^2 or higher. Now the equations of motion for $\mathbf{Z}$ are

$$\dot{X} = Y$$
$$\dot{Y} = -X. \tag{2.242}$$

These have solution

$$X = X_0 \cos t + Y_0 \sin t$$
$$Y = Y_0 \cos t - X_0 \sin t. \tag{2.243}$$

Thus, the solution for $z(t)$ is given by Eqs. (2.241) and (2.243). If desired, the initial conditions for $\mathbf{Z}$ can be expressed in terms of the initial conditions for $\mathbf{z}$ using Eq. (2.240).

Frequently, in physical applications of this formalism, it happens that the new coordinates have physical significance. For example, in guiding-center theory, we shall find a Lie transform that takes us from the phase space coordinates of a particle to those of a guiding center. In such a circumstance, very little is gained by expressing the initial conditions of the transformed problem in terms of those of the original problem. Instead, the new coordinates acquire their own physical significance, and we can speak of "the equations of motion of a guiding center"

and "the initial conditions of a guiding center," and forget all about the original single-particle coordinates.

For a less trivial example of the vector Lie transform technique, see Appendix C where the method is used to calculate the gyrofrequency shift for two-dimensional nonrelativistic guiding-center motion in a spatially nonuniform electromagnetic field.

2.4.4 Canonical Lie Transforms of a Hamiltonian System

When using perturbation theory to study a Hamiltonian dynamical system, the above technique of Lie transforming the dynamical vector field could be used, but there is a serious problem with this approach: There is no guarantee that the Lie transform of a Hamiltonian vector field will be another Hamiltonian vector field.

Recall that a Hamiltonian vector field is given by contracting the Poisson tensor with the gradient of a scalar function. This suggests the following solution to the above problem: Instead of Lie transforming the Hamiltonian vector field, Lie transform the Poisson tensor and Hamiltonian separately. This will insure that the transformed equations of motion are still in Hamiltonian form.

Let us examine a little more closely why this should work. Hamiltonian equations of motion are given by Eq. (2.111). If we write

$$\mathbf{Z} = \exp(\epsilon\mathcal{L})\mathbf{z},$$

$$\mathbf{J}' = \exp(-\epsilon\mathcal{L})\mathbf{J},$$

and

$$H' = \exp(-\epsilon\mathcal{L})H, \tag{2.244}$$

then since our equations of motion are in tensor form, we are guaranteed that

the new equations of motion will be

$$\dot{\mathbf{Z}} = \mathbf{J}' \cdot \frac{\partial H'}{\partial \mathbf{Z}}. \tag{2.245}$$

Furthermore, we are guaranteed that $\mathbf{J}'$ is antisymmetric and obeys the Jacobi identity because these requirements can also be written as tensorial equations (see Eqs. (2.103) and (2.104), respectively). Thus, Eq. (2.245) qualifies as a bona fide Hamiltonian system.

We can now prove a marvelous theorem that considerably simplifies the work involved in making *canonical* (bracket-preserving) Lie transformations of a Hamiltonian system, and is probably responsible for the popularity of the Lie transform technique: A Poisson tensor is a Lie-dragged tensor along *any* vector field that is Hamiltonian with respect to it. Suppose the Poisson tensor is denoted by $\mathbf{J}$. Let $\mathbf{V}$ be given by

$$V^\alpha = J^{\alpha\beta}\frac{\partial W}{\partial z^\beta} \tag{2.246}$$

for some (any) scalar field, W. Then the theorem states

$$\mathcal{L}_V \mathbf{J} = 0. \tag{2.247}$$

This is easily proved using the formula for the Lie derivative of a second rank contravariant tensor. We write

$$\begin{aligned}(\mathcal{L}_V \mathbf{J})^{\alpha\beta} &= V^\xi J^{\alpha\beta}{}_{,\xi} - V^\alpha{}_{,\xi} J^{\xi\beta} - V^\beta{}_{,\xi} J^{\alpha\xi} \\ &= -(J^{\alpha\xi} J^{\beta\gamma}{}_{,\xi} + J^{\gamma\xi} J^{\alpha\beta}{}_{,\xi} + J^{\beta\xi} J^{\gamma\alpha}{}_{,\xi}) W_{,\gamma} \\ &\quad - J^{\alpha\gamma} W_{,\xi\gamma} (J^{\xi\beta} + J^{\beta\xi}),\end{aligned} \tag{2.248}$$

where we have used Eq. (2.246) for $\mathbf{V}$. The first term vanishes by the Jacobi identity, the second term vanishes by antisymmetry, and the theorem is proved.

It immediately follows that a Lie transform along the vector field $\mathbf{V}$ leaves $\mathbf{J}$ unchanged. This is because a Lie transform is the exponentiation of a Lie

derivative (set $\mathcal{L}_m \mathbf{t}_n = 0$ in Eqs. (2.221) through (2.224) to recover $\mathbf{T} = \mathbf{t}$). Thus, Lie transforms generated by Hamiltonian vector fields are always canonical. Now Hamiltonian vector fields are in one-to-one correspondence with scalar phase functions, W, by Eq. (2.246), so we have found a way to generate canonical transformations with scalars.

Thus, to perform a canonical Lie transform of a Hamiltonian system, we need only to transform the Hamiltonian. Now the Lie derivative of a scalar with respect to a Hamiltonian vector field is given by

$$\mathcal{L}_V H = V^\alpha H_{,\alpha} = J^{\alpha\beta} W_{,\beta} H_{,\alpha} = -\{W, H\}. \tag{2.249}$$

Thus, for a canonical Lie transform of a Hamiltonian, we may rewrite Eqs. (2.221) through (2.224) as follows:

$$K_0 = H_0, \tag{2.250}$$

$$K_1 = H_1 + \{W_1 H_0\}, \tag{2.251}$$

$$K_2 = H_2 + \{W_2, H_0\} + \{W_1, H_1\} + \frac{1}{2}\{W_1, \{W_1, H_0\}\} \tag{2.252}$$

$$\begin{aligned} K_3 = H_3 &+ \{W_3, H_0\} + \{W_2, H_1\} + \{W_2, \{W_1, H_0\}\} + \{W_1, H_2\} \\ &+ \frac{1}{2}\{W_1, \{W_1, H_1\}\} + \frac{1}{6}\{W_1, \{W_1, \{W_1, H_0\}\}\}, \end{aligned} \tag{2.253}$$

etc. Here we have denoted the new Hamiltonian by K.

To see how this is used, consider the following example: We perturb a harmonic oscillator Hamiltonian by the addition of a nonlinear term,

$$H = \frac{1}{2}(q^2 + p^2) - \frac{6}{3}p^4. \tag{2.254}$$

Note that the unperturbed motion oscillates with unit frequency. We can introduce action-angle variables for the unperturbed Hamiltonian,

$$J = \frac{1}{2}(q^2 + p^2) \tag{2.255}$$

$$\theta = \arctan(q/p), \tag{2.256}$$

so that

$$H = J - \frac{\epsilon}{2}J^2\left(1 + \frac{4}{3}\cos(2\theta) + \frac{1}{3}\cos(4\theta)\right). \tag{2.257}$$

Thus we have

$$H_0 = J \tag{2.258}$$

and

$$H_1 = -\frac{1}{2}J^2\left(1 + \frac{4}{3}\cos(2\theta) + \frac{1}{3}\cos(4\theta)\right). \tag{2.259}$$

We now try to remove H_1 by a canonical Lie transform generated by the scalar, W_1 (we shall work only to order one in ϵ). We have $K_0 = H_0$, and

$$K_1 = H_1 + \{W_1, H_0\} = H_1 + \{W_1, J\} = H_1 + \frac{\partial W_1}{\partial \theta}. \tag{2.260}$$

Note that we cannot demand that $K_1 = 0$ since that would cause W_1 to be multivalued (that is, secular terms would appear in W_1). The best that we can hope for is to make K_1 equal to the θ-average of H_1. That is,

$$K_1 = -\frac{1}{2}J^2. \tag{2.261}$$

Then

$$\frac{\partial W_1}{\partial \theta} = \frac{1}{6}J^2\left[4\cos(2\theta) + \cos(4\theta)\right], \tag{2.262}$$

and this integrates to give

$$W_1 = \frac{1}{24}J^2\left[8\sin(2\theta) + \sin(4\theta)\right]. \tag{2.263}$$

Using this generator we can work out the transformation equations, and hence completely solve the problem (to order ϵ). For now we note that the perturbed frequency is given by

$$\Omega \equiv \frac{\partial K}{\partial J} = 1 - \epsilon J. \tag{2.264}$$

Note how the Lie transform has taken us to a new set of coordinates in which the perturbation is averaged; that is, independent of the angle variable. Since

the resulting Hamiltonian depends only on the action variable, it is integrable by definition. Furthermore, secular terms were avoided by this absorbing of the averaged part of the perturbation into the new Hamiltonian.

Aforementioned problems of resonant perturbations occur when the unperturbed motion has characteristic frequencies that vary with the action (this is true generically, but not in our above example). When this happens, $\partial W_1/\partial\theta$ can equal a quantity that is oscillatory but whose frequency passes through zero on some set of measure zero in phase space. Thus, in some neighborhood of this region, problems of secular behavior can develop. Various techniques exist for dealing with this problem, but we shall not consider such problematic regions of phase space in this thesis.

2.4.5 Noncanonical Lie Transforms of a Hamiltonian System

It sometimes happens that a canonical transformation is not the best way to solve a particular problem in perturbation theory. This may be because it is best to express the unperturbed problem in noncanonical coordinates for which the perturbation alters not only the Hamiltonian but also the Poisson structure. This is the case for both the guiding-center and oscillation-center problems whose solution forms the core of this thesis. In this case, we must resort to noncanonical transformations, but we demand that they preserve the Hamiltonian nature of the equations of motion. As has already been pointed out, this can be accomplished by Lie transforming the Poisson tensor along with the Hamiltonian; this means that the vector generator of the Lie transform should not be a Hamiltonian vector field.

Consider once again the harmonic oscillator Hamiltonian,

$$H = \frac{1}{2}(q^2 + p^2). \tag{2.265}$$

This time, we introduce a perturbation not in the Hamiltonian but rather in the Poisson structure. Suppose that the perturbed brackets are

$$\{q, p\} = 1 - \epsilon p^2. \tag{2.266}$$

Thus we have $\mathbf{J} = \mathbf{J}_0 + \epsilon \mathbf{J}_1$, where $\mathbf{J}_0$ is the canonical Poisson tensor. We wish to perform a Lie transform that will restore the bracket to its canonical form. We demand

$$0 = \mathbf{J}_1' = \mathbf{J}_1 - \mathcal{L}_g \mathbf{J}_0. \tag{2.267}$$

Straightforward computation shows that this imposes only one independent requirement on the generating vector field, $\mathbf{g}$, namely

$$\frac{\partial g^q}{\partial q} + \frac{\partial g^p}{\partial p} = p^2. \tag{2.268}$$

It is easy enough to solve this equation; for example, we could take

$$g^p = \frac{p^3}{3} \tag{2.269}$$

and

$$g^q = 0. \tag{2.270}$$

This effectively restores the bracket to canonical form, but it alters the Hamiltonian as follows:

$$K = H - \epsilon \mathcal{L}_g H = \frac{1}{2}(q^2 + p^2) - \frac{\epsilon}{3} p^4. \tag{2.271}$$

Note that this transformed problem is coincidentally the same one that we treated in the last subsection. Thus, we could now apply a second (this time canonical) Lie transform to finally solve it. Once again, we would find the perturbed frequency, $\Omega = 1 - \epsilon J$.

The important thing to note here is that g is not a Hamiltonian vector field. If it were, there would have to exist a scalar function W such that $0 = \partial W/\partial p$ and $p^3/3 = -\partial W/\partial q$. Examination of the mixed second derivatives shows these to be incompatible requirements.

2.4.6 Lie Transforming the Phase-Space Lagrangian

There is another way to go about making noncanonical transformations of a Hamiltonian system that is guaranteed to keep it Hamiltonian. Recall that specifying the action one form is equivalent to specifying the Poisson tensor (assuming that everything is nonsingular). We can simply take the exterior derivative of γ to get ω, and then invert ω to get $\mathbf{J}$. These are all tensorial relationships, so we could just as well Lie transform γ and H instead of $\mathbf{J}$ and H.

Indeed, there are several advantages to this approach. First, it is easier to take Lie derivatives of one forms than of second rank contravariant tensors; there is one less term to worry about, and, more importantly, we can use the homotopy formula to help us Lie differentiate one forms. Second, when we Lie transform the Poisson tensor, we are guaranteed that the resulting tensor will be a valid Poisson structure only to the order we are keeping. When we Lie transform the action one form on the other hand, its exterior derivative is still going to be closed even if we truncate it. Thus ω is exactly closed, so $\mathbf{J} = \omega^{-1}$ will obey the Jacobi identity *exactly.*

Consider a Lie transformation of the original action one form, γ, into a new action one form, Γ. Using the homotopy formula, Eqs. (2.221) through (2.224) become

$$\Gamma_0 = \gamma_0, \tag{2.272}$$

$$\Gamma_1 = \gamma_1 - i_1\omega_0 + dS_1, \tag{2.273}$$

$$\Gamma_2 = \gamma_2 - i_2\omega_0 - \frac{1}{2}i_1(\omega_1 + \Omega_1) + dS_2, \tag{2.274}$$

$$\Gamma_3 = \gamma_3 - i_3\omega_0 - i_2\Omega_1 - i_1[\omega_2 - \frac{1}{3}di_1(\omega_1 + \frac{1}{2}\Omega_1)] + dS_3, \tag{2.275}$$

etc. Here, we have defined $\omega_n \equiv d\gamma_n$, and $\Omega_n \equiv d\Gamma_n$. Note that in these equations, we have also made near-identity gauge transformations by adding dS_n at order n for all $n \geq 1$. In fact, any other one-forms in these equations that were given

by the exterior derivative of a scalar (typically arising from the second term on the right of Eqs. (2.75) and (2.76)), were absorbed in the definitions of the S_n.

Thus, these last transformation equations are capable of dealing with any near-identity coordinate or gauge transformations, and so it is these that we shall use in the sections to follow. The vectors g_n and the scalars S_n will be determined by certain desiderata: We want the transformation to average away the rapidly oscillating terms of the Hamiltonian and action one-form, and we want to avoid secular terms. For the guiding-center problem, we shall also want the action one-form to be invariant with respect to certain transformations called *gyrogauge* and *boostgauge* transformations. This will be explained in more detail later.

For now, we consider another simple example. Consider once again the harmonic oscillator Hamiltonian, and perturb the canonical action one form as follows:

$$\gamma = pdq + \frac{\epsilon}{3}p^3 dq. \tag{2.276}$$

We have

$$\omega = d\gamma = (1 + \epsilon p^2) dp \wedge dq. \tag{2.277}$$

This inverts to give $(1 + \epsilon p^2)^{-1}$ times the canonical Poisson tensor, and to order ϵ this is the same as the perturbation that was examined in the last subsection (which is why we chose it). We can now compare the two methods of doing the problem.

Demand that $\Gamma_1 = 0$, so Eq. (2.273) gives

$$0 = \Gamma_1 = \gamma_1 - i_1\omega_0 + dS_1 = (\frac{1}{3}p^3 - g^p)dq - g^q dp + dS_1. \tag{2.278}$$

Thus we can take $S_1 = 0$, and

$$g^p = \frac{p^3}{3} \tag{2.279}$$

and

$$g^q = 0. \tag{2.280}$$

These are precisely the same generators that we discovered in the last subsection, they have precisely the same effect on the Hamiltonian, and the rest of the problem follows in identical fashion. That is, a second canonical Lie transformation is necessary to get to averaged coordinates.

Chapter 3

Relativistic Guiding-Center Theory

3.1 Discussion

Relativistic guiding-center motion occurs in many applications of plasma physics, including controlled fusion, free-electron lasers, and astrophysics. The tandem mirror and bumpy torus plasma confinement devices, for example, utilize populations of magnetized electrons at relativistic energies in complicated field-line geometries. In free-electron lasers, relativistic electron beams travel along strong magnetic fields with superposed wiggler fields. Near a neutron star, relativistic plasma can be confined in strong electromagnetic and gravitational fields.

All these examples point out the need for a formalism that is able to treat general electromagnetic field geometries. Particle simulation codes used for studying the properties of guiding-center plasmas in controlled fusion confinement devices sometimes require the guiding-center equations of motion to one order higher than the usual drifts; this indicates the need for a simplified and systematic per-

turbative treatment, such as that afforded by the use of Lie transforms. The free-electron laser problem has no obvious preferred frame of reference, and this suggests that a manifestly covariant description would best reveal the essence of the physical processes involved. The neutron star problem involves coupling to a general relativistic gravitational field, and this absolutely requires a manifestly covariant formulation. All these desiderata will be satisfied by our theory.

Nonrelativistic theories of guiding-center motion in arbitrary magnetic geometry frequently make use of orthonormal triads of unit vectors at each point of three-dimensional physical space. One member of each such triad is required to lie in the direction of the magnetic field at that point. Such a basis affords great clarity and relative ease in the computation and exposition of the results of guiding-center theory.

One of the first problems to be addressed in any relativistic formulation of guiding-center theory is thus that of finding the relativistic analogs of these basis triads. Fortunately, this problem has been solved by Fradkin [13], who gives a straightforward method for finding orthonormal tetrads of unit vectors at each point of four-dimensional spacetime. In a frame for which the perpendicular electric field vanishes, one pair of unit vectors in these tetrads lies perpendicular to the magnetic field, while the other pair spans the two-dimensional subspace determined by the direction of the magnetic field and the direction of time.

Fradkin shows that these two two-dimensional subspaces are covariantly defined, and that the rapid gyration takes place in the first of these, while the slower parallel motion takes place in the second. This formalism is therefore useful for isolating the oscillatory motion so that it can be effectively averaged to obtain the guiding-center equations of motion. It is described from first principles in Sections 3.2, 3.3, and 3.4.

Lie transform perturbation theory is used to perform the averaging. Though this technique has been known for some time [40], its use for the guiding-center

problem poses special difficulties which were first overcome by Littlejohn [22]. The difficulties are due to the fact that the Poisson structure as well as the Hamiltonian depends upon the rapidly gyrating variables, so that the transformation required to gyroaverage the system of equations is not canonical.

A Lie transform in its most general sense is a coordinate transformation generated by a vector field on phase space. If this vector field generator is a Hamiltonian vector field (that is, a vector field that is the flow generated by some scalar Hamiltonian-like function) then the transformation it induces is canonical; in this case one often simply speaks of the transformation as being generated by the corresponding scalar function. For the guiding-center problem, however, the vector generator of the averaging transformation cannot be a Hamiltonian vector field, since it must generate a noncanonical transformation.

In the nonrelativistic guiding-center problem, it was found by Littlejohn [7] to be easiest to apply the general Lie transform to the action one form. This is the approach that is followed here; it was described from first principles in Chapter 2.

In any calculation that goes beyond the lowest order drifts, it was found by Littlejohn [41] to be necessary to worry about maintaining a certain gauge invariance property of the action one form which for the nonrelativistic case is known as *gyrogauge* invariance. If the averaging transformation does not preserve this invariance property, then the final guiding-center equations of motion will depend unavoidably on the arbitrarily chosen basis vectors used to set up the problem, as was noted by Hagan and Frieman [42]. In Section 3.6, we work out the relativistic generalization of this invariance property, and we find that the relativistic case admits another similar gauge invariance property which we call *boostgauge* invariance.

The Lie transforms are carried out in Sections 3.7, 3.8 and 3.9, and the guiding-center Lagrangian and Hamiltonian are presented. The Poisson bracket structure is then given in Section 3.10 and the equations of motion are pre-

sented and discussed in Section 3.11. In Section 3.12, a complete summary of the transformation equations is given for reference and the correction to the gyromomentum is derived. In Section 3.13, we show how to write our results in "1 + 3" notation, and we compare our results to those of Northrop [43]. In Section 3.14 we cast all our results in *manifestly* gyrogauge and boostgauge invariant format.

3.2 Conventions and Notation

In this work, we adopt the following conventions: The particle space-time coordinate will be denoted by r^μ, where $\mu = 0,\ldots,3$. The Minkowski metric, $g_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1)$, is used throughout our derivation of the guiding-center equations, but the results will be written in manifestly covariant form so that this assumption can be relaxed. The four potential is given by $A^\mu = (\phi, \mathbf{A})$, so the antisymmetric field tensor is $F = dA$, or

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}. \tag{3.281}$$

The dual field tensor, $\mathcal{F} = {}^*F$, is given by

$$\mathcal{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z & -E_y \\ B_y & -E_z & 0 & E_x \\ B_z & E_y & -E_x & 0 \end{pmatrix}. \tag{3.282}$$

where $\epsilon_{\mu\nu\alpha\beta}$ is the completely antisymmetric fourth rank Levi-Civita tensor with $\epsilon_{0123} = +1$. Note carefully that $\epsilon^{0123} = -1$, thanks to the Minkowski metric.

It is often convenient to use "1+3" notation. Then, the matrix of components

of the mixed field tensor, $F^{\mu}{}_{\nu}$, may be written

$$F = \begin{pmatrix} 0 & \mathbf{E} \\ \mathbf{E} & \mathbf{1} \times \mathbf{B} \end{pmatrix}, \tag{3.283}$$

and that of the mixed dual field tensor, $\mathcal{F}^{\mu}{}_{\nu}$, may be written

$$\mathcal{F} = \begin{pmatrix} 0 & -\mathbf{B} \\ -\mathbf{B} & \mathbf{1} \times \mathbf{E} \end{pmatrix}. \tag{3.284}$$

Note that we have used the notation $(\mathbf{1}\times\mathbf{B})_{ij} = \epsilon_{ikl}\delta_{jk}B_l = \epsilon_{ijl}B_l$. Also note that the mixed field tensors are neither symmetric nor antisymmetric. The advantage to dealing with the mixed tensors is that one may contract them with other tensors using ordinary matrix multiplication. Of course, we could equally well do this with the completely covariant or contravariant forms, but we would have to remember to use the Minkowski metric when multiplying a row by a column.

Thus, when the field tensor is applied to an arbitrary four-vector, the result may be written

$$F \cdot \begin{pmatrix} a \\ \mathbf{a} \end{pmatrix} = \begin{pmatrix} \mathbf{E} \cdot \mathbf{a} \\ a\mathbf{E} + \mathbf{a} \times \mathbf{B} \end{pmatrix}. \tag{3.285}$$

The analogous equation for the dual field tensor is

$$\mathcal{F} \cdot \begin{pmatrix} a \\ \mathbf{a} \end{pmatrix} = \begin{pmatrix} -\mathbf{B} \cdot \mathbf{a} \\ -a\mathbf{B} + \mathbf{a} \times \mathbf{E} \end{pmatrix}. \tag{3.286}$$

This "1 + 3" notation will prove to be useful and convenient throughout the remainder of this thesis.

The two familiar Lorentz scalars can be expressed in terms of these tensors by

$$\lambda_1 \equiv \frac{1}{2}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}F : F = B^2 - E^2, \tag{3.287}$$

and

$$\lambda_2 \equiv \frac{1}{4}\mathcal{F}_{\mu\nu}F^{\mu\nu} = \frac{1}{4}\mathcal{F} : F = \mathbf{E}\cdot\mathbf{B}. \tag{3.288}$$

Note carefully that $F : F \equiv F_{\mu\nu}F^{\mu\nu} = -F_{\mu\nu}F^{\nu\mu} = -\mathrm{Tr}(F\cdot F)$.

The Lorentz equation of motion may then be written

$$m\frac{du}{d\tau} = \frac{e}{c}F(r)\cdot u, \tag{3.289}$$

where

$$u = \frac{dr}{d\tau} \tag{3.290}$$

is the four-velocity, τ is the proper time, m is the rest mass and e is the charge.

Equation (3.289) makes it clear that if the field is independent of space-time position, then the frequencies of the motion are the eigenvalues of F times $-ie/mc$. Now the characteristic equation for the matrix F is

$$\det(F - \lambda\mathbf{1}) = \lambda^4 + \lambda_1\lambda^2 - \lambda_2^2 = 0. \tag{3.291}$$

This biquadratic in λ is easily solved to give $\lambda = \pm\lambda_E$, or $\lambda = \pm i\lambda_B$, where we have defined the Lorentz scalars

$$\lambda_E \equiv \mathrm{sgn}(\lambda_2)\sqrt{\frac{1}{2}(\sqrt{\lambda_1^2 + 4\lambda_2^2} - \lambda_1)}, \tag{3.292}$$

and

$$\lambda_B \equiv \sqrt{\frac{1}{2}(\sqrt{\lambda_1^2 + 4\lambda_2^2} + \lambda_1)}. \tag{3.293}$$

We can write λ_1 and λ_2 in terms of λ_E and λ_B as follows:

$$\lambda_1 = \lambda_B^2 - \lambda_E^2, \tag{3.294}$$

and

$$\lambda_2 = \lambda_B\lambda_E. \tag{3.295}$$

We can now define the two Lorentz scalars

$$\Omega_E \equiv \frac{e\lambda_E}{mc}, \tag{3.296}$$

and

$$\Omega_B \equiv \frac{e\lambda_B}{mc}. \tag{3.297}$$

The first of these is the inverse of the characteristic proper time required to accelerate to relativistic velocities along field lines, while the second is the gyrofrequency with respect to proper time.

3.3 The Electromagnetic Projection Operators

In this section, we summarize the work of Fradkin [13] that is relevent to this study. It is straightforward to verify the following identities:

$$F^2 - \mathcal{F}^2 = -\lambda_1 \mathbf{1}, \tag{3.298}$$

and

$$F \cdot \mathcal{F} = \mathcal{F} \cdot F = -\lambda_2 \mathbf{1}. \tag{3.299}$$

Premultiplying the first of these by F, and employing the second gives

$$F^3 = -\lambda_2 \mathcal{F} - \lambda_1 F. \tag{3.300}$$

Premultiplying by F once again gives

$$F^4 + \lambda_1 F^2 - \lambda_2^2 \mathbf{1} = 0. \tag{3.301}$$

Comparing this with Eq. (3.291), we see that we have proven that F obeys its own characteristic equation, as it must by the Hamilton-Cayley theorem. Now it is clear that Eq. (3.301) may be written as follows:

$$(F - \lambda_E \mathbf{1}) \cdot (F + \lambda_E \mathbf{1}) \cdot (F - i\lambda_B \mathbf{1}) \cdot (F + i\lambda_B \mathbf{1}) = 0, \tag{3.302}$$

and the four factors in this expression commute, so any of them could have been written first. Thus, if Ψ is an arbitrary column four-vector, then

$$(F - \lambda_E \mathbf{1}) \cdot \Big[(F + \lambda_E \mathbf{1}) \cdot (F - i\lambda_B \mathbf{1}) \cdot (F + i\lambda_B \mathbf{1}) \cdot \Psi \Big] = 0, \tag{3.303}$$

so that $(F+\lambda_E\mathbf{1})\cdot(F-i\lambda_B\mathbf{1})\cdot(F+i\lambda_B\mathbf{1})\cdot\Psi$ is an (unnormalized) eigenvector of F with eigenvalue λ_E. Thus, the operator $(F+\lambda_E\mathbf{1})\cdot(F-i\lambda_B\mathbf{1})\cdot(F+i\lambda_B\mathbf{1})$ is a (unnormalized) projection operator that projects arbitrary four-vectors onto the vector subspace spanned by the zeroth eigenvector of F. Proceeding in this manner, it is easy to see that the projection operator

$$P_{\|} = \frac{F^2+\lambda_B^2\mathbf{1}}{\lambda_B^2+\lambda_E^2} \tag{3.304}$$

projects arbitrary four-vectors onto the vector subspace spanned by the eigenvectors of F with eigenvalues $\pm\lambda_E$, while the projection operator

$$P_{\perp} = \frac{-F^2+\lambda_E^2\mathbf{1}}{\lambda_B^2+\lambda_E^2} \tag{3.305}$$

projects arbitrary four-vectors onto the vector subspace spanned by the eigenvectors of F with eigenvalues $\pm i\lambda_B$. The normalization constants were chosen to make the projection operators idempotent; that is

$$P_{\|}\cdot P_{\|} = P_{\|}, \tag{3.306}$$

$$P_{\perp}\cdot P_{\perp} = P_{\perp}, \tag{3.307}$$

$$P_{\|}\cdot P_{\perp} = P_{\perp}\cdot P_{\|} = 0, \tag{3.308}$$

and

$$P_{\|} + P_{\perp} = \mathbf{1}. \tag{3.309}$$

We have thus decomposed the tangent space at each point of space-time into the Cartesian product of two two-dimensional "two-flats." The rapid gyromotion takes place in the perpendicular two-flat since it is spanned by the eigenvectors corresponding to the imaginary eigenvalues, while the parallel motion takes place in the parallel two-flat since it is spanned by the eigenvectors corresponding to the real eigenvalues. These two-flats will play an indispensible role in our theory. We shall use them to isolate the gyrational components of the particle velocity in preparation for the guiding-center Lie transform.

In Section 3.7, we shall order the fields in an expansion parameter and, for reasons that will be explained at that time, we shall demand that our lowest-order field have $\lambda_E = 0$. Furthermore, the two-flats that we shall use will always be defined in terms of the zero-order field; that is, the field tensor that appears on the right hand side of Eqs. (3.304) and (3.305) is always the lowest-order field tensor with $\lambda_E = 0$. Thus, these equations can be simplified to read

$$P_{\parallel} = \mathbf{1} + \frac{F^2}{\lambda_B^2} = \frac{\mathcal{F}^2}{\lambda_B^2} \tag{3.310}$$

and

$$P_{\perp} = -\frac{F^2}{\lambda_B^2} = \mathbf{1} - \frac{\mathcal{F}^2}{\lambda_B^2}. \tag{3.311}$$

In "1 + 3" notation, Eqs. (3.310) and (3.311) become

$$P_{\parallel} = \frac{1}{B^2 - E^2} \begin{pmatrix} B^2 & -\mathbf{E} \times \mathbf{B} \\ \mathbf{E} \times \mathbf{B} & \mathbf{BB} + \mathbf{EE} - E^2 \mathbf{1} \end{pmatrix}, \tag{3.312}$$

and

$$P_{\perp} = \frac{1}{B^2 - E^2} \begin{pmatrix} -E^2 & \mathbf{E} \times \mathbf{B} \\ -\mathbf{E} \times \mathbf{B} & -\mathbf{BB} - \mathbf{EE} + B^2 \mathbf{1} \end{pmatrix}. \tag{3.313}$$

Henceforth, all our results concerning the nature of the two-flats and the unit vectors that span them will contain this assumption that the underlying field tensor has $\lambda_E = 0$.

3.4 The Orthonormal Basis Tetrad

We wish to show how to construct a tetrad of unit vectors such that one pair spans the parallel two-flat while the other pair spans the perpendicular two-flat. Clearly such a tetrad is not unique; it is defined only to within an arbitrary rotation in the perpendicular two-flat, and an arbitrary hyperbolic rotation (boost) in the parallel two-flat. We shall have much more to say about this nonuniqueness later; for now we are simply looking for a way to construct *any* such tetrad.

From the arguments presented in the last section, we know that one way to do this is to examine the eigenvectors of the field tensor. Here we shall take a different approach that is perhaps more physically motivated. Recall that we are dealing with fields for which $E_{\|} = 0$ (if this is true in any one frame, it will be true in all frames because $\mathbf{E} \cdot \mathbf{B}$ is a Lorentz scalar). There exist a set of local "preferred" reference frames for which $\mathbf{E}_{\perp}$ also vanishes; hence there is no electric field at all in these preferred frames. Thus, in a preferred frame, the field tensors may be written in "1 + 3" notation as follows:

$$F = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \times \mathbf{B} \end{pmatrix} \tag{3.314}$$

and

$$\mathcal{F} = \begin{pmatrix} 0 & -\mathbf{B} \\ -\mathbf{B} & \mathbf{0} \end{pmatrix}. \tag{3.315}$$

Also, in a preferred frame, the projection operators have the form

$$P_{\|} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{bb} \end{pmatrix} \tag{3.316}$$

and

$$P_{\perp} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{1} - \mathbf{bb} \end{pmatrix}, \tag{3.317}$$

where

$$\mathbf{b} \equiv \mathbf{B} / |\mathbf{B}| . \tag{3.318}$$

The above forms for the projection operators in a preferred frame make it clear that we can choose the following orthonormal basis tetrad for a preferred frame:

$$\hat{\mathbf{e}}_0 = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}, \quad \hat{\mathbf{e}}_1 = \begin{pmatrix} 0 \\ \mathbf{b} \end{pmatrix}, \tag{3.319}$$

and

$$\hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ \boldsymbol{\tau}_1 \end{pmatrix}, \quad \hat{\mathbf{e}}_3 = \begin{pmatrix} 0 \\ \boldsymbol{\tau}_2 \end{pmatrix}, \tag{3.320}$$

where $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are unit three-vectors perpendicular to $\mathbf{b}$, such that $\{\mathbf{b}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2\}$ constitutes an orthonormal triad in three-dimensional space. We reiterate that the above choice is not unique.

Of course, we would like to be able to construct an orthonormal basis tetrad in an *arbitrary* Lorentz frame. To see how to do this, we consider a Lorentz boost from the above-described preferred frame to a new frame. The Lorentz transformation matrix for a boost is

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\boldsymbol{\beta} \\ -\gamma\boldsymbol{\beta} & \mathbf{1} + (\gamma - 1)\beta^{-2}\boldsymbol{\beta}\boldsymbol{\beta} \end{pmatrix}, \tag{3.321}$$

where the three-vector $\boldsymbol{\beta}$ is the generator of the Lorentz boost (it is the relative velocity of the two reference frames divided by c), and where $\gamma \equiv (1 - \beta^2)^{-1/2}$. This matrix is an element of the Lorentz group because it satisfies $\Lambda^{-1} = g \cdot \Lambda^T \cdot g$ (here we have used a superscripted "T" to denote the transpose operation). See Jackson [44] for more details on the Lorentz group and its generators.

The new field tensor components are then

$$F' = \Lambda \cdot F \cdot \Lambda^{-1} = \begin{pmatrix} 0 & \gamma\boldsymbol{\beta} \times \mathbf{B} \\ \gamma\boldsymbol{\beta} \times \mathbf{B} & \mathbf{1} \times [\gamma\mathbf{B} - (\gamma - 1)\beta^{-2}\boldsymbol{\beta}\boldsymbol{\beta} \cdot \mathbf{B}] \end{pmatrix}. \tag{3.322}$$

In writing this result, we have made use of the vector identity,

$$\boldsymbol{\beta} \times \mathbf{B}\boldsymbol{\beta} - \boldsymbol{\beta}\boldsymbol{\beta} \times \mathbf{B} = \mathbf{1} \times (\beta^2\mathbf{B} - \boldsymbol{\beta}\boldsymbol{\beta} \cdot \mathbf{B}). \tag{3.323}$$

From this result for the field tensor, we see that we can identify the electric and magnetic fields in the new frame as

$$\mathbf{E}' = \gamma\boldsymbol{\beta} \times \mathbf{B} \tag{3.324}$$

and

$$\mathbf{B}' = \gamma\mathbf{B} - (\gamma - 1)\beta^{-2}\boldsymbol{\beta}\boldsymbol{\beta}\cdot\mathbf{B}. \tag{3.325}$$

At this point, there are a number of interesting observations to be made. First note that if $\boldsymbol{\beta}$ is parallel to $\mathbf{B}$ then $\mathbf{E}' = 0$, so the transformation takes us to another preferred frame. Next note that if $\boldsymbol{\beta}$ is perpendicular to $\mathbf{B}$ then $\mathbf{B}$ is parallel to $\mathbf{B}'$. Next note that it is possible to arrive at *any* desired $\mathbf{E}'$ by a transformation with $\boldsymbol{\beta}$ perpendicular to $\mathbf{B}$. Specifically, if we take

$$\boldsymbol{\beta} = -\boldsymbol{\beta}_E, \tag{3.326}$$

where

$$\boldsymbol{\beta}_E \equiv \frac{\mathbf{E}' \times \mathbf{B}'}{B'^2}, \tag{3.327}$$

then it is easy to see that the new electric field is $\mathbf{E}'$. Conversely, if we begin with a frame in which the (perpendicular) electric field is $\mathbf{E}'$, then a Lorentz boost with $\boldsymbol{\beta} = \boldsymbol{\beta}_E$ gets us to a preferred frame.

The orthonormal tetrad in the new frame is then

$$\begin{aligned}
\hat{\mathbf{e}}'_0 &= \Lambda\cdot\hat{\mathbf{e}}_0 = \begin{pmatrix} \gamma_E \\ \gamma_E\boldsymbol{\beta}_E \end{pmatrix} \\
\hat{\mathbf{e}}'_1 &= \Lambda\cdot\hat{\mathbf{e}}_1 = \begin{pmatrix} 0 \\ \mathbf{b} \end{pmatrix},
\end{aligned} \tag{3.328}$$

and

$$\begin{aligned}
\hat{\mathbf{e}}'_2 &= \Lambda\cdot\hat{\mathbf{e}}_2 = \begin{pmatrix} \gamma_E\boldsymbol{\beta}_E\cdot\boldsymbol{\tau}_1 \\ \boldsymbol{\tau}_1 + (\gamma_E - 1)\beta_E^{-2}\boldsymbol{\beta}_E\boldsymbol{\beta}_E\cdot\boldsymbol{\tau}_1 \end{pmatrix} \\
\hat{\mathbf{e}}'_3 &= \Lambda\cdot\hat{\mathbf{e}}_3 = \begin{pmatrix} \gamma_E\boldsymbol{\beta}_E\cdot\boldsymbol{\tau}_2 \\ \boldsymbol{\tau}_2 + (\gamma_E - 1)\beta_E^{-2}\boldsymbol{\beta}_E\boldsymbol{\beta}_E\cdot\boldsymbol{\tau}_2 \end{pmatrix},
\end{aligned} \tag{3.329}$$

where $\gamma_E \equiv (1 - \beta_E^2)^{-1/2}$.

At this point we note that we can choose $\boldsymbol{\tau}_1$ to lie along the direction of $\boldsymbol{\beta}_E$ without any loss of generality. We can now write the results for the unit tetrad

in the general frame, dropping the primes which are no longer needed because *all* quantities will refer to the general frame. Thus

$$\hat{\mathbf{e}}_0 = \begin{pmatrix} \gamma_E \\ \gamma_E \boldsymbol{\beta}_E \end{pmatrix}, \quad \hat{\mathbf{e}}_1 = \begin{pmatrix} 0 \\ \mathbf{b} \end{pmatrix}, \tag{3.330}$$

and

$$\hat{\mathbf{e}}_2 = \begin{pmatrix} \gamma_E \boldsymbol{\beta}_E \\ \gamma_E \hat{\boldsymbol{\beta}}_E \end{pmatrix}, \quad \hat{\mathbf{e}}_3 = \begin{pmatrix} 0 \\ \mathbf{b} \times \hat{\boldsymbol{\beta}}_E \end{pmatrix}, \tag{3.331}$$

where

$$\boldsymbol{\beta}_E \equiv \frac{\mathbf{E} \times \mathbf{B}}{B^2}, \tag{3.332}$$

and $\gamma_E \equiv (1 - \beta_E^2)^{-1/2}$. Here we have also introduced the notation $\hat{\boldsymbol{\beta}}_E$ for a unit vector in the direction of $\boldsymbol{\beta}_E$ if $\boldsymbol{\beta}_E \neq 0$. If $\boldsymbol{\beta}_E = 0$, one may choose $\hat{\boldsymbol{\beta}}_E$ to be any unit three-vector perpendicular to $\mathbf{b}$.

Using Eqs. (3.285) and (3.286), the following useful identities are readily demonstrated:

$$F \cdot \hat{\mathbf{e}}_0 = 0, \quad F \cdot \hat{\mathbf{e}}_1 = 0, \tag{3.333}$$

$$F \cdot \hat{\mathbf{e}}_2 = -\lambda_B \hat{\mathbf{e}}_3, \quad F \cdot \hat{\mathbf{e}}_3 = +\lambda_B \hat{\mathbf{e}}_2, \tag{3.334}$$

and

$$\mathcal{F} \cdot \hat{\mathbf{e}}_0 = -\lambda_B \hat{\mathbf{e}}_1, \quad \mathcal{F} \cdot \hat{\mathbf{e}}_1 = -\lambda_B \hat{\mathbf{e}}_0, \tag{3.335}$$

$$\mathcal{F} \cdot \hat{\mathbf{e}}_2 = 0, \quad \mathcal{F} \cdot \hat{\mathbf{e}}_3 = 0. \tag{3.336}$$

Thus, the field tensor and its dual have the effect of rotating these unit vectors *within* their respective two-flats.

Using Eqs. (3.310) and (3.311), it is easy to verify that $P_\parallel$ leaves $\hat{\mathbf{e}}_0$ and $\hat{\mathbf{e}}_1$ unchanged and annihilates $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$, while $P_\perp$ annihilates $\hat{\mathbf{e}}_0$ and $\hat{\mathbf{e}}_1$ and leaves $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$ unchanged. It is also easy to verify that this tetrad is orthonormal with respect to the Minkowski metric; that is, that

$$\hat{\mathbf{e}}_\mu \cdot \hat{\mathbf{e}}_\nu = g_{\mu\nu}. \tag{3.337}$$

So $\hat{\mathbf{e}}_0$ and $\hat{\mathbf{e}}_1$ span the parallel two-flat, and $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$ span the perpendicular two-flat, as asserted. The geometrical situation is illustrated schematically in Fig. 3.1.

In terms of the $\hat{\mathbf{e}}_\alpha$, the projection operators may be written

$$P_{\|} = -\hat{\mathbf{e}}_0\hat{\mathbf{e}}_0 + \hat{\mathbf{e}}_1\hat{\mathbf{e}}_1 \tag{3.338}$$

and

$$P_{\perp} = \hat{\mathbf{e}}_2\hat{\mathbf{e}}_2 + \hat{\mathbf{e}}_3\hat{\mathbf{e}}_3. \tag{3.339}$$

This should be clear from the geometrical picture, but may also be verified by direct algebra.

When applied to the particle four-velocity, these projection operators will allow us to isolate the rapid gyrational motion in the perpendicular two-flat from the nongyrational motion in the parallel two flat. Thus

$$u = u^\mu \hat{\mathbf{e}}_\mu, \tag{3.340}$$

or, if we introduce polar coordinates (w, θ) for the perpendicular four-velocity components and hyperbolic polar coordinates (k, β) for the parallel velocity components, then we may write

$$u = \hat{\mathbf{e}}_0 k\cosh\beta + \hat{\mathbf{e}}_1 k\sinh\beta - \hat{\mathbf{e}}_2 w\sin\theta - \hat{\mathbf{e}}_3 w\cos\theta \tag{3.341}$$

or

$$u = k\hat{\mathbf{t}} + w\hat{\mathbf{c}}, \tag{3.342}$$

where we have defined

$$\hat{\mathbf{t}} \equiv \hat{\mathbf{e}}_0\cosh\beta + \hat{\mathbf{e}}_1\sinh\beta, \tag{3.343}$$

and

$$\hat{\mathbf{c}} \equiv -\hat{\mathbf{e}}_2\sin\theta - \hat{\mathbf{e}}_3\cos\theta. \tag{3.344}$$

If we also define

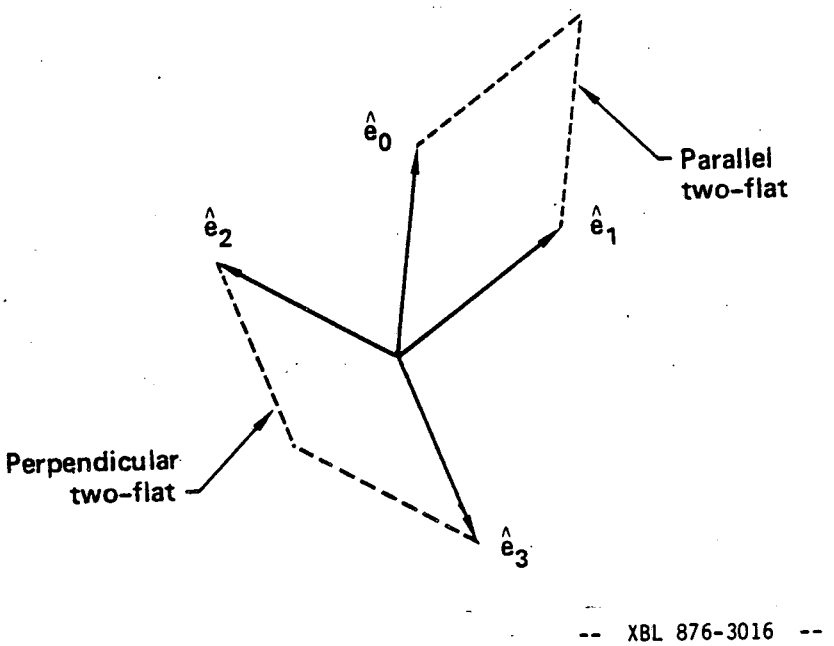


Figure 3.1: The Orthonormal Basis Tetrad

$$\hat{\mathbf{b}} \equiv \hat{\mathbf{e}}_0 \sinh\beta + \hat{\mathbf{e}}_1 \cosh\beta, \tag{3.345}$$

and

$$\hat{\mathbf{a}} \equiv \hat{\mathbf{e}}_2 \cos\theta - \hat{\mathbf{e}}_3 \sin\theta, \tag{3.346}$$

then $(\hat{\mathbf{t}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}, \hat{\mathbf{a}})$ form a new velocity-dependent basis tetrad that is also orthonormal with respect to the Minkowski metric. Please do not confuse the basis four-vector $\hat{\mathbf{b}}$ with the basis three-vector $\mathbf{b}$, and do not confuse the hyperbolic polar coordinate β with the Lorentz transformation generator $\boldsymbol{\beta}$.

Some useful relations among the elements of this new basis tetrad are

$$\frac{\partial \hat{\mathbf{t}}}{\partial \beta} = \hat{\mathbf{b}}, \quad \frac{\partial \hat{\mathbf{b}}}{\partial \beta} = \hat{\mathbf{t}}, \tag{3.347}$$

$$\frac{\partial \hat{\mathbf{c}}}{\partial \theta} = -\hat{\mathbf{a}}, \quad \frac{\partial \hat{\mathbf{a}}}{\partial \theta} = \hat{\mathbf{c}}, \tag{3.348}$$

and

$$F \cdot \hat{\mathbf{t}} = 0, \quad F \cdot \hat{\mathbf{b}} = 0, \tag{3.349}$$

$$F \cdot \hat{\mathbf{c}} = -\lambda_B \hat{\mathbf{a}}, \quad F \cdot \hat{\mathbf{a}} = +\lambda_B \hat{\mathbf{c}}, \tag{3.350}$$

and

$$\mathcal{F} \cdot \hat{\mathbf{t}} = -\lambda_B \hat{\mathbf{b}}, \quad \mathcal{F} \cdot \hat{\mathbf{b}} = -\lambda_B \hat{\mathbf{t}}, \tag{3.351}$$

$$\mathcal{F} \cdot \hat{\mathbf{c}} = 0, \quad \mathcal{F} \cdot \hat{\mathbf{a}} = 0. \tag{3.352}$$

Also, the projection operators may now be written

$$P_{\parallel} = -\hat{\mathbf{t}}\hat{\mathbf{t}} + \hat{\mathbf{b}}\hat{\mathbf{b}} \tag{3.353}$$

and

$$P_{\perp} = \hat{\mathbf{c}}\hat{\mathbf{c}} + \hat{\mathbf{a}}\hat{\mathbf{a}}. \tag{3.354}$$

It is useful to compare the above description of the four-velocity in terms of (k, β, w, θ) with the more conventional "1 + 3" representation, $u = c(\gamma_v, \gamma_v \boldsymbol{\beta}_v)$,

where $\boldsymbol{\beta}_v \equiv \mathbf{v}/c$. We shall do this using the unit tetrad that we constructed above. Combining Eqs. (3.330), (3.331) and (3.341), we find

$$c\gamma_v = \gamma_E(k\cosh\beta - \beta_E w\sin\theta) \tag{3.355}$$

and

$$c\gamma_v\boldsymbol{\beta}_v = \gamma_E(\boldsymbol{\beta}_E k\cosh\beta - \hat{\boldsymbol{\beta}}_E w\sin\theta) - \mathbf{b}\times\hat{\boldsymbol{\beta}}_E w\cos\theta + \mathbf{b}k\sinh\beta. \tag{3.356}$$

From these equations, it follows that

$$\beta_{v1} \equiv \boldsymbol{\beta}_v\cdot\mathbf{b} = \frac{k\sinh\beta}{\gamma_E(k\cosh\beta - \beta_E w\sin\theta)} \tag{3.357}$$

$$\beta_{v2} \equiv \boldsymbol{\beta}_v\cdot\hat{\boldsymbol{\beta}}_E = \frac{\beta_E k\cosh\beta - w\sin\theta}{k\cosh\beta - \beta_E w\sin\theta} \tag{3.358}$$

$$\beta_{v3} \equiv \boldsymbol{\beta}_v\cdot(\mathbf{b}\times\hat{\boldsymbol{\beta}}_E) = \frac{-w\cos\theta}{\gamma_E(k\cosh\beta - \beta_E w\sin\theta)} \tag{3.359}$$

and

$$k = c\gamma_E\gamma_v\sqrt{1-\beta_{v1}^2 - 2\beta_E\beta_{v2} + \beta_E^2\beta_{v1}^2 + \beta_E^2\beta_{v2}^2} \tag{3.360}$$

$$\beta = \tanh^{-1}\left(\frac{\beta_{v1}}{\gamma_E(1-\beta_E\beta_{v2})}\right) \tag{3.361}$$

$$w = c\gamma_E\gamma_v\sqrt{\beta_{v2}^2 + \beta_{v3}^2 - 2\beta_E\beta_{v2} + \beta_E^2 - \beta_E^2\beta_{v3}^2} \tag{3.362}$$

$$\theta = \arg\left(-\beta_{v3} - i\gamma_E(\beta_{v2} - \beta_E)\right) \tag{3.363}$$

Note that the four coordinates (k, β, w, θ) obey the constraint $k^2 - w^2 = c^2$, and this is why they can be determined by the three components of $\boldsymbol{\beta}_v$. Naturally, the above transformation equations depend upon the choice we made for the unit tetrad. This arbitrariness will be discussed further in Section 3.6. These transformation equations will be most useful when we want to compare our results to those of other authors who have used "1 + 3" notation; this will be done in Section 3.13.

3.5 Phase Space Lagrangian for a Charged Particle in an Electromagnetic Field

For a relativistic charged particle in an electromagnetic field, one possible choice for the Hamiltonian, H, in canonical coordinates, (q,p), is given by [44]

$$H(q,p) = \frac{1}{2m}\left(p - \frac{e}{c}A(q)\right)^2, \tag{3.364}$$

and the action one form for canonical coordinates is, by Eq. (2.163)

$$\gamma = p \cdot dq. \tag{3.365}$$

Note that the independent variable is the particle's proper time; the equations of motion are thus of the form of Eq. (2.165), but the dot in that equation now denotes differentiation with respect to proper time.

We begin by making a noncanonical transformation to the new coordinates (r,u), where

$$\begin{cases} r = q \\ u = \frac{1}{m}\left(p - \frac{e}{c}A(q)\right). \end{cases} \tag{3.366}$$

Thus we have eliminated the unphysical canonical momentum, p, in favor of the particle velocity, u. The new Hamiltonian is

$$H'(r,u) = \frac{m}{2}u^2 \tag{3.367}$$

and the new action one form is

$$\gamma' = \left(mu + \frac{e}{c}A(r)\right) \cdot dr. \tag{3.368}$$

If we now use Eq. (3.342) to eliminate the four components of u in favor of (k,β,w,θ), then the new Hamiltonian is

$$H''(r,k,\beta,w,\theta) = \frac{m}{2}(-k^2 + w^2) \tag{3.369}$$

and the new action one form is

$$\gamma'' = \left(mk\hat{\mathbf{t}} + mw\hat{\mathbf{c}} + \frac{e}{c}A(r)\right) \cdot dr. \tag{3.370}$$

It is important to remember that $\hat{\mathbf{t}}$ and $\hat{\mathbf{b}}$ are functions of r and β, and $\hat{\mathbf{c}}$ and $\hat{\mathbf{a}}$ are functions of r and θ. Thus, the second term in the parenthesis on the right hand side of Eq. (3.370) is rapidly oscillating due to its dependence on θ (this will be made more precise shortly). We are now ready to apply the Lie transform procedure that will effectively average H'' and γ'' by transforming to gyrocoordinates in which θ is ignorable.

3.6 Gyrogauge and Boostgauge Transformations

We now discuss the afore-mentioned arbitrariness in choosing the orthonormal unit vectors, $\hat{\mathbf{e}}_\alpha$. A *boostgauge* transformation replaces our choices for $\hat{\mathbf{e}}_0$ and $\hat{\mathbf{e}}_1$ as follows:

$$\hat{\mathbf{e}}_0' = \hat{\mathbf{e}}_0 \cosh\Phi(r) - \hat{\mathbf{e}}_1 \sinh\Phi(r), \tag{3.371}$$

$$\hat{\mathbf{e}}_1' = \hat{\mathbf{e}}_1 \cosh\Phi(r) - \hat{\mathbf{e}}_0 \sinh\Phi(r), \tag{3.372}$$

while a *gyrogauge* transformation replaces our choices for $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$ as follows:

$$\hat{\mathbf{e}}_2' = \hat{\mathbf{e}}_2 \cos\Psi(r) + \hat{\mathbf{e}}_3 \sin\Psi(r), \tag{3.373}$$

$$\hat{\mathbf{e}}_3' = \hat{\mathbf{e}}_3 \cos\Psi(r) - \hat{\mathbf{e}}_2 \sin\Psi(r). \tag{3.374}$$

Note that the new unit vectors are still orthonormal, that $\hat{\mathbf{e}}_0$ and $\hat{\mathbf{e}}_1$ still span the parallel two-flat, and that $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$ still span the perpendicular two-flat. The gyrogauge and boostgauge transformations have simply given each of these two pairs of unit vectors a rotation within its respective two-flat. The amount of

rotation is measured by Φ in the parallel two-flat, and by Ψ in the perpendicular two-flat. Note that these can be functions of the particle's spacetime position, r.

Recall that we used the unit tetrad to decompose the particle velocity into parallel and perpendicular parts, and to coordinatize these by (k, β) and (w, θ), respectively. It is fairly easy to see that the transformation given by Eqs. (3.371) through (3.374) will have no effect on k and w, but will shift β and θ. Hence, we add

$$\beta' = \beta + \Phi(r) \tag{3.375}$$

to our boostgauge transformation equations, and

$$\theta' = \theta + \Psi(r) \tag{3.376}$$

to our gyrogauge transformation equatons. None of the other phase space coordinates are affected by the transformations.

Equations (3.371) through (3.376) constitute the full gyrogauge and boostgauge transformation equations. A quantity that is left unchanged by these transformation equations will be said to be gyrogauge or boostgauge invariant, respectively. The concept of gyrogauge invariance has a nonrelativistic analog which was first discussed by Littlejohn [41]. In the remainder of this section, we shall extend his methods to our relativistic problem.

To begin with, we note that the unit vectors $(\hat{\mathbf{t}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}, \hat{\mathbf{a}})$ are all gyrogauge and boostgauge invariant. This is demonstrated for $\hat{\mathbf{t}}$ as follows:

$$\begin{aligned}
\hat{\mathbf{t}}' &= \hat{\mathbf{e}}_0' \cosh\beta' + \hat{\mathbf{e}}_1' \sinh\beta' \\
&= (\hat{\mathbf{e}}_0 \cosh\Phi - \hat{\mathbf{e}}_1 \sinh\Phi)\cosh(\beta+\Phi) + (\hat{\mathbf{e}}_1 \cosh\Phi - \hat{\mathbf{e}}_0 \sinh\Phi)\sinh(\beta+\Phi) \\
&= \hat{\mathbf{e}}_0[\cosh\Phi\cosh(\beta+\Phi) - \sinh\Phi\sinh(\beta+\Phi)] \\
&\quad + \hat{\mathbf{e}}_1[-\sinh\Phi\cosh(\beta+\Phi) + \cosh\Phi\sinh(\beta+\Phi)] \\
&= \hat{\mathbf{e}}_0 \cosh\beta + \hat{\mathbf{e}}_1 \sinh\beta \\
&= \hat{\mathbf{t}};
\end{aligned} \tag{3.377}$$

the demonstration for the other three unit vectors follows similarly. Because the parallel and perpendicular projection operators may be written in the form of Eqs. (3.353) and (3.354), their gyrogauge and boostgauge invariance is manifest.

The fact that the quantities above are gyrogauge and boostgauge invariant means that they may be expressed in terms of purely physical tensor quantities; more precisely, they may be expressed in terms of quantities that are completely independent of our choice of the orientation of the basis tetrad, $\hat{\mathbf{e}}_\alpha$, at each point in spacetime. For example, $P_\parallel$ and $P_\perp$ can be expressed in terms of the field tensor, as was done in Eqs. (3.304) and (3.305). The gyrogauge and boostgauge invariant quantities k and w can be written in terms of the projection operators and the particle four-velocity with the help of Eq. (3.342)

$$k = \sqrt{-u \cdot P_\parallel \cdot u}, \tag{3.378}$$

and

$$w = \sqrt{u \cdot P_\perp \cdot u}. \tag{3.379}$$

Finally, the members of the tetrad $(\hat{\mathbf{t}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}, \hat{\mathbf{a}})$ can all be expressed in terms of the field tensor and the particle four-velocity, with the help of Eqs. (3.342), (3.350), and (3.351)

$$\hat{\mathbf{t}} = \frac{1}{k} P_\parallel \cdot u, \tag{3.380}$$

$$\hat{\mathbf{b}} = -\frac{1}{\lambda_B} \mathcal{F} \cdot \hat{\mathbf{t}}, \tag{3.381}$$

$$\hat{\mathbf{c}} = \frac{1}{w} P_\perp \cdot u, \tag{3.382}$$

$$\hat{\mathbf{a}} = -\frac{1}{\lambda_B} F \cdot \hat{\mathbf{c}}. \tag{3.383}$$

Now consider the pair of one-forms:

$$\mathcal{Q} \equiv (\vec{\nabla}\, \hat{\mathbf{e}}_1) \cdot \hat{\mathbf{e}}_0 = -(\vec{\nabla}\, \hat{\mathbf{e}}_0) \cdot \hat{\mathbf{e}}_1 = (\vec{\nabla}\, \hat{\mathbf{b}}) \cdot \hat{\mathbf{t}} = -(\vec{\nabla}\, \hat{\mathbf{t}}) \cdot \hat{\mathbf{b}}, \tag{3.384}$$

and

$$\mathcal{R} \equiv (\vec{\nabla}\, \hat{\mathbf{e}}_2) \cdot \hat{\mathbf{e}}_3 = -(\vec{\nabla}\, \hat{\mathbf{e}}_3) \cdot \hat{\mathbf{e}}_2 = (\vec{\nabla}\, \hat{\mathbf{c}}) \cdot \hat{\mathbf{a}} = -(\vec{\nabla}\, \hat{\mathbf{a}}) \cdot \hat{\mathbf{c}}, \tag{3.385}$$

where $\vec{\nabla}$ is a shorthand for the spacetime gradient. It is a straightforward exercise to show that $\mathcal{Q}$ is *not* boostgauge invariant, and that $\mathcal{R}$ is *not* gyrogauge invariant; this is essentially because the spacetime derivatives are taken at constant β and θ, and these latter two quantities are obviously not boostgauge and gyrogauge invariant, respectively. First note that $\vec{\nabla}$ transforms under a general boostgauge and gyrogauge transformation as follows:

$$\vec{\nabla}' = \vec{\nabla} - (\vec{\nabla}\,\Phi)\frac{\partial}{\partial\beta} - (\vec{\nabla}\,\Psi)\frac{\partial}{\partial\theta}, \tag{3.386}$$

where we have made use of Eqs. (3.375) and (3.376). Thus we have

$$\mathcal{Q}' = (\vec{\nabla}'\,\hat{\mathbf{b}}')\cdot\hat{\mathbf{t}}' = [\vec{\nabla}\,\hat{\mathbf{b}} - (\vec{\nabla}\,\Phi)\hat{\mathbf{t}}]\cdot\hat{\mathbf{t}} = \mathcal{Q} + \vec{\nabla}\,\Phi, \tag{3.387}$$

and

$$\mathcal{R}' = (\vec{\nabla}'\,\hat{\mathbf{c}}')\cdot\hat{\mathbf{a}}' = [\vec{\nabla}\,\hat{\mathbf{c}} + (\vec{\nabla}\,\Psi)\hat{\mathbf{a}}]\cdot\hat{\mathbf{a}} = \mathcal{R} + \vec{\nabla}\,\Psi. \tag{3.388}$$

Here we have used Eqs. (3.347) and (3.348). The one-forms $\mathcal{Q}$ and $\mathcal{R}$ will be useful to us momentarily. Furthermore, they have great geometrical significance as will become clear later when we discuss the guiding-center equations of motion.

We now ask what it means for a general one-form in our phase space to be boostgauge and gyrogauge invariant. Using Eq. (2.170), we find that the r component of the one-form transforms as follows:

$$\begin{aligned}\Gamma_r &= \frac{\partial r}{\partial r'}\gamma_r + \frac{\partial\beta}{\partial r'}\gamma_\beta + \frac{\partial\theta}{\partial r'}\gamma_\theta \\ &= \gamma_r - (\vec{\nabla}\,\Phi)\gamma_\beta - (\vec{\nabla}\,\Psi)\gamma_\theta,\end{aligned} \tag{3.389}$$

while all of the other components (k, β, w, and θ) are unchanged. Thus it is clear that the charged particle Hamiltonian and action one form given by Eqs. (3.369) and (3.370) are boostgauge and gyrogauge invariant, since they have no β or θ components.

Now we demand that our Lie transformations, when applied to gauge invariant quantities, preserve their gauge invariance. This, coupled with the established boostgauge and gyrogauge invariance of the particle action one-form, will

guarantee the boostgauge and gyrogauge invariance of the guiding-center action one-form. Suppose that we have a boostgauge and gyrogauge invariant scalar field, f. Applying the Lie derivative operator, $\mathcal{L}_g$, we find from Eq. (2.30)

$$\mathcal{L}_g f = g^r \cdot \vec{\nabla} f + g^k \frac{\partial f}{\partial k} + g^\beta \frac{\partial f}{\partial \beta} + g^w \frac{\partial f}{\partial w} + g^\theta \frac{\partial f}{\partial \theta}. \tag{3.390}$$

If we now subject this to a general boostgauge and gyrogauge transformation, we find

$$\begin{aligned}(\mathcal{L}_g f)' &= g'^r \cdot \vec{\nabla}' f + g'^k \frac{\partial f}{\partial k} + g'^\beta \frac{\partial f}{\partial \beta} + g'^w \frac{\partial f}{\partial w} + g'^\theta \frac{\partial f}{\partial \theta} \\ &= g'^r \cdot \vec{\nabla} f + g'^k \frac{\partial f}{\partial k} + (g'^\beta - \vec{\nabla} \Phi \cdot g'^r) \frac{\partial f}{\partial \beta} + g'^w \frac{\partial f}{\partial w} \\ &\quad + (g'^\theta - \vec{\nabla} \Psi \cdot g'^r) \frac{\partial f}{\partial \theta},\end{aligned} \tag{3.391}$$

where we have made use of the assumed gauge invariance of f. Thus, $\mathcal{L}_g f$ will be gauge invariant if all the components of g are gauge invariant, with the exception of g^β and g^θ which must transform as follows:

$$g'^\beta = g^\beta + \vec{\nabla} \Phi \cdot g'^r, \tag{3.392}$$

and

$$g'^\theta = g^\theta + \vec{\nabla} \Psi \cdot g'^r. \tag{3.393}$$

Thus, if we use a subscripted "0" to denote a gauge invariant quantity, we see that the components of the vector g must be of the form

$$\begin{aligned} g^r &= (g^r)_0 \\ g^k &= (g^k)_0 \\ g^\beta &= (g^\beta)_0 + \mathcal{Q} \cdot (g^r)_0 \\ g^w &= (g^w)_0 \\ g^\theta &= (g^\theta)_0 + \mathcal{R} \cdot (g^r)_0 \end{aligned} \tag{3.394}$$

Using the homotopy formula, it is a straightforward exercise to show that this result is valid not only for gauge invariant scalars, but also for *any* gauge invariant n-form. In particular, this restriction on the form of g is necessary to guarantee the gauge invariance of the Lie transformed action one-form, so we shall demand that it hold in the sections to follow.

3.7 The Zero-Order Problem

We order the particle Hamiltonian and action one-form with the prescription $e \mapsto e/\epsilon$; equivalently, we could say that we are ordering the electromagnetic field at order ϵ^{-1}. The electromagnetic contribution to the canonical momentum thus dominates the kinetic contribution. This ordering procedure has been discussed at length by Kruskal [45] and by Littlejohn [41].

We shall also order the four potential of the electromagnetic field in the parameter ϵ, so

$$A = \sum_{i=0}^{\infty} \epsilon^i A_i. \tag{3.395}$$

Clearly, this induces an ordering of the field itself

$$F = \sum_{i=0}^{\infty} \epsilon^i F_i, \tag{3.396}$$

where

$$F_i = dA_i. \tag{3.397}$$

Henceforth, when we refer to the Lorentz scalars ($\lambda_1, \lambda_2, \lambda_E$ and λ_B) or to the unit basis tetrads or to the projection operators, it is to be understood that they are calculated on the basis of the *zero order* field tensor, F_0.

The Hamiltonian, Eq. (3.369), is thus an order unity scalar. The particle action one-form, Eq. (3.370), may be written

$$\gamma = \frac{1}{\epsilon} \sum_{i=0}^{\infty} \epsilon^i \gamma_i, \tag{3.398}$$

where γ_0 has the component

$$\gamma_{0r} = \frac{e}{c} A_0(r), \tag{3.399}$$

γ_1 has the component

$$\gamma_{1r} = \frac{e}{c} A_1(r) + mk\hat{\mathbf{t}} + mw\hat{\mathbf{c}}, \tag{3.400}$$

and γ_i has the component

$$\gamma_{ir} = \frac{e}{c} A_i(r) \tag{3.401}$$

for $i \geq 2$. All components not listed above are zero.

Suppose that we now write the equations of motion to lowest order as $\omega_0 \cdot \dot{z} = 0$, where $\omega_0 \equiv d\gamma_0$. This turns out to be an instructive exercise even though, as we shall see in a moment, it is somewhat misleading. We see that the only surviving component of ω_0 is

$$\omega_{0rr} = \frac{e}{c} F_0, \tag{3.402}$$

so we get the following equation of motion:

$$F_0 \cdot \dot{r} = 0. \tag{3.403}$$

Now we know that $\dot{t}$ is never zero, so F_0 must have at least one null eigenvector with nonzero time component. In particular, this must be true in a preferred frame, for which $\boldsymbol{\beta}_E = 0$. Thus the parallel two-flat must be the nullspace of F_0. So we demand that

$$\lambda_E = 0, \tag{3.404}$$

where we again emphasize that λ_E is computed from Eqs. (3.282), (3.287), (3.288) and (3.292) using F_0 in place of F. This is a restriction on the allowed zero order fields. It is the relativistic analog of the usual nonrelativistic restriction that $E_{\|} = 0$ to lowest order. Recall that we used this assumption in Section 3.4 when we first discussed the basis tetrads.

Thus, when we order the four potential in ϵ, we must keep in mind that the field derived from A_0 should have no $E_\parallel$. If we have a problem in which there is nonzero $E_\parallel$, then it must be included in A_n where $n \geq 1$. In particular, it could all be put into A_1. The only reason for keeping A_n where $n \geq 2$ in our theory is that sometimes a problem admits another expansion parameter in the field geometry (the stellarator expansion parameter and the long-thin parameter in mirrors are examples), and in some asymptotic theories that other expansion parameter may be taken to be equal to the guiding-center expansion parameter. In such cases, one might want to expand the field in a general power series in ϵ, rather than just restrict oneself to the use of A_0 and A_1.

Thus, Eq. (3.403) constitutes only two independent conditions on the four components of $\dot{r}$. Dotting it with $\hat{\mathbf{c}}$ and $\hat{\mathbf{a}}$ and using Eq. (3.350) gives $\hat{\mathbf{c}} \cdot \dot{r} = \hat{\mathbf{a}} \cdot \dot{r} = 0$, so $\dot{r}$ must lie in the parallel two-flat; that is, the particle motion is constrained to lie along the field lines like that of a bead sliding along a wire. The rapid oscillatory motion is then considered to be a modification to this motion along the field lines, to be transformed away except for the residual perpendicular drifting motion.

What is perhaps most disturbing about Eq. (3.403) is that it gives only two dynamical equations of motion when there are really eight independent phase space coordinates. It gives us no description of the motion along the field lines, and no description of the rate of change of the velocity components. This is because the matrix of components of the zero order Lagrangian two-form is a eight by eight matrix whose rank is only two. This is thus an example of a problem in asymptotics with no well-defined limit problem; this phenomenon is by no means rare and has been discussed in a general context by Kruskal [46].

To get a better idea of what is going on here, we should consider the full particle equations of motion, retaining the lowest order nonzero contributions to

each component of $\omega = d\gamma$, even if some are higher order than others. We find

$$\omega_{rr} = \frac{e}{\epsilon c} F_0 + \mathcal{O}(1), \tag{3.405}$$

$$\omega_{rk} = -m\hat{\mathbf{t}}, \tag{3.406}$$

$$\omega_{r\beta} = -mk\hat{\mathbf{b}}, \tag{3.407}$$

$$\omega_{rw} = -m\hat{\mathbf{c}}, \tag{3.408}$$

$$\omega_{r\theta} = +mw\hat{\mathbf{a}}, \tag{3.409}$$

with all other components vanishing. Forming the equations of motion, $\omega \cdot \dot{z} = \partial H/\partial z$, we find that

$$\dot{r} = k\hat{\mathbf{t}} + w\hat{\mathbf{c}}, \tag{3.410}$$

so there is no longer any ambiguity in the parallel motion. Similarly we can now find the equations of motion for the velocity components. We get

$$\dot{k} = \mathcal{O}(1), \tag{3.411}$$

$$\dot{\beta} = \mathcal{O}(1), \tag{3.412}$$

$$\dot{w} = \mathcal{O}(1), \tag{3.413}$$

and

$$\dot{\theta} = \frac{1}{\epsilon}\Omega_B + \mathcal{O}(1). \tag{3.414}$$

This makes it clear that the dominant motion at lowest order is the gyration, in accordance with our intuition. Thus, as $\epsilon \to 0$, we have the rate of change of θ dominating that of all the other dynamical variables, including r. Hence, averages over the unperturbed motion will simply be averages over θ.

Note that in order to get this zero order equation of motion, we needed γ_r only to order ϵ^{-1}, while all the other components of γ were needed to order unity. This peculiar mixing of orders persists to higher order; so to obtain the n-th order guiding-center equations of motion, we will need γ_r only to order $n-1$, while all the other components of γ will be needed to order n.

3.8 The Preparatory Lie Transform

All treatments of guiding-center motion share one feature in common: In the transformation from particle position, r, to guiding-center position, R, they all include the term, $-w\hat{\mathbf{a}}/\Omega_B$. This is the gyroradius vector, and it is the most intuitive term in the entire guiding-center transformation (indeed, one might argue that it is the *only* intuitive term in the entire guiding-center transformation). We shall make this transformation before we do anything else, as this was found to facilitate the remainder of the calculation in Littlejohn's nonrelativistic treatment [7].

From Eqs. (2.30) and (2.215), we see that, to first order, the difference between z and Z is simply given by the components of the generator vector, g. So since we want to have $R = r - w\hat{\mathbf{a}}/\Omega_B$, we see that we should choose

$$g_p^r = -\frac{w}{\Omega_B}\hat{\mathbf{a}}, \tag{3.415}$$

where the subscript "p" denotes "preparatory."

Now g_p^r is clearly boostgauge and gyrogauge invariant, but from Eq. (3.394) we see that a Lie transform generated by this vector alone would not preserve the gauge invariance of the action one-form. Consequently, we must append the following additional components to g_p:

$$g_p^\beta = -\frac{w}{\Omega_B}\hat{\mathbf{a}}\cdot\mathcal{Q}, \tag{3.416}$$

and

$$g_p^\theta = -\frac{w}{\Omega_B}\hat{\mathbf{a}}\cdot\mathcal{R}. \tag{3.417}$$

First note that the Hamiltonian, Eq. (3.369), is unaffected by the preparatory Lie transform because it is independent of r, β and θ (so $\mathcal{L}_p H'' = 0$). Next, using Eqs. (2.272) through (2.275), we calculate the new action one-form resulting from the transformation generated by this vector. This transformation takes place at

first order only, so we may set $g_1 = g_p$ and $g_2 = g_3 = 0$ in those equations. Also, since we are interested in calculating the guiding-center equations of motion to third order (this turns out to be one order higher than the usual perpendicular drifts), we do not need Γ_{3r}.

At zero order, we have the obvious

$$\Gamma_0 = \gamma_0. \tag{3.418}$$

This has the single nonzero component,

$$\Gamma_{0r} = \frac{e}{c} A_0. \tag{3.419}$$

The corresponding Lagrangian two-form, ω_0, was given in Eq. (3.402).

Moving on to first order, it is readily found that $i_p \omega_0$ (where, in keeping with past convention, $i_p \equiv i_{g_p}$) has only one nonzero component,

$$(i_p \omega_0)_r = mw\hat{\mathbf{c}}. \tag{3.420}$$

We take $S_1 = 0$, so Eq. (2.273) gives the following nonzero component for Γ_1:

$$\Gamma_{1r} = \frac{e}{c} A_1 + mk\hat{\mathbf{t}}. \tag{3.421}$$

Note that the aforementioned rapidly oscillating term, $mw\hat{\mathbf{c}}$, has been removed from γ_{1r} by the transformation.

Before proceeding to second order, we need to calculate $\omega_1 \equiv d\gamma_1$ and $\Omega_1 \equiv d\Gamma_1$. The first of these has the following nonzero components:

$$\begin{aligned} \omega_{1rr} = \frac{e}{c} F_1 &+ mk(\vec{\nabla}\, \hat{\mathbf{t}} - \hat{\mathbf{t}}\, \overleftarrow{\nabla}) \\ &+ mw(\vec{\nabla}\, \hat{\mathbf{c}} - \hat{\mathbf{c}}\, \overleftarrow{\nabla}), \end{aligned} \tag{3.422}$$

$$\omega_{1rk} = -m\hat{\mathbf{t}}, \tag{3.423}$$

$$\omega_{1r\beta} = -mk\hat{\mathbf{b}}, \tag{3.424}$$

$$\omega_{1rw} = -m\hat{\mathbf{c}}, \tag{3.425}$$

$$\omega_{1r\theta} = +mw\hat{\mathbf{a}}. \tag{3.426}$$

The second has the following nonzero components:

$$\Omega_{1rr} = \frac{e}{c}F_1 + mk(\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla}), \tag{3.427}$$

$$\Omega_{1rk} = -m\hat{\mathbf{t}}, \tag{3.428}$$

$$\Omega_{1r\beta} = -mk\hat{\mathbf{b}}. \tag{3.429}$$

Note that we have introduced the notation $\hat{\mathbf{t}}\,\overleftarrow{\nabla}$ for the transpose of $\vec{\nabla}\,\hat{\mathbf{t}}$.

We are now ready to proceed to second order. First note that $\frac{1}{2}i_p\omega_1$ has the following nonzero components:

$$(\frac{1}{2}i_p\omega_1)_r = -\frac{1}{2}\frac{w}{\Omega_B}\hat{\mathbf{a}}\cdot\Big[\frac{e}{c}F_1 + mk(\vec{\nabla}\,\hat{\mathbf{t}}\cdot P_\perp - \hat{\mathbf{t}}\,\overleftarrow{\nabla}) + mw(\vec{\nabla}\,\hat{\mathbf{c}}\cdot P_\parallel - \hat{\mathbf{c}}\,\overleftarrow{\nabla})\Big], \tag{3.430}$$

and

$$(\frac{1}{2}i_p\omega_1)_\theta = -\frac{mw^2}{2\Omega_B}. \tag{3.431}$$

Next note that $\frac{1}{2}i_p\Omega_1$ has the single nonzero component,

$$(\frac{1}{2}i_p\Omega_1)_r = -\frac{1}{2}\frac{w}{\Omega_B}\hat{\mathbf{a}}\cdot\Big[\frac{e}{c}F_1 + mk(\vec{\nabla}\,\hat{\mathbf{t}}\cdot P_\perp - \hat{\mathbf{t}}\,\overleftarrow{\nabla})\Big]. \tag{3.432}$$

Now, using Eq. (2.274) and choosing $S_2 = 0$, we can write down the nonzero components of Γ_2,

$$\Gamma_{2r} = \frac{e}{c}A_2 + \frac{w}{\Omega_B}\hat{\mathbf{a}}\cdot\Big[\frac{e}{c}F_1 + mk(\vec{\nabla}\,\hat{\mathbf{t}}\cdot P_\perp - \hat{\mathbf{t}}\,\overleftarrow{\nabla}) + \frac{mw}{2}(\vec{\nabla}\,\hat{\mathbf{c}}\cdot P_\parallel - \hat{\mathbf{c}}\,\overleftarrow{\nabla})\Big], \tag{3.433}$$

and

$$\Gamma_{2\theta} = \frac{mw^2}{2\Omega_B}. \tag{3.434}$$

Note that Γ_{2r} has rapidly oscillating terms; these will be removed by subsequent Lie transforms. Also note the appearance of the gyromomentum as the θ component of Γ_2.

Moving on to third order, we recall that we do not need Γ_{3r}. Referring to Eq. (2.275), it is easily seen that γ_3 and $i_p\omega_2$ both have only an r-component, so we do not bother with these terms. Then $\frac{1}{3}i_p di_p\omega_1$ has a nonzero r-component which we shall not calculate, and it also has a nonzero θ component given by

$$(\frac{1}{3}i_p di_p\omega_1)_\theta = -\frac{mw^3}{3\Omega_B^3}\hat{\mathbf{a}}\cdot\vec{\nabla}\,\Omega_B - \frac{w^2}{3\Omega_B^2}\hat{\mathbf{c}}\cdot\left[\frac{e}{c}F_1 + mk(\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla})\right]\cdot\hat{\mathbf{a}}. \quad (3.435)$$

Similarly, $\frac{1}{6}i_p di_p\Omega_1$ has a nonzero r-component which we shall not calculate, and it also has a nonzero θ component given by

$$(\frac{1}{6}i_p di_p\Omega_1)_\theta = -\frac{w^2}{6\Omega_B^2}\hat{\mathbf{c}}\cdot\left[\frac{e}{c}F_1 + mk(\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla})\right]\cdot\hat{\mathbf{a}}. \quad (3.436)$$

Taking $S_3 = 0$, we see that the nonzero components of Γ_3 are Γ_{3r} and

$$\Gamma_{3\theta} = -\frac{mw^3}{3\Omega_B^3}\hat{\mathbf{a}}\cdot\vec{\nabla}\,\Omega_B - \frac{w^2}{2\Omega_B^2}\hat{\mathbf{c}}\cdot\left[\frac{e}{c}F_1 + mk\,(\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla})\right]\cdot\hat{\mathbf{a}}. \quad (3.437)$$

Note that this has rapidly oscillating terms which will have to be removed by subsequent Lie transforms. This completes the preparatory transformation.

3.9 The Averaging Lie Transforms

We now perform the averaging Lie transformations that will take us to the guiding-center action one-form. These are somewhat more difficult than the preparatory transformation, since we do not know the generators in advance. For economy of notation, we reset our variables as follows: We shall henceforth refer to the Hamiltonian and action one-form that resulted from the preparatory transformation as H'' and γ, respectively, and these new Lie transforms will take us to H''' and Γ.

First consider the action one form. Once again, nothing changes at order zero, so

$$\Gamma_0 = \gamma_0, \quad (3.438)$$

and the only nonzero component of this is

$$\Gamma_{0r} = \frac{e}{c} A_0. \tag{3.439}$$

The corresponding Lagrangian two-form, ω_0, was given in Eq. (3.402); its only nonzero component was ω_{0rr}.

At order one, we take $g_1^r = 0$ and $S_1 = 0$ because we have already succeeded in averaging Γ_{1r} by the preparatory transformation, and we don't want to ruin this. It follows that $i_1\omega_0 = 0$, and so $\Gamma_1 = \gamma_1$. The only nonvanishing component of Γ_1 is then

$$\Gamma_{1r} = \frac{e}{c} A_1 + mk\hat{\mathbf{t}}. \tag{3.440}$$

Note that we have not yet had to specify g_1^k, g_1^β, g_1^w, or g_1^θ, since it is clear that these have no effect on Γ_1. These components of g_1 will be useful in the averaging of Γ_2. Also note that $\Omega_1 = \omega_1$ is given by Eqs. (3.427) through (3.429).

A word of caution is in order concerning the coordinate τ. It is not altered in any way by the transformation. This means that after we complete the transformation to guiding-center coordinates, τ will still be the *single-particle* proper time; it will *not* be the guiding-center proper time. So $g_{\mu\nu} dr^\mu dr^\nu = -d\tau^2$, but $g_{\mu\nu} dR^\mu dR^\nu \neq -d\tau^2$. Thus, throughout the remainder of this calculation, it is best to regard τ as simply an orbit parameter, devoid of relevant physical significance.

Now we proceed to second order. Note that $i_2\omega_0$ has only an r-component,

$$(i_2\omega_0)_r = \frac{e}{c} g_2^r \cdot F_0. \tag{3.441}$$

Next note that $\frac{1}{2} i_1\Omega_1 = \frac{1}{2} i_1\omega_1$ has the following nonzero component:

$$(\frac{1}{2} i_1\omega_1)_r = \frac{1}{2}(mkg_1^\beta \hat{\mathbf{b}} + mg_1^k \hat{\mathbf{t}}), \tag{3.442}$$

We then take $S_2 = 0$ because we have already succeeded in averaging $\Gamma_{2\theta}$ by the preparatory transformation, and we don't want to ruin this. Equation (2.274)

then gives the following nonzero components for Γ_2:

$$\Gamma_{2r} = \frac{e}{c}A_2 + \frac{w}{\Omega_B}\hat{\mathbf{a}}\cdot\left[\frac{e}{c}F_1 + mk(\vec{\nabla}\,\hat{\mathbf{t}}\cdot P_\perp - \hat{\mathbf{t}}\,\overleftarrow{\nabla}) + \frac{mw}{2}(\vec{\nabla}\,\hat{\mathbf{c}}\cdot P_\parallel - \hat{\mathbf{c}}\,\overleftarrow{\nabla})\right] - \frac{e}{c}g_2^r\cdot F_0 - mkg_1^\beta\hat{\mathbf{b}} - mg_1^k\hat{\mathbf{t}}, \quad (3.443)$$

and

$$\Gamma_{2\theta} = \frac{mw^2}{2\Omega_B}. \quad (3.444)$$

We now proceed to third order, and once again we do not need the r-component of Γ_3. Referring to Eq. (2.275), it is easily seen that $i_3\omega_0$ has only an r-component, so we do not bother with this term. Then $i_2\Omega_1 = i_2\omega_1$ has a nonzero r-component which we shall not calculate; its other nonzero components are

$$(i_2\omega_1)_k = -mg_2^r\cdot\hat{\mathbf{t}}, \quad (3.445)$$

and

$$(i_2\omega_1)_\beta = -mkg_2^r\cdot\hat{\mathbf{b}}. \quad (3.446)$$

Next, $i_1\omega_2$ has a nonzero r-component which we shall not calculate; its other nonzero components are

$$(i_1\omega_2)_w = -\frac{mw}{\Omega_B}g_1^\theta, \quad (3.447)$$

and

$$(i_1\omega_2)_\theta = +\frac{mw}{\Omega_B}g_1^w. \quad (3.448)$$

Next, $\frac{1}{3}i_1di_1(\omega_1 + \frac{1}{2}\Omega_1) = \frac{1}{2}i_1di_1\omega_1$ has a nonzero r-component which we shall not calculate; it has no other nonzero components. From Eq. (2.275) we see that the nonzero components of Γ_3 are Γ_{3r} and the following:

$$\Gamma_{3k} = mg_2^r\cdot\hat{\mathbf{t}} + \frac{\partial S_3}{\partial k}, \quad (3.449)$$

$$\Gamma_{3\beta} = mkg_2^r\cdot\hat{\mathbf{b}} + \frac{\partial S_3}{\partial\beta}, \quad (3.450)$$

$$\Gamma_{3w} = \frac{mw}{\Omega_B} g_1^\theta + \frac{\partial S_3}{\partial w}, \tag{3.451}$$

and

$$\Gamma_{3\theta} = -\frac{w^2}{2\Omega_B^2} \hat{\mathbf{c}} \cdot \left[\frac{e}{c} F_1 + mk(\vec{\nabla}\, \hat{\mathbf{t}} - \hat{\mathbf{t}}\, \overleftarrow{\nabla}) \right] \cdot \hat{\mathbf{a}} - \frac{mw^3}{3\Omega_B^3} \hat{\mathbf{a}} \cdot \vec{\nabla}\, \Omega_B - \frac{mw}{\Omega_B} g_1^w + \frac{\partial S_3}{\partial \theta}. \tag{3.452}$$

Now we apply the Lie transform to the Hamiltonian. This is straightforward, and we get

$$H''' = H_1''' + \epsilon H_2''' + \mathcal{O}(\epsilon^2), \tag{3.453}$$

where

$$H_1''' = H'' = m(-k^2 + w^2)/2, \tag{3.454}$$

and

$$H_2''' = mkg_1^k - mwg_1^w. \tag{3.455}$$

Thus, the Hamiltonian, which emerged unscathed from the preparatory Lie transform, may indeed be modified by the averaging Lie transform.

We must now choose the vector generator components, $g_1^k, g_1^\beta, g_1^w, g_1^\theta$,and g_2^r, and the scalar gauge transformation generator, S_3, in order to average and maximally simplify $\Gamma_{2r}, H_2''', \Gamma_{3k}, \Gamma_{3\beta}, \Gamma_{3w}$, and $\Gamma_{3\theta}$. These are given by Eqs. (3.443), (3.455), (3.449), (3.450), (3.451), and (3.452), respectively. We proceed by taking the averaged parts of these equations,

$$\Gamma_{2r} = \frac{e}{c} A_2 - \frac{mw^2}{2\Omega_B} \left[\mathcal{R} - \frac{1}{2} \left(\hat{\mathbf{a}} \cdot \vec{\nabla}\, \hat{\mathbf{c}} - \hat{\mathbf{c}} \cdot \vec{\nabla}\, \hat{\mathbf{a}} \right) \cdot P_\parallel \right] - \frac{e}{c} \bar{g}_2^r \cdot F_0 - m\bar{g}_1^k \hat{\mathbf{t}} - mk\bar{g}_1^\beta \hat{\mathbf{b}}, \tag{3.456}$$

$$H_2''' = mk\bar{g}_1^k - mw\bar{g}_1^w, \tag{3.457}$$

$$\Gamma_{3k} = m\bar{g}_2^r \cdot \hat{\mathbf{t}} + \frac{\partial \bar{S}_3}{\partial k}, \tag{3.458}$$

$$\Gamma_{3\beta} = mk\bar{g}_2^r \cdot \hat{\mathbf{b}} + \frac{\partial \bar{S}_3}{\partial \beta}, \tag{3.459}$$

$$\Gamma_{3w} = \frac{mw}{\Omega_B}\bar{g}_1^{\theta} + \frac{\partial \bar{S}_3}{\partial w}, \tag{3.460}$$

$$\Gamma_{3\theta} = -\frac{mw^2}{4\Omega_B^3}\left(\frac{e}{mc}\right) F_0 : \left[\left(\frac{e}{mc}\right) F_1 + mk\left(\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla}\right)\right] - \frac{mw}{\Omega_B}\bar{g}_1^{\bar{w}} + \frac{\partial \bar{S}_3}{\partial \theta}, \tag{3.461}$$

and the fluctuating parts,

$$0 = \frac{w}{\Omega_B}\hat{\mathbf{a}}\cdot\left[\frac{e}{c} + mk\left(\vec{\nabla}\,\hat{\mathbf{t}}\cdot P_\perp - \hat{\mathbf{t}}\,\overleftarrow{\nabla}\right)\right] + \frac{mw^2}{4\Omega_B}\left(\hat{\mathbf{a}}\cdot\vec{\nabla}\,\hat{\mathbf{c}} + \hat{\mathbf{c}}\cdot\vec{\nabla}\,\hat{\mathbf{a}}\right)\cdot P_\parallel - \frac{e}{c}\tilde{g}_2^{r}\cdot F_0 - m\tilde{g}_1^{k}\hat{\mathbf{t}} - mk\tilde{g}_1^{\beta}\hat{\mathbf{b}}, \tag{3.462}$$

$$0 = mk\tilde{g}_1^{k} - mw\tilde{g}_1^{\tilde{w}}, \tag{3.463}$$

$$0 = m\tilde{g}_2^{r}\cdot\hat{\mathbf{t}} + \frac{\partial \tilde{S}_3}{\partial k}, \tag{3.464}$$

$$0 = mk\tilde{g}_2^{r}\cdot\hat{\mathbf{b}} + \frac{\partial \tilde{S}_3}{\partial \beta}, \tag{3.465}$$

$$0 = \frac{mw}{\Omega_B}\tilde{g}_1^{\theta} + \frac{\partial \tilde{S}_3}{\partial w}, \tag{3.466}$$

$$0 = -\frac{mw^3}{3\Omega_B^3} - \frac{mw}{\Omega_B}\tilde{g}_1^{\tilde{w}} + \frac{\partial \tilde{S}_3}{\partial \theta}, \tag{3.467}$$

where we have demanded that the Hamiltonian and one-form components themselves be purely averaged. In the above equations, an overbar denotes the averaged part of a quantity, while an overtilde denotes the fluctuating part.

Solve Eq. (3.467) for $\partial\tilde{S}_3/\partial\theta$ in terms of $\tilde{g}_1^{\tilde{w}}$. Then use Eq. (3.463) to get $\tilde{g}_1^{\tilde{w}}$ in terms of $\tilde{g}_1^{k}$. Then dot Eq. (3.462) with $\hat{\mathbf{t}}$ in order to get $\tilde{g}_1^{k}$. The result is

$$\frac{\partial \tilde{S}_3}{\partial \theta} = \frac{mw^3}{3\Omega_B^3}\hat{\mathbf{a}}\cdot\vec{\nabla}\,\Omega_B - \frac{k}{\Omega_B}\left\{\frac{w}{\Omega_B}\left[\frac{e}{c}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{t}} - mk(tta)\right] + \frac{mw^2}{4\Omega_B}\left[(act) + (cat)\right]\right\}, \tag{3.468}$$

where the abbreviation (act) is shorthand for $\hat{\mathbf{a}}\cdot\vec{\nabla}\,\hat{\mathbf{c}}\cdot\hat{\mathbf{t}}$, etc. Now this equation

is easily integrated to give

$$\tilde{S}_3 = -\frac{mw^3}{3\Omega_B^3}\hat{\mathbf{c}}\cdot\vec{\nabla}\,\Omega_B + \frac{wk}{\Omega_B^2}\left[\frac{e}{c}\hat{\mathbf{c}}\cdot F_1\cdot\hat{\mathbf{t}} - mk(ttc)\right] + \frac{mw^2k}{8\Omega_B^2}\left[(ata)+(ctc)\right]. \tag{3.469}$$

We can now back substitute to get the oscillatory parts of the vector generator components,

$$\tilde{g_1^k} = \frac{w}{\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}} + \frac{kw}{\Omega_B}(tta) - \frac{w^2}{4\Omega_B}\left[(act)+(cat)\right], \tag{3.470}$$

$$\tilde{g_1^\beta} = \frac{w}{k\lambda_B}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{b}} - \frac{w}{\Omega_B}(bta) + \frac{w^2}{4k\Omega_B}\left[(acb)+(cab)\right], \tag{3.471}$$

$$\tilde{g_1^w} = \frac{k}{\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}} + \frac{k^2}{\Omega_B}(tta) - \frac{kw}{4\Omega_B}\left[(act)+(cat)\right], \tag{3.472}$$

$$\tilde{g_1^\theta} = \frac{w}{\Omega_B^2}\hat{\mathbf{c}}\cdot\vec{\nabla}\,\Omega_B - \frac{k}{w\lambda_B}\hat{\mathbf{c}}\cdot F_1\cdot\hat{\mathbf{t}} + \frac{k^2}{w\Omega_B}(ttc) - \frac{k}{4\Omega_B}\left[(ata)-(ctc)\right], \tag{3.473}$$

and

$$\tilde{g_2^r} = \frac{w}{\lambda_B\Omega_B}\left(P_\parallel - P_\perp\right)\cdot F_1\cdot\hat{\mathbf{c}} + \frac{w^2}{8\Omega_B^2}\left(\hat{\mathbf{a}}\cdot\vec{\nabla}\,\hat{\mathbf{a}} - \hat{\mathbf{c}}\cdot\vec{\nabla}\,\hat{\mathbf{c}}\right)\cdot P_\parallel + \frac{kw}{\Omega_B^2}\Big[\left(\hat{\mathbf{c}}\cdot\vec{\nabla}\,\hat{\mathbf{t}}\cdot P_\perp - P_\perp\cdot\vec{\nabla}\,\hat{\mathbf{t}}\cdot\hat{\mathbf{c}}\right) - \left(\hat{\mathbf{t}}\cdot\vec{\nabla}\,\hat{\mathbf{c}}\cdot P_\parallel - P_\parallel\cdot\vec{\nabla}\,\hat{\mathbf{c}}\cdot\hat{\mathbf{t}}\right)\Big]. \tag{3.474}$$

Next we consider the equations for the averaged parts of the generators, Eqs. (3.456) through (3.461). These constitute nine equations (Eq. (3.456) is really four equations) in seventeen unknowns (the nine components of Γ, and the eight components of $\bar{g}$). Thus, we can choose eight unknowns at will. So we demand

$$\Gamma_{2r} = \frac{e}{c}A_2 - \frac{mw^2}{2\Omega_B}\mathcal{R}, \tag{3.475}$$

$$\Gamma_{3k} = 0, \tag{3.476}$$

$$\Gamma_{3\beta} = 0, \tag{3.477}$$

$$\Gamma_{3w} = 0, \tag{3.478}$$

and

$$\Gamma_{3\theta} = 0. \tag{3.479}$$

Here we have retained the term involving $\mathcal{R}$ in Γ_{2r} in order to preserve boostgauge and gyrogauge invariance, according to Eq. (3.389). Taking $\bar{S}_3 = 0$, we can now solve for $\bar{g}$. We get

$$\bar{g_2^r} = 0, \tag{3.480}$$

$$\bar{g_1^k} = \frac{w^2}{4\Omega_B}\left[(cat) - (act)\right], \tag{3.481}$$

$$\bar{g_1^\beta} = -\frac{w^2}{4k\Omega_B}\left[(cab) - (acb)\right], \tag{3.482}$$

$$\bar{g_1^w} = \frac{w}{2\lambda_B}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{c}} + \frac{kw}{2\Omega_B}\left[(atc) - (cta)\right], \tag{3.483}$$

$$\bar{g_1^\theta} = 0. \tag{3.484}$$

We can now solve for H_2''' using Eq. (3.457) to get

$$\begin{aligned} H_2''' &= -\frac{mw^2}{2\lambda_B}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{c}} - \frac{mkw^2}{4\Omega_B}\left[(atc) - (cta)\right] \\ &= \frac{mw^2}{4\Omega_B^2}\left(\frac{e}{mc}\right)F_0 : \left[\left(\frac{e}{mc}\right)F_1 + \frac{k}{2}\left(\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla}\right)\right]. \end{aligned} \tag{3.485}$$

This completes the averaging transformation.

Henceforth, we shall write transformed quantities as functions of the guiding-center variables (R, K, $\mathcal{B}$, W, θ) instead of their lower-case counterparts. Note that this has no mathematical significance, and is done only to emphasize the *physical* interpretation of the various quantities that emerge from the theory. We regard functions in the mathematicians' sense of the word: functional arguments are nothing more than dummy placeholders.

We may now write out the full guiding-center Hamiltonian and action one form to the above-described order. We have

$$H_2''' = \frac{m}{2}(-K^2 + W^2) + \epsilon\frac{mW^2}{4\Omega_B^2}\left(\frac{e}{mc}\right)F_0 : \left[\left(\frac{e}{mc}\right)F_1\right.$$

$$+\frac{K}{2}\left(\vec{\nabla}\,\hat{\mathbf{t}}-\hat{\mathbf{t}}\,\overleftarrow{\nabla}\right)\Big]+O(\epsilon^2). \tag{3.486}$$

and

$$\Gamma=\left[\frac{e}{\epsilon c}\left(A_0+\epsilon A_1+\epsilon^2 A_2\right)+mK\hat{\mathbf{t}}-\frac{\epsilon mW^2}{2\Omega_B}\mathcal{R}+O(\epsilon^2)\right]\cdot dR$$
$$+\epsilon\frac{mW^2}{2\Omega_B}d\Theta+\mathcal{O}(\epsilon^3). \tag{3.487}$$

Note that θ is an ignorable coordinate, so that its canonically conjugate momentum, $\mu\equiv mW^2/2\Omega_B$, is conserved. This can now be identified as the gyromomentum, and it is useful to eliminate the coordinate W in favor of μ. The results will be denoted

$$H_{gc}=-\frac{m}{2}K^2+\mu\Omega_B$$
$$+\frac{\epsilon\mu}{2\lambda_B}F_0:\left[\left(\frac{e}{mc}\right)F_1+\frac{K}{2}\left(\vec{\nabla}\,\hat{\mathbf{t}}-\hat{\mathbf{t}}\,\overleftarrow{\nabla}\right)\right]+O(\epsilon^2) \tag{3.488}$$

and

$$\Gamma_{gc}=\left[\frac{e}{\epsilon c}A+mK\hat{\mathbf{t}}-\epsilon\mu\mathcal{R}+O(\epsilon^2)\right]\cdot dR+\epsilon\mu d\Theta. \tag{3.489}$$

This is the form of the guiding-center Hamiltonian and action one form that will be used in subsequent sections. Note that the order ϵ term in the Hamiltonian may be neglected if only the classical drifts (usual gradient, polarization and curvature drifts) are desired.

3.10 The Guiding-Center Poisson Brackets

As a first step towards writing down the guiding-center equations of motion, we form the guiding-center Lagrangian two-form. The nonzero components are

$$\Omega_{RR}=\frac{e}{\epsilon c}(F_0+\epsilon F')+mK(\vec{\nabla}\,\hat{\mathbf{t}}-\hat{\mathbf{t}}\,\overleftarrow{\nabla}), \tag{3.490}$$

$$\Omega_{RK}=-m\hat{\mathbf{t}}, \tag{3.491}$$

$$\Omega_{RB}=-mK\hat{\mathbf{b}}, \tag{3.492}$$

$$\Omega_{R\mu} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\mathcal{R} & \text{(higher order)}, \end{cases} \tag{3.493}$$

$$\Omega_{\mu\Theta} = \epsilon, \tag{3.494}$$

where

$$F' \equiv \begin{cases} F_1 & \text{(classical order)} \\ F_1 + \epsilon F_2 - \frac{\epsilon c}{e}\mu\mathcal{N} & \text{(higher order)}, \end{cases} \tag{3.495}$$

and

$$\mathcal{N} \equiv d\mathcal{R}. \tag{3.496}$$

Here we have drawn a distinction between two cases, just as we did with the Hamiltonian. Terms of *classical* order are all that are necessary to retain if only the usual gradient, curvature and polarization drifts are desired. If one would like the equations of motion to one order higher than that, one must also retain the terms labelled *higher* order. This makes a difference only in $\Omega_{R\mu}$ and in the definition of F'.

Now we can get the Poisson brackets using Eq. (2.168). We do this by inverting the eight by eight matrix consisting of the components of Ω. This is a tedious but straightforward exercise, and the nonvanishing results are presented below. We have performed this matrix inversion for both the classical-order and the higher-order cases separately.

$$\{R, R\} = -\frac{\epsilon F_0}{m\lambda_B\Omega_B\Upsilon}, \tag{3.497}$$

$$\{R, K\} = -\frac{\hat{\mathbf{t}}}{m}\cdot\Xi, \tag{3.498}$$

$$\{R, \mathcal{B}\} = \frac{\hat{\mathbf{b}}}{mK}\cdot\Xi, \tag{3.499}$$

$$\{R, \Theta\} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\{R, R\}\cdot\mathcal{R} & \text{(higher order)}, \end{cases} \tag{3.500}$$

$$\{K, \mathcal{B}\} = -\frac{e}{m^2ck}\hat{\mathbf{t}}\cdot\Xi\cdot F''\cdot\hat{\mathbf{b}}, \tag{3.501}$$

$$\{K,\Theta\} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\{K,R\}\cdot\mathcal{R} & \text{(higher order)}, \end{cases} \tag{3.502}$$

$$\{\mathcal{B},\Theta\} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\{\mathcal{B},R\}\cdot\mathcal{R} & \text{(higher order)}, \end{cases} \tag{3.503}$$

and

$$\{\Theta,\mu\} = \epsilon^{-1}, \tag{3.504}$$

where we have defined the scalar

$$\Upsilon \equiv 1 + \frac{\epsilon F_0 : F''}{2\lambda_B^2}, \tag{3.505}$$

and the tensors

$$\Xi \equiv \mathbf{1} + \frac{\epsilon F''\cdot F_0}{\lambda_B^2\Upsilon} \tag{3.506}$$

and

$$F'' \equiv F' + \frac{mcK}{e}(\overrightarrow{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla}), \tag{3.507}$$

and where F' is given by Eq. (3.495). Note carefully that the bracket of R with R is nonzero because R is really four coordinates; thus $\{R,R\}$ is a four by four antisymmetric matrix and, consequently, its diagonal elements vanish but the rest of it may be nonzero.

Note that Θ and μ are decoupled from the other dynamical variables at the classical order, but that Θ is not decoupled at higher order. The reason for this will be clarified shortly, but for now we note that this coupling is not at all problematic. The important point is that the set of functions of R, K and $\mathcal{B}$ form a subset of the set of all phase functions that is a closed Lie subalgebra under the operation of these Poisson brackets. Then, since our Hamiltonian is independent of Θ, we can eliminate that degree of freedom and still have a valid Hamiltonian system for guiding centers. This is an example of the *reduction* of a Hamiltonian system, discussed in Chapter 2.

Next note that we could have expanded all of the above expressions in pure power series in ϵ. For example, Υ appears in the denominators of several brackets, and consists of an order one term and an order ϵ term. One might argue that, since our expressions are valid only to a certain power of ϵ anyway, we ought to expand this in powers of ϵ. There is, however, a compelling reason not to do this: The above brackets are guaranteed to obey the Jacobi identity *exactly* because they are elements of the inverse matrix of the matrix of components of the Lagrange tensor which obeys $d\Omega = dd\Gamma = 0$. If we were to expand the brackets in ϵ, and retain ϵ only to a certain power, then the Jacobi identity would be satisfied only to that power of ϵ. Now one might counter that in an asymptotic theory of this nature, that is all we have a right to demand. In practice, however, guiding-center equations of motion are often integrated numerically, and violations of the Jacobi identity invalidate Liouville's theorem which guarantees phase space area preservation. This, in turn, can lead to an observed "fuzziness" of KAM tori which might cause one to draw erroneous conclusions about the presence of stochasticity.

To elaborate on this last point, in studies of mirror-confined plasmas, for example, one might integrate the guiding-center equations numerically and produce a "puncture plot" of the places where the trajectory of the guiding center intersects the midplane of the device. If such a plot exhibits stochasticity, one might well expect the radial transport of the plasma to be enhanced significantly as compared to a case for which the plot is a smooth KAM surface. Thus, in a study of mirror plasma radial transport, one might vary some parameter to see for what value this transition from regular to stochastic motion takes place. The decision might be made by comparing the numerically-generated puncture-plots for several different parameter values in some range. Yet if one uses guiding-center equations of motion that do not satisfy Liouville's theorem *exactly*, one runs the risk of misinterpreting "fuzziness" in plots that is due only to violations

of Liouville's theorem (which is, after all, the only reason that KAM tori exist in the first place) as the presence of true stochasticity.

This is why we inverted the Lagrange tensor for the classical and the higher-order cases *separately*, rather than do a single inversion for the higher-order case and truncate to get the classical case. As things stand, the brackets for both cases presented above are guaranteed to satisfy the Jacobi identity *exactly*.

3.11 Guiding-Center Equations of Motion

These brackets together with the Hamiltonian, Eq. (3.488), give the guiding-center equations of motion according to Eq. (2.169). First consider the equation for $\dot{R}$. To the classical order, this may be written

$$\begin{aligned}\dot{R} &= \{R,R\}\cdot\mu\,\vec{\nabla}\,\Omega_B - \{R,K\}mK \\ &= K\hat{\mathbf{t}} + \frac{\epsilon}{\lambda_B^2\Upsilon}\left(K\hat{\mathbf{t}}\cdot F'' + \frac{c}{e}\mu\,\vec{\nabla}\,\Omega_B\right)\cdot F_0. \end{aligned} \tag{3.508}$$

The first term contains the usual parallel motion and the $\mathbf{E}\times\mathbf{B}$ drift. The order ϵ contribution consists of two parts: The first contains the relativistic analog of the curvature and polarization drifts (they are in F''), and the second is the relativistic analog of the grad-B drift; these statements will be clarified when we cast these results in "1 + 3" notation. Of course, the above apparatus is sufficient to get $\dot{R}$ to one order higher than this, but the expression itself is rather unenlightening to look at, so we shall not bother to write it down.

The equations for $\dot{K}$ and $\dot{\mathcal{B}}$ are then

$$\begin{aligned}\dot{K} &= \{K,R\}\cdot\mu\,\vec{\nabla}\,\Omega_B \\ &= \frac{\mu}{m}\hat{\mathbf{t}}\cdot\Xi\cdot\vec{\nabla}\,\Omega_B \end{aligned} \tag{3.509}$$

and

$$\dot{\mathcal{B}} = \{\mathcal{B},R\}\cdot\mu\,\vec{\nabla}\,\Omega_B - \{\mathcal{B},K\}mK$$

$$= -\frac{\mu}{mK}\hat{\mathbf{b}}\cdot\Xi\cdot\vec{\nabla}\,\Omega_B - \frac{e}{mc}\hat{\mathbf{t}}\cdot\Xi\cdot F''\cdot\hat{\mathbf{b}}. \tag{3.510}$$

The terms containing $\vec{\nabla}\,\Omega_B$ contain the mirroring force, and the contribution of F_1 contains the force due to the parallel electric field; once again, these statements will be clarified when we cast these results in "1 + 3" notation.

Next note that $\dot{\mu}$ is exactly zero, even at the higher order; this, of course, was our aim all along. The higher order equation of motion for Θ is

$$\dot{\Theta} = \frac{1}{\epsilon}\Omega_B + \epsilon\mathcal{R}\cdot\dot{R} + \frac{\epsilon}{2\lambda_B}F_0 : \left[\left(\frac{e}{mc}\right)F_1 + \frac{K}{2}\left(\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla}\right)\right]. \tag{3.511}$$

The first term is the lowest-order gyromotion. The second term arises from the bracket structure, and corrects for the possibility that as the guiding-center moves in R, the perpendicular unit vectors upon which the definition of Θ is based may rotate within the perpendicular two-flat. This term arose from our demand of boostgauge and gyrogauge invariance, and it is the reason that the Poisson bracket of Θ with R, K, and $\mathcal{B}$ cannot vanish at higher order. The necessity of this has been discussed by Littlejohn [41] and by Hagan and Frieman [42].

The third term on the right side of Eq. (3.511) arises from the first-order piece of the Hamiltonian and consists of two subterms in the square brackets. The first of these subterms is the correction to the gyrofrequency due to F_1. To see this, define the *total gyrofrequency* due to both F_0 and F_1 by $\Omega_{BT} \equiv e\lambda_{BT}/mc$, where λ_{BT} is given by Eq. (3.293). We quickly find

$$\begin{aligned}\Omega_{BT} &= \frac{e}{mc}\sqrt{\frac{1}{2}(F_0+\epsilon F_1) : (F_0+\epsilon F_1)} + \mathcal{O}(\epsilon^2)\\ &= \frac{e}{mc}\sqrt{\lambda_B^2 + \epsilon F_0 : F_1} + \mathcal{O}(\epsilon^2)\\ &= \Omega_B + \frac{\epsilon}{2\lambda_B}\left(\frac{e}{mc}\right)F_0 : F_1 + \mathcal{O}(\epsilon^2).\end{aligned} \tag{3.512}$$

The second subterm of the third term on the right of Eq. (3.511) is the gyrofrequency shift due to gradients of the perpendicular electric field. This is not expected to be obvious, and will be discussed further in Section 3.13, when we cast our results in "1 + 3" notation.

The geometrical significance of the second term in Eq. (3.511) is illustrated in Fig. 3.2 (here we temporarily revert to using lower-case r and θ). In order to compare the unit tetrad at one point in spacetime, r, with that at another point, $r + \delta r$, (to see how much it rotated) we need some way of transporting the unit vectors from one point to another. The correct way of doing this was elucidated by Littlejohn [41]. Since we have assumed flat spacetime throughout this calculation, we can simply translate the unit vector $\hat{\mathbf{e}}_2$ from r to $r + \delta r$ in the usual manner of Euclidean geometry. Of course, when we arrive at $r + \delta r$, the translated unit vector, called $\hat{\mathbf{e}}_2^*$, will not be the same as the unit vector $\hat{\mathbf{e}}_2$. Furthermore, it need not even lie in the perpendicular two-flat. To remedy this, we project it onto the perpendicular two-flat and normalize the result to get a new unit vector, called $\hat{\mathbf{e}}_2^{**}$.The angle between $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_2^{**}$ at the point $r + \delta r$ is defined to be $\delta\theta$. The calculation goes as follows:

$$\begin{aligned}
\hat{\mathbf{e}}_2^*(r+\delta r) &= \hat{\mathbf{e}}_2(r) \\
&= \hat{\mathbf{e}}_2(r+\delta r-\delta r) \\
&= \hat{\mathbf{e}}_2(r+\delta r) - \delta r\cdot\vec{\nabla}\,\hat{\mathbf{e}}_2(r+\delta r)+ \\
&\quad \frac{1}{2}\delta r\delta r :\vec{\nabla}\vec{\nabla}\,\hat{\mathbf{e}}_2(r+\delta r)+\cdots .
\end{aligned} \tag{3.513}$$

Henceforth, all quantities are evaluated at the point $r + \delta r$ so this will not be noted explicitly. Continuing,

$$\begin{aligned}
\hat{\mathbf{e}}_2^{**} &\equiv \frac{P_\perp\cdot\hat{\mathbf{e}}_2^*}{|P_\perp\cdot\hat{\mathbf{e}}_2^*|} \\
&= \hat{\mathbf{e}}_2 - \delta r\cdot\vec{\nabla}\,\hat{\mathbf{e}}_2\cdot\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3 + \frac{1}{2}\delta r\delta r :\vec{\nabla}\vec{\nabla}\,\hat{\mathbf{e}}_2\cdot\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3 \\
&\quad -\frac{1}{2}\left(\delta r\cdot\vec{\nabla}\,\hat{\mathbf{e}}_2\cdot\hat{\mathbf{e}}_3\right)^2\hat{\mathbf{e}}_2+\cdots .
\end{aligned} \tag{3.514}$$

Thus

$$\cos\delta\theta = 1 - \frac{\delta\theta^2}{2} + \cdots$$

$$
\begin{aligned}
&= \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_2^{**} \\
&= 1 - \frac{1}{2}\left(\delta r \cdot \vec{\nabla}\, \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_3\right)^2 + \cdots \\
&= 1 - \frac{1}{2}\left(\mathcal{R} \cdot \delta r\right)^2 + \cdots,
\end{aligned} \tag{3.515}
$$

so we identify

$$\delta\theta = \mathcal{R} \cdot \delta r. \tag{3.516}$$

This is the change in θ due to the rotation of the unit vectors alone, and it explains the second term on the right of Eq. (3.511). A similar term, $\mathcal{Q} \cdot \dot{R}$, would appear in the equation of motion of $\mathcal{B}$ if we went to higher order.

It was noted by Littlejohn [41] that the one-form, $\mathcal{R}$ is the potential for the gauge field $\mathcal{N} = d\mathcal{R}$ which obeys the field equation $d\mathcal{N} = dd\mathcal{R} = 0$. In the relativistic problem, we also have the gauge field $\mathcal{M} = d\mathcal{Q}$, and this also obeys $d\mathcal{M} = dd\mathcal{Q} = 0$. These are the gauge fields corresponding to the boostgauge and gyrogauge gauge groups. Note that $\mathcal{M}$ and $\mathcal{N}$ are gauge invariant even though $\mathcal{Q}$ and $\mathcal{R}$ are not. Thus, they can be expressed in terms of the field tensor directly; in index notation

$$\mathcal{M}_{\mu\nu} = \frac{1}{\lambda_B} \mathcal{F}^{\alpha}{}_{\beta} P_{\parallel}^{\gamma\beta}{}_{,\mu} P_{\parallel\gamma\alpha,\nu}, \tag{3.517}$$

and

$$\mathcal{N}_{\mu\nu} = \frac{1}{\lambda_B} F^{\alpha}{}_{\beta} P_{\perp}^{\gamma\beta}{}_{,\mu} P_{\perp\gamma\alpha,\nu}. \tag{3.518}$$

The $\dot{R}$ term of the guiding-center Lagrangian, Eq. (3.489), thus couples the two gauge potentials, A and $\mathcal{R}$, and the coupling constant is the gyromomentum.

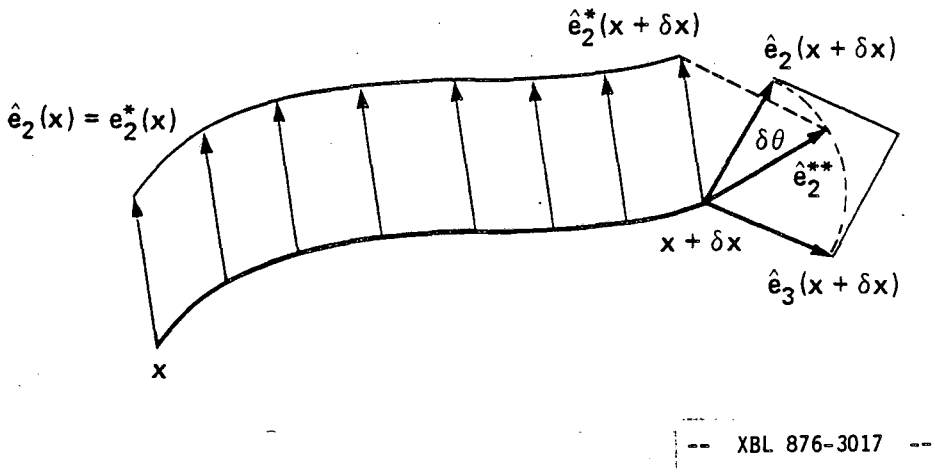


Figure 3.2: Change in Gyroangle due to Rotation of Basis Tetrad as Guiding Center Moves in Spacetime

3.12 Summary of Guiding-Center Transformation

The entire transformation that we have made from the particle coordinates may be written in the form of Eq. (2.214) as follows:

$$Z = \exp(\epsilon\mathcal{L}_{g_p})\exp(\epsilon\mathcal{L}_{g_1})\exp(\epsilon^2\mathcal{L}_{g_2})\exp(\epsilon^3\mathcal{L}_{g_3})\cdots z. \tag{3.519}$$

It is possible to expand these equations in ϵ, and plug in our expressions for the generators to get the coordinate transformation equations. For reference, we present these here:

$$\begin{aligned} R = r &- \frac{\epsilon w\hat{\mathbf{a}}}{\Omega_B} + \epsilon^2\Big\{\frac{w^2}{2\Omega_B}\hat{\mathbf{a}}\cdot\vec{\nabla}\left(\frac{\hat{\mathbf{a}}}{\Omega_B}\right) \\ &+ \frac{w^2\hat{\mathbf{a}}\cdot\mathcal{R}\hat{\mathbf{c}}}{2\Omega_B^2} + \frac{w}{\lambda_B\Omega_B}(P_\parallel - P_\perp)\cdot F_1\cdot\hat{\mathbf{c}} \\ &- \frac{2kw}{\Omega_B^2}(ttc)\hat{\mathbf{t}} + \frac{w^2}{8\Omega_B^2}\left[(ata)-(ctc)\right]\hat{\mathbf{t}} \\ &+ \frac{kw}{\Omega_B^2}\left[(btc)+(tbc)\right]\hat{\mathbf{b}} - \frac{w^2}{8\Omega_B^2}\left[(aba)-(cbc)\right]\hat{\mathbf{b}} \\ &+ \frac{kw}{\Omega_B^2}\left[(cta)-(atc)\right]\hat{\mathbf{a}}\Big\} + O(\epsilon^3), \end{aligned} \tag{3.520}$$

$$\begin{aligned} K = k + \epsilon\Big[&\frac{kw}{\Omega_B}(tta) - \frac{w^2}{2\Omega_B}(act) \\ &+ \frac{w}{\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}}\Big] + O(\epsilon^2), \end{aligned} \tag{3.521}$$

$$\begin{aligned} \mathcal{B} = \beta + \epsilon\Big[&\frac{w^2}{2k\Omega_B}(acb) - \frac{w}{\Omega_B}\left[(bta)-(atb)\right] \\ &- \frac{w}{k\lambda_B}\hat{\mathbf{b}}\cdot F_1\cdot\hat{\mathbf{a}}\Big] + O(\epsilon^2), \end{aligned} \tag{3.522}$$

$$\begin{aligned} W = w + \epsilon\Big[&\frac{k^2}{\Omega_B}(tta) - \frac{kw}{4\Omega_B}\left[3(act)-(cat)\right] \\ &+ \frac{k}{\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}} + \frac{w}{2\lambda_B}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{c}}\Big] + O(\epsilon^2), \end{aligned} \tag{3.523}$$

and

$$\Theta = \theta + \epsilon\Big[\frac{k^2}{w\Omega_B}(ttc) - \frac{k}{4\Omega_B}[(ata) - (ctc)] - \frac{w}{\Omega_B}(aca) + \frac{w}{\Omega_B^2}\hat{\mathbf{c}}\cdot\vec{\nabla}\,\Omega_B + \frac{k}{w\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{c}}\Big] + O(\epsilon^2). \tag{3.524}$$

In the above equations, the capitalized variables are the guiding-center coordinates and the lower case variables are the particle coordinates; it is emphasized that *all* quantities on the right hand sides of these equations (e.g. unit vectors, field tensor, etc.) are evaluated at the particle coordinates. The inverse transformation is given by

$$\begin{aligned} r = R + \frac{\epsilon W\hat{\mathbf{a}}}{\Omega_B} + \epsilon^2\Big\{ & \frac{W^2}{2\Omega_B}\hat{\mathbf{a}}\cdot\vec{\nabla}\left(\frac{\hat{\mathbf{a}}}{\Omega_B}\right) - \frac{W}{\lambda_B\Omega_B}(P_\parallel - P_\perp)\cdot F_1\cdot\hat{\mathbf{c}} \\ & + \hat{\mathbf{t}}\Big[\frac{2KW}{\Omega_B^2}(ttc) - \frac{W^2}{8\Omega_B^2}[(ata) - (ctc)]\Big] \\ & + \hat{\mathbf{b}}\Big[\frac{W^2}{8\Omega_B^2}[(aba) - (cbc)] - \frac{KW}{\Omega_B^2}[(btc) + (tbc)]\Big] \\ & + \hat{\mathbf{c}}\Big[-\frac{W^2}{\Omega_B^3}\hat{\mathbf{c}}\cdot\vec{\nabla}\,\Omega_B + \frac{K}{\lambda_B\Omega_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{c}} - \frac{K^2}{\Omega^2}(ttc) \\ & + \frac{KW}{4\Omega_B^2}[(ata) - (ctc)] + \frac{W^2}{2\Omega_B^2}(aca)\Big] \\ & + \hat{\mathbf{a}}\Big[-\frac{W}{2\lambda_B\Omega_B}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{c}} - \frac{K}{\lambda_B\Omega_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}} \\ & - \frac{K^2}{\Omega^2}(tta) + \frac{KW}{4\Omega_B^2}[3(cat) - (act)]\Big]\Big\} + O(\epsilon^3), \end{aligned} \tag{3.525}$$

$$k = K - \epsilon\Big[\frac{KW}{\Omega_B}(tta) - \frac{W^2}{2\Omega_B}(act) + \frac{W}{\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}}\Big] + O(\epsilon^2), \tag{3.526}$$

$$\beta = \mathcal{B} - \epsilon\Big[\frac{W^2}{2K\Omega_B}(acb) - \frac{W}{\Omega_B}[(bta) - (atb)] - \frac{W}{K\lambda_B}\hat{\mathbf{b}}\cdot F_1\cdot\hat{\mathbf{a}}\Big] + O(\epsilon^2), \tag{3.527}$$

$$w = W - \epsilon\Big[\frac{K^2}{\Omega_B}(tta) - \frac{KW}{4\Omega_B}[3(act) - (cat)] + \frac{K}{\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}} + \frac{W}{2\lambda_B}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{c}}\Big] + O(\epsilon^2), \tag{3.528}$$

and

$$\theta = \Theta - \epsilon\Big[\frac{K^2}{W\Omega_B}(ttc) - \frac{K}{4\Omega_B}[(ata) - (ctc)] - \frac{W}{\Omega_B}(aca) + \frac{W}{\Omega_B^2}\hat{\mathbf{c}}\cdot\vec{\nabla}\,\Omega_B + \frac{K}{W\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{c}}\Big] + O(\epsilon^2). \tag{3.529}$$

In the above equations, everything on the right is evaluated at the guiding-center position.

Recall that the gyromomentum in guiding-center coordinates is given by $mW^2/2\Omega_B$. In particle coordinates, this may be written

$$\mu_{\text{part}} = \frac{mw^2}{2\Omega_B} + \epsilon\Big\{\frac{mw^3}{2\Omega_B^3}\hat{\mathbf{a}}\cdot\vec{\nabla}\,\Omega_B + \frac{mw}{\Omega_B}\Big[\frac{w}{2\lambda_B}\hat{\mathbf{a}}\cdot F_1\cdot\hat{\mathbf{c}} + \frac{k}{\lambda_B}\hat{\mathbf{t}}\cdot F_1\cdot\hat{\mathbf{a}} + \frac{k^2}{\Omega_B}(tta) - \frac{kw}{4\Omega_B}[3(act) - (cat)]\Big]\Big\} + O(\epsilon^2). \tag{3.530}$$

This expression is useful because it gives the conserved quantity in terms of particle coordinates.

3.13 Comparison with Three-Vector Formulations

In order to compare our results with the three-vector formulation given by Northrop [43], we must be able to cast our results into "1 + 3" notation. We learned how to do this for the particle coordinates back at the end of Section 3.4 where we gave the explicit transformation equations, Eqs. (3.357) through (3.363). These are scalar equations in phase space, and so they will retain their form under the guiding-center Lie transform. We need only to replace (k, β, w, θ) by $(K, \mathcal{B}, W, \Theta)$, and to reinterpret $\boldsymbol{\beta}_v$ as the *guiding-center* three-velocity (divided by c). Then we can write down the equations of motion for $\boldsymbol{\beta}_v$ by differentiating Eqs. (3.357) through (3.359) with respect to proper time, using the known equations of motion for the guiding-center coordinates, and expressing the results back in terms of $\boldsymbol{\beta}_v$ by using Eqs. (3.360) through (3.363).

The above-described program seems rather tedious. Fortunately, there are two things that we can do to simplify the task. First, we need only check our results to the order of the classical drifts. This is the order given in the text by Northrop [43]. Second, we can check our results in one of the "preferred" frames of reference, as were described back in Section 3.4. If they hold there, they have to hold in all other frames as well because our results are in manifestly covariant format. These two simplifications make the problem straightforward.

First note that in a preferred frame $\boldsymbol{\beta}_E = 0$, so Eqs. (3.357) through (3.363) become

$$\beta_{v1} = \tanh \mathcal{B} \tag{3.531}$$

$$\beta_{v2} = -\frac{W \sin \Theta}{K \cosh \mathcal{B}} \tag{3.532}$$

$$\beta_{v3} = -\frac{W \cos \Theta}{K \cosh \mathcal{B}} \tag{3.533}$$

and

$$K = c\gamma_v \sqrt{1 - \beta_{v1}^2} \tag{3.534}$$

$$\mathcal{B} = \tanh^{-1} \beta_{v1} \tag{3.535}$$

$$W = c\gamma_v \sqrt{\beta_{v2}^2 + \beta_{v3}^2} \tag{3.536}$$

$$\Theta = \arg(-\beta_{v3} - i\beta_{v2}), \tag{3.537}$$

where, as noted in the last paragraph, all variables are now *guiding-center* variables. In particular, the equations

$$K \cosh \mathcal{B} = c\gamma_v \tag{3.538}$$

and

$$K \sinh \mathcal{B} = c\gamma_v \beta_{v1} = \gamma_v v_\parallel, \tag{3.539}$$

where $v_\parallel \equiv c\beta_{v1}$, will turn out to be particularly useful. The quantity $\mathcal{B}$ is sometimes called the *rapidity*.

Next note that, in a preferred frame, the unit vectors that we constructed in Eqs. (3.330) and (3.331) can be inserted into Eqs. (3.343) and (3.345) to yield

$$\hat{\mathbf{t}} = \begin{pmatrix} \cosh\mathcal{B} \\ \mathbf{b}\sinh\mathcal{B} \end{pmatrix}, \quad \hat{\mathbf{b}} = \begin{pmatrix} \sinh\mathcal{B} \\ \mathbf{b}\cosh\mathcal{B} \end{pmatrix}. \tag{3.540}$$

These will also be useful in what follows.

Now examine Eq. (3.508). We can consider the terms individually. First

$$K\hat{\mathbf{t}} = \begin{pmatrix} \gamma_v c \\ \gamma_v v_\parallel \mathbf{b} \end{pmatrix} \tag{3.541}$$

follows immediately. Next

$$\begin{aligned} \hat{\mathbf{t}}\cdot F'' &= \hat{\mathbf{t}}\cdot F_1 + \frac{mcK}{e}\hat{\mathbf{t}}\cdot\vec{\nabla}\,\hat{\mathbf{t}} \\ &= \hat{\mathbf{t}}\cdot F_1 + \frac{mcK}{e}\left(\cosh\mathcal{B}\frac{1}{c}\frac{\partial}{\partial t} + \sinh\mathcal{B}\mathbf{b}\cdot\nabla\right)\hat{\mathbf{t}} \\ &= \hat{\mathbf{t}}\cdot F_1 + \frac{mcv_\parallel}{eK}\gamma_v^2\begin{pmatrix} 0 \\ \frac{\partial\mathbf{b}}{\partial t} + v_\parallel\mathbf{b}\cdot\nabla\mathbf{b} \end{pmatrix} \\ &\quad + \frac{mc}{eK}\gamma_v^2\begin{pmatrix} 0 \\ \frac{\partial\mathbf{u}_E}{\partial t} + v_\parallel\mathbf{b}\cdot\nabla\mathbf{u}_E \end{pmatrix}, \end{aligned} \tag{3.542}$$

where

$$\mathbf{u}_E \equiv c\frac{\mathbf{E}\times\mathbf{B}}{B^2}, \tag{3.543}$$

also follows after a short computation. Note that $\mathbf{u}_E$ vanishes in a preferred frame, but its derivatives may not; thus we had to apply the derivative to $\hat{\mathbf{t}}$ *before* specializing to a preferred frame.

Next we write the components of F_1 as follows

$$F_1 = \begin{pmatrix} 0 & \mathbf{E}_1 \\ \mathbf{E}_1 & \mathbf{1}\times\mathbf{B}_1 \end{pmatrix}. \tag{3.544}$$

Recall that $\mathbf{E}_1$ must contain all of the parallel electric field.

It now follows from Eq. (3.508) that

$$c\dot{t} = c\gamma_v + \mathcal{O}(\epsilon) \tag{3.545}$$

and

$$\dot{\mathbf{R}} = \gamma_v v_\parallel \mathbf{b}_T + \frac{\epsilon}{B}\mathbf{b}_T \times \left\{ \frac{mc}{e}\gamma_v^2 \left[v_\parallel \left(\frac{\partial \mathbf{b}_T}{\partial t} + v_\parallel \mathbf{b}_T \cdot \nabla \mathbf{b}_T \right) + \left(\frac{\partial \mathbf{u}_E}{\partial t} + v_\parallel \mathbf{b}_T \cdot \nabla \mathbf{u}_E \right) \right] + \frac{\mu}{m}\nabla B \right\} + \mathcal{O}(\epsilon^2), \tag{3.546}$$

where

$$\mathbf{b}_T \equiv \frac{\mathbf{B} + \epsilon \mathbf{B}_1}{|\mathbf{B} + \epsilon \mathbf{B}_1|}. \tag{3.547}$$

Now take the perpendicular part of $\dot{\mathbf{R}}$ by dotting it with $\mathbf{1} - \mathbf{b}_T\mathbf{b}_T$, then divide by $\dot{t}$ to get

$$\frac{d\mathbf{R}_\perp}{dt} = \frac{\epsilon}{\Omega_B}\mathbf{b}_T \times \left\{ \gamma_v \left[v_\parallel \left(\frac{\partial \mathbf{b}_T}{\partial t} + v_\parallel \mathbf{b}_T \cdot \nabla \mathbf{b}_T \right) + \left(\frac{\partial \mathbf{u}_E}{\partial t} + v_\parallel \mathbf{b}_T \cdot \nabla \mathbf{u}_E \right) \right] + \frac{\mu}{m\gamma_v}\nabla \Omega_B \right\} + \mathcal{O}(\epsilon^2). \tag{3.548}$$

This is identical to Eq. (1.76) in the text by Northrop [43] in a preferred frame. Recall that $\lambda_B = B$ in a preferred frame, so that Ω_B in the above equation is simply eB/mc. The classical curvature, gradient and polarization drifts are readily visible in the above equation. If we had instead done the calculation for a general frame of reference, the $\mathbf{E} \times \mathbf{B}$ drift would appear as well. The reader is referred to Northrop [43] for a good discussion of these results.

Next differentiate $\gamma_v v_\parallel = K \sinh \mathcal{B}$ to get

$$\frac{d}{dt}(\gamma_v v_\parallel) = \frac{1}{\gamma_v}(\dot{K} \sinh \mathcal{B} + K\dot{\mathcal{B}} \cosh \mathcal{B}). \tag{3.549}$$

Insert Eqs. (3.509) and (3.510) for $\dot{K}$ and $\dot{\mathcal{B}}$, respectively, and after a little algebra we find

$$\frac{d}{dt}(\gamma_v v_\parallel) = \frac{1}{\gamma_v}\left(-\frac{\mu}{m}\mathbf{b}_T \cdot \nabla \Omega_B - \frac{e}{m}\gamma_v \hat{\mathbf{t}} \cdot F_1 \cdot \hat{\mathbf{b}} \right) + \mathcal{O}(\epsilon). \tag{3.550}$$

Now it follows from Eq. (3.544) that

$$\hat{\mathbf{t}} \cdot F_1 \cdot \hat{\mathbf{b}} = -\mathbf{b} \cdot \mathbf{E}_1 = -E_\parallel, \tag{3.551}$$

So we finally have

$$\frac{d}{dt}(\gamma_v v_\parallel) = -\frac{\mu}{m\gamma_v}\mathbf{b}_T \cdot \nabla\Omega_B + \frac{e}{m}E_\parallel + \mathcal{O}(\epsilon). \tag{3.552}$$

This is identical to Eq. (1.77) in the text by Northrop [43] in a preferred frame. The terms on the right are the mirroring force and the force due to the parallel electric field, respectively.

Northrop's Eq. (1.78) is immediately seen to be equivalent to the fact that our gyromomentum μ is a constant of the motion. Note that Northrop's magnetic moment M_r is related to our μ as follows: $M_r = e\mu/mc$.

Next, we know from Eq. (3.538) that $c\gamma_v = K\cosh\mathcal{B}$, so

$$\frac{d}{dt}(mc^2\gamma_v) = \frac{mc}{\gamma_v}(\dot{K}\cosh\mathcal{B} + K\sinh\mathcal{B}\dot{\mathcal{B}}). \tag{3.553}$$

Now use Eqs. (3.509), (3.510) and (3.551) to get

$$\frac{d}{dt}(mc^2\gamma_v) = \frac{\mu}{\gamma_v}\frac{\partial\Omega_B}{\partial t} + ev_\parallel E_\parallel + \mathcal{O}(\epsilon) \tag{3.554}$$

after a short calculation. This is identical to Eq. (1.79) in the text by Northrop [43] in a preferred frame.

Finally, as promised, we discuss the nonrelativistic limit of the second subterm of the third term on the right side of Eq. (3.511). This term is given by $(K/4\lambda_B)F_0 : (\overrightarrow{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla})$. To simplify the evaluation of this term, we specialize to a preferred frame where the perpendicular electric field vanishes (though we shall be careful to retain its gradient). We also specialize to the case of time-independent fields, spatially uniform magnetic field, and zero parallel velocity. These assumptions are not at all necessary; they serve only to simplify an otherwise tedious calculation, to aid the reader in seeing an effect that would otherwise

be masked by lots of other less interesting terms, and to facilitate comparison with Appendix C. Under these circumstances, we find that

$$\vec{\nabla}\, K\hat{\mathbf{t}} = c\gamma_v \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \nabla\boldsymbol{\beta}_E \end{pmatrix}, \tag{3.555}$$

and

$$F_0 = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{1}\times\mathbf{b} \end{pmatrix}. \tag{3.556}$$

It then follows after a short calculation that

$$\frac{K}{4\lambda_B} F_0 : (\vec{\nabla}\,\hat{\mathbf{t}} - \hat{\mathbf{t}}\,\overleftarrow{\nabla}) = \frac{c\gamma_v}{2\lambda_B}\mathbf{B}\cdot(\nabla\times\boldsymbol{\beta}_E) = -\frac{e\gamma_v}{2m\Omega_B}\nabla_\perp\cdot\mathbf{E}_\perp. \tag{3.557}$$

Except for the factor γ_v, which is clearly a relativistic effect, this is identical to the gyrofrequency shift due to perpendicular electric fields that is derived in Appendix C. This shift was discovered by Kaufman [47] in 1960, who also showed that it gives rise to the phenomenon of *gyroviscosity.*

The reader is urged to consult the text by Northrop [43] as well as a paper by Vandervoort [48] for a further discussion and alternative presentation of the above results.

3.14 Manifestly Boostgauge and Gyrogauge Invariant Format

The guiding-center equations of motion presented above contain expressions, such as $\vec{\nabla}\,\hat{\mathbf{t}}$, that are not boostgauge or gyrogauge invariant. Of course, the equations as a whole are guaranteed to be gauge invariant by our method of derivation; but they are not *manifestly* so. This is due to the fact that our chosen coordinates, namely $(R, K, \mathcal{B}, \mu, \Theta)$, are themselves not gauge invariant, thanks to the inclusion of $\mathcal{B}$ and Θ. This observation suggests that if we were to transform to a new

set of gauge invariant coordinates, we could write our results in manifestly gauge invariant format; that is, without any mention of the unit vectors, $\hat{\mathbf{e}}_\alpha$. In this section, we shall derive two new versions of the Poisson brackets: The first will be manifestly boostgauge invariant, but it will not be manifestly gyrogauge invariant. The second will be both manifestly boostgauge invariant and manifestly gyrogauge invariant.

3.14.1 Manifest Boostgauge Invariance

To get manifestly boostgauge invariant results, we would like to replace K and $\mathcal{B}$ by the new boostgauge invariant coordinate

$$U \equiv K\hat{\mathbf{t}}. \tag{3.558}$$

The inverse transformation would then be

$$K = \sqrt{-U^2} \tag{3.559}$$

and

$$\mathcal{B} = \tanh^{-1}\left(-\frac{U\cdot\hat{\mathbf{e}}_1(R)}{U\cdot\hat{\mathbf{e}}_0(R)}\right). \tag{3.560}$$

Alas, there is a problem with this approach. Since the new coordinate U is a four vector, it contains four degrees of freedom, whereas K and $\mathcal{B}$ represent only two degrees of freedom. This discrepency stems from the fact that U is not an arbitrary four vector because it is constrained to lie in the parallel two flat; that is, it obeys the constraint equation

$$P_\perp(R)\cdot U = 0. \tag{3.561}$$

This constraint restricts U to two degrees of freedom, but it also means that the coordinates R and U are no longer *independent* variables. The coordinate transformation is not a diffeomorphism (it is injective rather than bijective) and so we cannot proceed in the usual manner.

We can remedy this difficulty by temporarily relaxing the constraint in Eq. (3.561). We make the following coordinate transformation (where, for clarity, we use primes to distinguish the new coordinates):

$$\begin{aligned} R' &= R \\ U' &= K\hat{\mathbf{t}}(R,\mathcal{B}) + \frac{C_{1a}}{\lambda_B(R)}\hat{\mathbf{c}}(R,\Theta) + C_{1b}\hat{\mathbf{a}}(R,\Theta) \\ \mu' &= \mu \\ \Theta' &= \Theta. \end{aligned} \tag{3.562}$$

The reason for including λ_B in the second term on the right hand side of the equation for U' will become clear in the next subsection. The inverse transformation is then

$$\begin{aligned} R &= R' \\ K &= \sqrt{-U' \cdot P_{\parallel}(R') \cdot U'} \\ \mathcal{B} &= \tanh^{-1}\left(-\frac{U' \cdot \hat{\mathbf{e}}_1(R')}{U' \cdot \hat{\mathbf{e}}_0(R')}\right) \\ \mu &= \mu' \\ \Theta &= \Theta' \\ C_{1a} &= \lambda_B(R')U' \cdot \hat{\mathbf{c}}(R',\Theta') \\ C_{1b} &= U' \cdot \hat{\mathbf{a}}(R',\Theta'). \end{aligned} \tag{3.563}$$

Here, U' is no longer constrained to lie in the parallel two-flat, and its perpendicular components are called C_{1a}/λ_B and C_{1b}. In order to have the same number of variables before and after the transformation, we have appended C_{1a} and C_{1b} to our usual set of variables before making the transformation.

We now have a diffeomorphism, but we still have to decide how to deal with these two new variables in the unprimed system. Our strategy will be to demand that they are Casimir functions. That way, the dynamics is constrained to lie on hypersurfaces for which they both are constant. If we start the phase space

trajectory on the hypersurface for which they are both zero, it will remain on that hypersurface. Of course, the equations of motion that we end up with will also be capable of describing dynamics on other hypersurfaces for which they are nonzero, but we ignore these other orbits as physically irrelevent.

So our phase space coordinates before this transformation are now taken to be $(R, K, \mathcal{B}, \mu, \Theta, C_{1a}, C_{1b})$. The bracket relations among these coordinates are given by Eqs. (3.497) through (3.504) for the brackets not involving C_{1a} and C_{1b}. Then, following the strategy discussed in the last paragraph, we simply say that the bracket of C_{1a} or C_{1b} with *any* of the other coordinates is zero. We now have dynamics in a ten dimensional phase space, but we are interested in what is going on only in the eight dimensional subspace defined by $C_{1a} = C_{1b} = 0$. We have simply imbedded the guiding-center dynamics in a higher dimensional phase space. It is clear that the Poisson bracket still obeys antisymmetry and the Jacobi identity.

It is now straightforward to write the Poisson bracket relations among the new set of coordinates, (R', U', μ', Θ'). Once we are finished doing this, it will be alright to set C_{1a} and C_{1b} equal to zero, but not until we have taken *every* derivative that needs to be taken in the process; derivatives get messed up by coordinate transformations that are not diffeomorphisms.

We illustrate this calculation for the $\{R', U'\}$ bracket as follows:

$$\begin{aligned}\{R', U'\} &= \{R, K\hat{\mathbf{t}}\} + \{R, C_{1a}\hat{\mathbf{c}}/\lambda_B(R)\} + \{R, C_{1b}\hat{\mathbf{a}}\} \\ &= \{R, K\hat{\mathbf{t}}\} + \{R, C_{1a}\}\hat{\mathbf{c}}/\lambda_B(R) + \{R, C_{1b}\}\hat{\mathbf{a}} \\ &\quad + \{R, \hat{\mathbf{c}}/\lambda_B(R)\}C_{1a} + \{R, \hat{\mathbf{a}}\}C_{1b} \\ &= \{R, K\}\hat{\mathbf{t}} + K\{R, R\}\cdot\vec{\nabla}\,\hat{\mathbf{t}} + K\{R, \mathcal{B}\}\hat{\mathbf{b}}. \end{aligned} \tag{3.564}$$

Note that all quantities on the right hand side in the above equation are expressed in the old coordinate system. Note also that all terms involving C_{1a} or C_{1b} have vanished, either because they are bracketed with something (recall that they are

Casimir functions), or because they appear in a term outside of all derivatives and so we have set them to zero.

Eqs. (3.497) through (3.499) can now be substituted into the right hand side of Eq. (3.564). The result will still contain objects such as $\vec{\nabla}\,\hat{\mathbf{t}}$ and $\vec{\nabla}\,\hat{\mathbf{b}}$. Eliminate these by means of the easily verified relations

$$\vec{\nabla}\,\hat{\mathbf{t}} = (\vec{\nabla}\,P_{\|})\cdot\hat{\mathbf{t}} - \mathcal{Q}\hat{\mathbf{b}} \tag{3.565}$$

$$\vec{\nabla}\,\hat{\mathbf{b}} = (\vec{\nabla}\,P_{\|})\cdot\hat{\mathbf{b}} - \mathcal{Q}\hat{\mathbf{t}}. \tag{3.566}$$

Because our results are guaranteed to be boostgauge invariant, all terms involving $\mathcal{Q}$ will cancel, leaving a *manifestly* boostgauge invariant result. This being the case, the result can be expressed in terms of the new coordinates.

Before presenting these results, a word of warning is in order. When the term $K(\vec{\nabla}\,P_{\|})\cdot\hat{\mathbf{t}}$ is expressed in the new coordinates, the result is easily found to be

$$K(\vec{\nabla}\,P_{\|})\cdot\hat{\mathbf{t}} = \left(\vec{\nabla}'\,P_{\|}(R')\right)\cdot P_{\|}(R')\cdot U'. \tag{3.567}$$

Upon applying the constraint, $P_{\|}(R')\cdot U'$ can be replaced by simply U'. One might thus be tempted to pull the following dubious maneuver:

$$\left(\vec{\nabla}'\,P_{\|}(R')\right)\cdot P_{\|}(R')\cdot U' = \left(\vec{\nabla}'\,P_{\|}(R')\right)\cdot U' = \vec{\nabla}'\,\left(P_{\|}(R')\cdot U'\right) = \vec{\nabla}'\,U' = 0. \tag{3.568}$$

This is incorrect because *after the constraint is applied, R' and U' are no longer independent variables.* We thus had no right to pull U' inside the $\vec{\nabla}'$ operator, nor did we have a right to say that $\vec{\nabla}'\,U' = 0$. This is subtle but important, as the brackets below are full of things that look like $\left(\vec{\nabla}'\,P_{\|}(R')\right)\cdot U'$, and they are definitely *not* zero.

We now present the full set of brackets in the new coordinate system (omitting the primes since ambiguity should no longer result from doing so). We find

$$\{R, R\} = -\frac{\epsilon F_0}{m\lambda_B\Omega_B\Upsilon'}, \tag{3.569}$$

$$\{R,U\} = \frac{1}{m}P_\parallel + \frac{\epsilon}{m\lambda_B^2\Upsilon'}F_0\cdot\left[F'''\cdot P_\parallel - \frac{mc}{e}(\vec{\nabla}\,P_\parallel\cdot U)\right], \tag{3.570}$$

$$\{R,\Theta\} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\{R,R\}\cdot\mathcal{R} & \text{(higher order)}, \end{cases} \tag{3.571}$$

$$\begin{aligned}\{U,U\} = &-\frac{\Omega_B}{2m\lambda_B^3}\mathcal{F}_0 : (F'''\cdot\Xi'^T)\mathcal{F}_0 \\ &-\frac{1}{m}\left[\left(P_\parallel\cdot\Xi'\cdot\left(\vec{\nabla}\,P_\parallel\cdot U\right)\right) - \left(P_\parallel\cdot\Xi'\cdot\left(\vec{\nabla}\,P_\parallel\cdot U\right)\right)^T\right] \\ &-\frac{\epsilon}{m\lambda_B\Omega_B\Upsilon'}(\vec{\nabla}\,P_\parallel\cdot U)^T\cdot F_0\cdot\left(\vec{\nabla}\,P_\parallel\cdot U\right)\end{aligned} \tag{3.572}$$

$$\{U,\Theta\} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\{U,R\}\cdot\mathcal{R} & \text{(higher order)}, \end{cases} \tag{3.573}$$

and

$$\{\Theta,\mu\} = \epsilon^{-1}, \tag{3.574}$$

where we have defined

$$\Upsilon' \equiv 1 + \frac{\epsilon F_0 : F'''}{2\lambda_B^2}, \tag{3.575}$$

$$\Xi' \equiv \mathbf{1} + \frac{\epsilon F'''\cdot F_0}{\lambda_B^2\Upsilon'}, \tag{3.576}$$

$$F''' \equiv F' + \frac{mc}{e}\left((\vec{\nabla}\,P_\parallel\cdot U) - (\vec{\nabla}\,P_\parallel\cdot U\,)^T\right), \tag{3.577}$$

where F' was defined in Eq. (3.495), and where the superscripted T means "transpose." Note that Υ', Ξ' and F''' are the boostgauge invariant portions of Υ, Ξ and F''; that is, they are related by

$$\Upsilon = \Upsilon', \tag{3.578}$$

$$\Xi = \Xi' + \frac{\epsilon k}{\lambda_B\Omega_B\Upsilon'}\hat{\mathbf{b}}\mathcal{Q}\cdot F_0, \tag{3.579}$$

and

$$F'' = F''' + \frac{mcK}{e}(\hat{\mathbf{b}}\mathcal{Q} - \mathcal{Q}\hat{\mathbf{b}}). \tag{3.580}$$

These new brackets may be compared to those for the old coordinates, given in Eqs. (3.497) to (3.504).

This Poisson structure has the Casimir function, $P_\perp \cdot U$, so the constraint Eq. (3.561) is guaranteed to hold for all times if it holds initially. The physical motion takes place on the hypersurface for which this Casimir function has the value zero.

The guiding-center Hamiltonian, Eq. (3.488), can now be expressed in the new boostgauge invariant coordinates:

$$H_{gc}(R,U,\mu) = \mu\Omega_B + \frac{m}{2}U^2 + \frac{\epsilon\mu}{2\lambda_B} \times \left[\left(\frac{e}{mc}\right) F_0 : F_1 + P_\perp : \left((\vec{\nabla}\, P_\perp \cdot U)\cdot F_0\right)\right]. \quad (3.581)$$

Note that this Hamiltonian is also gyrogauge invariant, since it does not involve Θ.

There is another way to derive the above manifestly boostgauge invariant Poisson brackets. We can write the phase space Lagrangian corresponding to Eq. (3.489) in manifestly boostgauge invariant form as follows:

$$L_{gc}(R,U,\mu,\Theta,\dot{R},\dot{\Theta}) = \left[\frac{e}{\epsilon c}A + mU - \epsilon\mu\mathcal{R} + O(\epsilon^2)\right]\cdot\dot{R} + \epsilon\mu\dot{\Theta} - \lambda_{1a}U\cdot\hat{\mathbf{c}}(R,\Theta) - \lambda_{1b}U\cdot\hat{\mathbf{a}}(R,\Theta) - H_{gc}(R,U,\mu). \quad (3.582)$$

The action associated with this Lagrangian may be varied to yield the same equations of motion given by the manifestly boostgauge invariant brackets and Hamiltonian, but the variation of the action must be performed subject to the constraint, Eq. (3.561). Hence we have introduced the Lagrange multipliers, λ_{1a} and λ_{1b}. Note that varying an action subject to a constraint causes the constraint to appear as a Casimir of the resulting Poisson structure; recall the example of this phenomenon given in Subsection 2.3.6.

The equations of motion in this coordinate system are then easily found either by using the Poisson brackets given in Eqs. (3.569) through (3.574) with the

Hamiltonian given in Eq. (3.581), or by finding the Euler-Lagrange equations from the phase space Lagrangian given in Eq. (3.582). The results are

$$\dot{R} = U - \frac{\epsilon\mu F_0 \cdot \vec{\nabla}\,\Omega_B}{m\lambda_B\Omega_B\Upsilon'} + \frac{\epsilon}{\lambda_B^2\Upsilon'}F_0 \cdot \left(F''' \cdot P_\parallel - \frac{mc}{e}(\vec{\nabla}\,P_\parallel \cdot U)\right) \cdot U \tag{3.583}$$

$$\begin{aligned}\dot{U} = &-\frac{\mu}{m}P_\parallel \cdot \vec{\nabla}\,\Omega_B + \frac{\epsilon\mu}{m\lambda_B^2\Upsilon'}\left(F''' \cdot P_\parallel - \frac{mc}{e}(\vec{\nabla}\,P_\parallel \cdot U)\right)^T \cdot F_0 \cdot \vec{\nabla}\,\Omega_B \\ &- \frac{\Omega_B}{2\lambda_B^3}\mathcal{F}_0 : (F''' \cdot \Xi'^T)\mathcal{F}_0 \cdot U \\ &- \left[P_\parallel \cdot \Xi' \cdot (\vec{\nabla}\,P_\parallel \cdot U) - (\vec{\nabla}\,P_\parallel \cdot U\,)^T \cdot \Xi'^T \cdot P_\parallel\right] \cdot U \\ &- \frac{\epsilon}{\lambda_B\Omega_B\Upsilon'}\left[(\vec{\nabla}\,P_\parallel \cdot U)^T \cdot F_0 \cdot (\vec{\nabla}\,P_\parallel \cdot U\,)\right] \cdot U\end{aligned} \tag{3.584}$$

$$\dot{\mu} = 0 \tag{3.585}$$

$$\dot{\Theta} = \frac{\Omega_B}{\epsilon} + \epsilon\mathcal{R} \cdot \dot{R} + \frac{\epsilon}{2\lambda_B}\left[\left(\frac{e}{mc}\right)F_0 : F_1 + P_\perp : \left((\vec{\nabla}\,P_\perp \cdot U) \cdot F_0\right)\right]. \tag{3.586}$$

These equations of motion may be compared term for term with Eqs. (3.508) through (3.511). In the equation for $\dot{R}$, note that the parallel motion is given simply by U. The second term contains the grad-B drift, and the third term contains the curvature and polarization drifts. The first term of $\dot{U}$ contains the mirroring force, and the force due to the parallel electric field arises from the terms that contain F_1 (via their dependence on F'''). Of course, $\dot{\mu}$ still vanishes, and the equation for $\dot{\Theta}$ compares term for term with Eq. (3.511) in an obvious way.

3.14.2 Manifest Boostgauge and Gyrogauge Invariance

Now we can use the same techniques to make our results gyrogauge invariant as well. To do this, we would like to replace the coordinate Θ by the new coordinate

$$\hat{\alpha} \equiv \hat{\mathbf{a}}(R, \Theta). \tag{3.587}$$

The inverse transformation would then be

$$\Theta = \arctan\left(-\frac{\hat{\alpha}\cdot\hat{\mathbf{e}}_3(R)}{\hat{\alpha}\cdot\hat{\mathbf{e}}_2(R)}\right). \tag{3.588}$$

Note that $\hat{\alpha}$, like Θ, has only one degree of freedom, even though it is a four vector. This is because it is subject to the constraints

$$P_{\|}(R)\cdot\hat{\alpha} = 0, \tag{3.589}$$

and

$$\hat{\alpha}\cdot\hat{\alpha} = 1. \tag{3.590}$$

In order to deal with this in a proper fashion, we have to use the same techniques that we used above to get boostgauge invariant brackets. Write the coordinate transformation

$$\begin{aligned}
R' &= R \\
U' &= K\hat{\mathbf{t}}(R,\mathcal{B}) + \frac{1}{\sqrt{C_3}}\left(\frac{C_{1a}}{\lambda_B(R)}\hat{\mathbf{c}}(R,\Theta) + C_{1b}\hat{\mathbf{a}}(R,\Theta)\right) \\
\mu' &= \mu \\
\hat{\alpha}' &= \sqrt{C_3}\hat{\mathbf{a}}(R,\Theta) + \frac{1}{K}\left(-C_{2a}\hat{\mathbf{t}}(R,\mathcal{B}) + \frac{C_{2b}}{\lambda_B(R)}\hat{\mathbf{b}}(R,\mathcal{B})\right)
\end{aligned} \tag{3.591}$$

The inverse transformation is then

$$\begin{aligned}
R &= R' \\
K &= \sqrt{-U'\cdot P_{\|}(R')\cdot U'} \\
\mathcal{B} &= \tanh^{-1}\left(-\frac{U'\cdot\hat{\mathbf{e}}_1(R')}{U'\cdot\hat{\mathbf{e}}_0(R')}\right) \\
\mu &= \mu' \\
\Theta &= \arctan\left(-\frac{\hat{\alpha}'\cdot\hat{\mathbf{e}}_3(R')}{\hat{\alpha}'\cdot\hat{\mathbf{e}}_2(R')}\right) \\
C_{1a} &= U'\cdot F_0(R')\cdot\hat{\alpha}' \\
C_{1b} &= U'\cdot P_{\perp}(R')\cdot\hat{\alpha}' \\
C_{2a} &= U'\cdot P_{\|}(R')\cdot\hat{\alpha}'
\end{aligned}$$

$$C_{2b} = U' \cdot \mathcal{F}(R') \cdot \hat{\alpha}'$$
$$C_3 = \hat{\alpha}' \cdot P_\perp(R') \cdot \hat{\alpha}' \tag{3.592}$$

We demand that C_{1a}, C_{1b}, C_{2a}, C_{2b} and C_3 are Casimir functions, and that the physical motion takes place on the submanifold defined by $C_{1a} = C_{1b} = C_{2a} = C_{2b} = 0$ and $C_3 = 1$.

We can now write the Poisson bracket relations among the new coordinates. We use the easily verified relations

$$\vec{\nabla}\,\hat{\mathbf{t}} = (\vec{\nabla}\,P_\parallel)\cdot\hat{\mathbf{t}} + \mathcal{Q}\hat{\mathbf{b}} \tag{3.593}$$

$$\vec{\nabla}\,\hat{\mathbf{b}} = (\vec{\nabla}\,P_\parallel)\cdot\hat{\mathbf{b}} - \mathcal{Q}\hat{\mathbf{t}} \tag{3.594}$$

$$\vec{\nabla}\,\hat{\mathbf{c}} = (\vec{\nabla}\,P_\perp)\cdot\hat{\mathbf{c}} + \mathcal{R}\hat{\mathbf{a}} \tag{3.595}$$

$$\vec{\nabla}\,\hat{\mathbf{a}} = (\vec{\nabla}\,P_\perp)\cdot\hat{\mathbf{a}} - \mathcal{R}\hat{\mathbf{c}}. \tag{3.596}$$

Note that, because our results are guaranteed to be both boostgauge and gyrogauge invariant, all terms involving $\mathcal{Q}$ and $\mathcal{R}$ will cancel, leaving a *manifestly* boostgauge and gyrogauge invariant result. Also note that the Hamiltonian $H_{gc}(R, U, \mu)$, given by Eq. (3.581), is already manifestly gyrogauge invariant (this is because it is Θ-independent). The new manifestly boostgauge and gyrogauge invariant brackets are then

$$\{R, R\} = -\frac{\epsilon F_0}{m\lambda_B\Omega_B\Upsilon'}, \tag{3.597}$$

$$\{R, U\} = \frac{1}{m}P_\parallel + \frac{\epsilon}{m\lambda_B^2\Upsilon'}F_0\cdot\left[F'''\cdot P_\parallel - \frac{mc}{e}(\vec{\nabla}\,P_\parallel\cdot U)\right], \tag{3.598}$$

$$\begin{aligned}\{U, U\} = &-\frac{\Omega_B}{2m\lambda_B^3}\mathcal{F}_0 : (F'''\cdot\Xi'^T)\mathcal{F}_0 \\ &-\frac{1}{m}\left(P_\parallel\cdot\Xi'\cdot(\vec{\nabla}\,P_\parallel\cdot U) - (\vec{\nabla}\,P_\parallel\cdot U\,)^T\cdot\Xi'^T\cdot P_\parallel\right) \\ &-\frac{\epsilon}{m\lambda_B\Omega_B\Upsilon'}(\vec{\nabla}\,P_\parallel\cdot U)^T\cdot F_0\cdot(\vec{\nabla}\,P_\parallel\cdot U\,)\end{aligned} \tag{3.599}$$

$$\{R,\hat{\alpha}\} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\{R,R\}\cdot\vec{\nabla}\, P_\perp\cdot\hat{\alpha} & \text{(higher order),} \end{cases} \tag{3.600}$$

$$\{U,\hat{\alpha}\} = \begin{cases} 0 & \text{(classical order)} \\ \epsilon\{U,R\}\cdot\vec{\nabla}\, P_\perp\cdot\hat{\alpha} & \text{(higher order),} \end{cases} \tag{3.601}$$

$$\{\hat{\alpha},\mu\} = \frac{1}{\epsilon\lambda_B} F_0\cdot\hat{\alpha}, \tag{3.602}$$

and

$$\{\hat{\alpha},\hat{\alpha}\} = -\frac{\epsilon}{m\lambda_B\Omega_B\Upsilon}(\vec{\nabla}\, P_\perp\cdot\hat{\alpha})^T\cdot F_0\cdot(\vec{\nabla}\, P_\perp\cdot\hat{\alpha}). \tag{3.603}$$

This Poisson structure has the Casimir functions, $P_\perp\cdot U$, $P_\parallel\cdot\hat{\alpha}$, and $\hat{\alpha}\cdot\hat{\alpha}$. This insures that the constraint Eqs. (3.561), (3.589) and (3.590) will hold at all times if they hold initially. The physical motion takes place on the hypersurface for which the first two of these Casimir functions have the value zero and the third has the value one.

Note that $\hat{\alpha}$, like Θ, has nonvanishing brackets with R and U at higher order. Once again, however, the set of functions of R and U form a subset of the set of all possible phase functions that is closed under the operation of these Poisson brackets; also, H_{gc} is independent of $\hat{\alpha}$. So we can still reduce to the guiding-center description.

Next, we note that these results could have been derived by varying the action corresponding to the phase space Lagrangian obtained by rewriting Eq. (3.489) in manifestly boostgauge and gyrogauge invariant format,

$$\begin{aligned} L_{gc} = \Big[\frac{e}{\epsilon c}A + mU + O(\epsilon^2)\Big]\cdot\dot{R} - \frac{\epsilon\mu}{\lambda_B}\hat{\alpha}\cdot F_0\cdot\dot{\hat{\alpha}} \\ - \lambda_{1a}U\cdot F_0\cdot\hat{\alpha} - \lambda_{1b}U\cdot P_\perp\cdot\hat{\alpha} \\ - \lambda_{2a}U\cdot P_\parallel\cdot\hat{\alpha} - \lambda_{2b}U\cdot\mathcal{F}_0\cdot\hat{\alpha} \\ - \lambda_3\hat{\alpha}\cdot P_\perp\cdot\hat{\alpha} - H_{gc}(R,U,\mu). \end{aligned} \tag{3.604}$$

This must be varied subject to the constraints, Eqs. (3.561), (3.589) and (3.590). We have enforced these constraints by introducing the scalar Lagrange multi-

pliers, λ_{1a}, λ_{1b}, λ_{2a}, λ_{2b}, and λ_3. Note that the term involving $\mathcal{R}$ has disappeared from Γ_{gc} when written in these coordinates, because $-\mu\mathcal{R}\cdot\dot{R}+\mu\dot{\Theta}=-\mu\hat{\alpha}\cdot F_0\cdot\dot{\hat{\alpha}}/\lambda_B$.

We are going to need these Lagrange multipliers in Chapter 5, so we compute them here for reference. They are rather easy to calculate, especially since we already know the Poisson brackets. The Euler-Lagrange equations for coordinates U and $\hat{\alpha}$ are

$$0=m\dot{R}-\lambda_{1a}F_0\cdot\hat{\alpha}-\lambda_{1b}P_\perp\cdot\hat{\alpha}-\frac{\partial H_{gc}}{\partial U} \tag{3.605}$$

and

$$\frac{d}{d\tau}\left(\frac{\epsilon\mu}{\lambda_B}\hat{\alpha}\cdot F_0\right)=\frac{\epsilon\mu}{\lambda_B}F_0\cdot\dot{\hat{\alpha}}+\lambda_{2a}U\cdot P_\parallel+\lambda_{2b}U\cdot\mathcal{F}_0+2\lambda_3P_\perp\cdot\hat{\alpha}+\frac{\partial H_{gc}}{\partial\hat{\alpha}}, \tag{3.606}$$

respectively. Upon multiplication by $\hat{\alpha}\cdot F_0$ and $\hat{\alpha}$, the first of these yields

$$\lambda_{1a}=-\frac{1}{\lambda_B^2}\hat{\alpha}\cdot F_0\cdot\left(m\dot{R}-\frac{\partial H_{gc}}{\partial U}\right) \tag{3.607}$$

and

$$\lambda_{1b}=\hat{\alpha}\cdot\left(m\dot{R}-\frac{\partial H_{gc}}{\partial U}\right), \tag{3.608}$$

respectively. Upon multiplication by U, $U\cdot\mathcal{F}_0$ and $\hat{\alpha}$, the second yields

$$\lambda_{2a}=\frac{1}{U^2}U\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\dot{R}-\frac{\partial H_{gc}}{\partial\hat{\alpha}}\right] \tag{3.609}$$

and

$$\lambda_{2b}=\frac{-1}{\lambda_B^2U^2}U\cdot\mathcal{F}_0\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\dot{R}-\frac{\partial H_{gc}}{\partial\hat{\alpha}}\right] \tag{3.610}$$

and

$$\lambda_3=\frac{\epsilon\mu}{\lambda_B}\dot{\hat{\alpha}}\cdot F_0\cdot\hat{\alpha}+\frac{1}{2}\hat{\alpha}\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\dot{R}-\frac{\partial H_{gc}}{\partial\hat{\alpha}}\right], \tag{3.611}$$

respectively. Note that, in perfect analogy with Eq. (2.184), these results can be cast in the form

$$\lambda_\nu=\xi_\nu^\alpha\frac{\partial H_{gc}}{\partial Z^\alpha}, \tag{3.612}$$

where the label ν runs over all the constraints present (1a, 1b, 2a, 2b, 3), and where

$$\xi^R_{1a} = -\frac{m}{\lambda_B^2}\hat{\alpha}\cdot F_0\cdot\{R,R\}$$
$$\xi^U_{1a} = -\frac{m}{\lambda_B^2}\hat{\alpha}\cdot F_0\cdot\left(\{R,U\}-\frac{1}{m}\mathbf{1}\right)$$
$$\xi^\mu_{1a} = 0$$
$$\xi^{\hat{\alpha}}_{1a} = -\frac{m}{\lambda_B^2}\hat{\alpha}\cdot F_0\cdot\{R,\hat{\alpha}\}$$

and

$$\xi^R_{1b} = m\hat{\alpha}\cdot\{R,R\}$$
$$\xi^U_{1b} = m\hat{\alpha}\cdot\left(\{R,U\}-\frac{1}{m}\mathbf{1}\right)$$
$$\xi^\mu_{1b} = 0$$
$$\xi^{\hat{\alpha}}_{1b} = m\hat{\alpha}\cdot\{R,\hat{\alpha}\}$$

and

$$\xi^R_{2a} = \frac{1}{U^2}U\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,R\}\right]$$
$$\xi^U_{2a} = \frac{1}{U^2}U\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,U\}\right]$$
$$\xi^\mu_{2a} = 0$$
$$\xi^{\hat{\alpha}}_{2a} = \frac{1}{U^2}U\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,\hat{\alpha}\}-\mathbf{1}\right]$$

and

$$\xi^R_{2b} = \frac{-1}{U^2}U\cdot\mathcal{F}_0\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,R\}\right]$$
$$\xi^U_{2b} = \frac{-1}{U^2}U\cdot\mathcal{F}_0\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,U\}\right]$$
$$\xi^\mu_{2b} = 0$$
$$\xi^{\hat{\alpha}}_{2b} = \frac{-1}{\lambda_B^2U^2}U\cdot\mathcal{F}_0\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,\hat{\alpha}\}-\mathbf{1}\right]$$

and

$$\xi^R_3 = -\frac{\epsilon\mu}{\lambda_B}\hat{\alpha}\cdot F_0\cdot\{\hat{\alpha},R\}+\frac{1}{2}\hat{\alpha}\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,R\}\right]$$

$$\xi_3^U = -\frac{\epsilon\mu}{\lambda_B}\hat{\alpha}\cdot F_0\cdot\{\hat{\alpha},U\} + \frac{1}{2}\hat{\alpha}\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,U\}\right]$$
$$\xi_3^\mu = -\frac{\epsilon\mu}{\lambda_B}\hat{\alpha}\cdot F_0\cdot\{\hat{\alpha},\mu\}$$
$$\xi_3^{\hat{\alpha}} = -\frac{\epsilon\mu}{\lambda_B}\hat{\alpha}\cdot F_0\cdot\{\hat{\alpha},\hat{\alpha}\} + \frac{1}{2}\hat{\alpha}\cdot\left[\epsilon\mu\hat{\alpha}\cdot\left(\frac{F_0}{\lambda_B}\right)\overleftarrow{\nabla}\cdot\{R,\hat{\alpha}\} - \mathbf{1}\right]. \quad (3.613)$$

Finally, we note that the equations of motion in these coordinates are easily found either by using the Poisson brackets given in Eqs. (3.597) through (3.603) with the Hamiltonian given in Eq. (3.581), or by finding the Euler-Lagrange equations from the phase space Lagrangian given in Eq. (3.604). The results are

$$\dot{R} = U - \frac{\epsilon\mu F_0\cdot\vec{\nabla}\,\Omega_B}{m\lambda_B\Omega_B\Upsilon'} + \frac{\epsilon}{\lambda_B^2\Upsilon'}F_0\cdot\left(F'''\cdot P_\parallel - \frac{mc}{e}(\vec{\nabla}\,P_\parallel\cdot U)\right)\cdot U \quad (3.614)$$

$$\dot{U} = -\frac{\mu}{m}P_\parallel\cdot\vec{\nabla}\,\Omega_B + \frac{\epsilon\mu}{m\lambda_B^2\Upsilon'}\left(F'''\cdot P_\parallel - \frac{mc}{e}(\vec{\nabla}\,P_\parallel\cdot U)\right)^T\cdot F_0\cdot\vec{\nabla}\,\Omega_B$$
$$- \frac{\Omega_B}{2\lambda_B^3}\mathcal{F}_0 : (F'''\cdot\Xi'^T)\mathcal{F}_0\cdot U$$
$$- \left[P_\parallel\cdot\Xi'\cdot(\vec{\nabla}\,P_\parallel\cdot U) - (\vec{\nabla}\,P_\parallel\cdot U\,)^T\cdot\Xi'^T\cdot P_\parallel\right]\cdot U$$
$$- \frac{\epsilon}{\lambda_B\Omega_B\Upsilon'}\left[(\vec{\nabla}\,P_\parallel\cdot U)^T\cdot F_0\cdot(\vec{\nabla}\,P_\parallel\cdot U\,)\right]\cdot U \quad (3.615)$$

$$\dot{\mu} = 0 \quad (3.616)$$

$$\dot{\hat{\alpha}} = \frac{e}{\epsilon mc}F_0\cdot\hat{\alpha} + \hat{\alpha}\cdot(P_\perp\overleftarrow{\nabla})\cdot\dot{R} - \frac{\epsilon}{2\lambda_B^2}\hat{\alpha}\cdot F_0\cdot\left[\left(\frac{e}{mc}\right)F_0 : F_1 + P_\perp : \left((\vec{\nabla}\,P_\perp\cdot U)\cdot F_0\right)\right]. \quad (3.617)$$

Note that Eqs. (3.614) through (3.616) are identical to the corresponding equations in the last subsection. These were gyrogauge invariant anyway, and so were unaffected by the manipulations carried out in this subsection. The equation for $\dot{\Theta}$ has been replaced by an equation for $\dot{\hat{\alpha}}$; the two may, however, be compared term for term in an obvious way.

Chapter 4

Relativistic Oscillation-Center Theory

4.1 Discussion

In this chapter, we shall consider the perturbation of a guiding center due to the presence of an electromagnetic wave of eikonal form. In doing so, we shall take as our unperturbed problem the guiding-center equations of motion, as derived in Chapter 3. Thus we are effectively using the superconvergent Lie transform procedure as described in Subsection 2.4.1.

We are interested in understanding the response of the guiding center to the presence of the wave. Towards this end, we seek a transformation to a new system of coordinates in which the wave perturbation is removed. Neglecting resonant phenomena, it turns out that it is possible to do this to first order, but not to second order. At second order, there remains an averaged residual perturbation to the Hamiltonian that gives rise to the ponderomotive force exerted by the wave on the guiding center. Thus, after we transform away the rapid fluctuations in

the guiding-center motion, we are left with the slower ponderomotive effects.

An analogy with the guiding-center problem may be helpful here. In that calculation, we averaged over the rapid gyromotion to find the slower drift motion. The thing that is drifting is then called a "guiding center." A guiding center is a fictitious object whose position and momentum are the gyroaverage of the particle position and momentum, respectively. Furthermore, a guiding center may be thought of as having an intrinsic or *spin* angular momentum equal to the *orbital* angular momentum of the underlying gyrating particle. Thus, by finding the averaging transformation that eliminates the fast degree of freedom, we have discovered a new "macroparticle" that lives on the slow time scale, but whose properties derive from those of the original charged particle gyrating on the fast time scale.

Similarly, when a perturbing wave is present and we transform away the associated rapid fluctuations, the residual ponderomotive forces may be thought of as acting on a new "macroparticle" that is averaged over a wave oscillation time scale. We call this new object an "oscillation center." Whereas an individual charged particle feels wave fluctuations on a rapid time scale, an oscillation center feels only the slower ponderomotive effects; it also feels resonant effects (since these are also slow and do not average away), but we shall ignore these in our treatment. Thus, a kinetic equation for a plasma of oscillation centers would contain only ponderomotive forces and resonant effects.

The averaged nth-order part of the ponderomotive Hamiltonian is called K_n, and we shall derive this for a relativistic guiding center. As has already been noted, K_1 vanishes if we neglect resonant effects. It was discovered by Cary and Kaufman that there exists an intimate connection between the ponderomotive Hamiltonian and the plasma's response to a wave. Specifically, K_2 is a quadratic form in the amplitude of the perturbing wave, and the kernel of this quadratic form is the functional derivative of the linear susceptibility with respect to the

distribution function. Subsequently, it was found by Kaufman that this relationship persists to higher order; that is, nonlinear corrections to the susceptibility are related to K_3, etc.

In the traditional approach to studying plasma response to a wave, one begins with the field equations and the kinetic equation, and studies perturbations in the fields and the distribution function about an equilibrium. Though this approach is not as systematic as ours, it has at least one advantage: The vector potential never appears, so all results obtained by such an analysis are guaranteed to be *manifestly* gauge invariant. In contrast, Hamiltonian or Lagrangian approaches to ponderomotive theory seem to require the use of the vector potential, so past attempts along these lines have produced results whose gauge invariance was either not established, or established only by laborious calculation after the fact.

In this chapter, we shall find that eikonal wave perturbations to the Lagrangian action for a relativistic charged particle in the guiding-center representation can be written in *manifestly* gauge-invariant form. To do this, it is necessary to abandon the usual approach of expanding the eikonal wave perturbation in a series of Bessel functions of $k_\perp \rho$. Instead, we first perform a Lagrangian gauge transformation, and then we expand in a series of functions that are related to indefinite integrals of Bessel functions. This allows us to develop an oscillation-center theory to arbitrarily high order in the wave amplitude expansion parameter, and be guaranteed of *manifest* gauge invariance at every step of the way. Thus, we can enjoy the benefits of the systematic Lie transform approach to ponderomotive theory without fear of losing manifest gauge invariance.

4.2 Eikonal Wave Perturbation

In single-particle phase space coordinates, an eikonal wave has a four potential of the form

$$A_w(r) = \tilde{A}(r) \exp\left(\frac{i}{\epsilon}\psi(r)\right) + \text{c.c.}, \tag{4.618}$$

where $\tilde{A}$ is the amplitude and ψ is the phase, and where c.c. denotes the expression's complex conjugate. The derivative of ψ with respect to spacetime position is the four wavevector, $\mathbf{k}$:

$$\mathbf{k} = \vec{\nabla}\,\psi(r). \tag{4.619}$$

Both $\tilde{A}$ and $\mathbf{k}$ are slowly varying functions of r. That is, an eikonal wave is *locally* a plane wave. To reflect this, we have placed $1/\epsilon$ in front of the phase. Thus, the derivative of A_w with respect to r is $i\mathbf{k}A_w/\epsilon$ plus terms of order unity that involve derivatives of $\tilde{A}$ or of $\mathbf{k}$.

Furthermore, in this work, we shall take this eikonal expansion parameter to be equal to the guiding-center expansion parameter (hence, it is no coincidence that we are calling it ϵ). This means that we are considering waves whose characteristic wavelengths are on the order of a gyroradius, and whose characteristic frequencies are on the order of a gyrofrequency.

We shall now consider the effect of such a wave on the single particle action one form in Eq. (3.370). Replacing A in that equation by $A + \lambda A_w$, we write

$$\gamma = \gamma'' + \lambda\gamma_w, \tag{4.620}$$

where γ_w is the perturbation in the action one form due to the wave, or

$$\gamma_w = \frac{e}{c}\tilde{A}(r) \cdot dr \exp\left(\frac{i}{\epsilon}\psi(r)\right) + \text{c.c.} \tag{4.621}$$

Note that we have introduced a new expansion parameter, λ, to order the wave amplitude. For the time being, we shall not compare λ and ϵ, though more will be said about this later.

As was remarked earlier, our starting point for the oscillation center Lie transform will be the guiding-center equations of motion. Hence, it is necessary to write γ_w in guiding-center coordinates (the above form for γ_w is in particle coordinates). We apply the guiding-center Lie transform to the above equation for γ to get

$$\Gamma = \Gamma_{gc} + \lambda\Gamma_w, \tag{4.622}$$

where Γ_{gc} is the guiding-center action one form, calculated in Chapter 3. Then, Γ_w is given by

$$\Gamma_w = \exp(-\epsilon\mathcal{L}_g)\gamma_w, \tag{4.623}$$

where g is the generator for the guiding-center transformation.

Note that we are working only to first order in ϵ. To this order we can take $g^r = -\rho\hat{\mathbf{a}}$, where $\rho \equiv w/\Omega_B$. All other components of g are unnecessary, and may be ignored. We shall use the boostgauge invariant set of coordinates (R, U, μ, θ) described in Section 3.14.

4.3 Manifest Gauge Invariance

At this point in the calculation, the usual approach is to apply the Lie transform in Eq. (4.623) by simply substituting $R + \rho\hat{\mathbf{a}}$ for r in Eq. (4.621). This is straightforward, and the result is

$$\Gamma_w = \frac{e}{c}\left(\tilde{A}\cdot dR + \frac{\epsilon\tilde{A}\cdot\hat{\mathbf{a}}d\mu}{m\rho\Omega_B} + \epsilon\rho\tilde{A}\cdot\hat{\mathbf{c}}d\theta\right)\exp\left(\frac{i}{\epsilon}\psi\right)\exp(i\rho\hat{\mathbf{a}}\cdot\mathbf{k}) + \text{c.c.}, \tag{4.624}$$

where we have retained the leading nonvanishing order for each component of the one form, and where it is understood that all quantities on the right (such as $\tilde{A}$ and $\hat{\mathbf{c}}$) are now evaluated at R. Since $\hat{\mathbf{a}}\cdot\mathbf{k}$ is oscillatory, the second exponential in the above expression gives rise to a series of Bessel functions of $\mathrm{k}_\perp\rho$.

Unfortunately, the above expression for Γ_w does *not* possess manifest gauge invariance. To understand why this is, we must qualify what we mean by "man-

ifest gauge invariance." A term in the action one form is gauge invariant if it is unchanged to within a Lagrangian gauge transformation when $\tilde{A}$ is replaced by $\tilde{A} + i\mathbf{k}\Lambda$, where Λ is any slowly varying scalar function of position. Thus, the quantity

$$\tilde{F} \equiv i(\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k}) \tag{4.625}$$

is gauge invariant since it is unchanged by this transformation. The quantity $\tilde{A} \cdot dR \exp(i\psi/\epsilon)$ is also gauge invariant since it transforms to itself plus the term

$$i\Lambda\mathbf{k} \cdot dR \exp\left(\frac{i}{\epsilon}\psi\right) = d\left[\epsilon\Lambda \exp\left(\frac{i}{\epsilon}\psi\right)\right] \tag{4.626}$$

(where we have neglected higher-order terms in ϵ), and this can be removed by a Lagrangian gauge transformation. We shall say that a term is *manifestly* gauge invariant if it has the form $\tilde{A} \cdot dR \exp(i\psi/\epsilon)$, or if it depends on $\tilde{A}$ only through its dependence on $\tilde{F}$.

Thus the first term on the right hand side of Eq. (4.624) is manifestly gauge invariant, but the other two terms are not. They are gauge invariant (as they must be), since to leading order in ϵ we have

$$i\left(\frac{\epsilon\Lambda\mathbf{k} \cdot \hat{\mathbf{a}} d\mu}{m\rho\Omega_B} + \epsilon\rho\Lambda\mathbf{k} \cdot \hat{\mathbf{c}} d\theta\right) \exp\left(\frac{i}{\epsilon}\psi\right) \exp(i\rho\hat{\mathbf{a}} \cdot \mathbf{k})$$
$$= d\left[\epsilon^2\rho\Lambda\mathbf{k} \cdot \hat{\mathbf{a}} \exp\left(\frac{i}{\epsilon}\psi\right) \exp(i\rho\hat{\mathbf{a}} \cdot \mathbf{k})\right] \tag{4.627}$$

and this can be removed by a Lagrangian gauge transformation, but they are not *manifestly* gauge invariant.

If we were to use Eq. (4.624) as the starting point for our ponderomotive theory, we would obtain results for K_n that are not manifestly gauge invariant. We could get around this problem if there were some way of manipulating Eq. (4.624) into manifestly gauge-invariant form. It turns out that this can be done by making a particular Lagrangian gauge transformation, but this transformation is far from obvious and needs to be motivated. As we shall now see, this motivation comes from the homotopy formula.

Return to Eq. (4.623), and expand the exponential in a series of Lie derivatives

$$\Gamma_w = \sum_{j=0}^{\infty} \frac{(-\epsilon)^j}{j!} \mathcal{L}_g^j \gamma_w. \tag{4.628}$$

Applying the generalized homotopy formula, Eq. (2.76), we get

$$\Gamma_w = \gamma_w + \sum_{j=1}^{\infty} \frac{(-\epsilon)^j}{j!} \left[(i_g d)^j + (d i_g)^j \right] \gamma_w. \tag{4.629}$$

Note that we have split off the $j = 0$ term from the sum because Eq. (2.76) is valid only for $j \geq 1$. The above may now be written in the suggestive form

$$\Gamma_w = \left[\gamma_w + \sum_{j=1}^{\infty} \frac{(-\epsilon)^j}{j!} (i_g d)^j \gamma_w \right] + d \left[\sum_{j=1}^{\infty} \frac{(-\epsilon)^j}{j!} i_g (d i_g)^{j-1} \gamma_w \right]. \tag{4.630}$$

Note that the second term in square brackets is an exact one form, and may therefore be removed by a Lagrangian gauge transformation. The first term in square brackets has two pieces: The first is γ_w itself, which we know is manifestly gauge invariant. The second is a series of terms all of which have the operator $i_g d$, raised to some power, operating on γ_w. Thus, in all these terms, the very first operator to be applied to γ_w is the exterior derivative. Now

$$d\gamma_w = \frac{e}{2\epsilon c} \tilde{F}\colon \; dr \wedge dr \exp\left(\frac{i}{\epsilon} \psi(r) \right) + \text{c.c.} \tag{4.631}$$

(plus higher-order terms), and this is manifestly gauge invariant. Subsequent applications of i_g and d preserve this manifest gauge invariance. Thus the term in the first square brackets on the right hand side of Eq. (4.630) is manifestly gauge invariant. Thus, Eq. (4.630) gives us the Lagrangian gauge transformation that leaves Γ_w in *manifestly* gauge invariant form.

At this point, one may wonder why we have bothered to keep all the terms in the above series when we have said that we are interested in only the lowest nonvanishing order in ϵ. Note that when we apply differential operators to γ_w, as given by Eq. (4.621), we pull out factors of $1/\epsilon$. This means that even terms with

very high j can make order unity contributions. Thus, it is important to keep *all* the terms of the series as given above. This situation arises as a consequence of the nonanalyticity of γ_w in ϵ. It will become more clear momentarily.

To proceed, we need expressions for $(i_g d)^j \gamma_w$ and $i_g (d i_g)^j \gamma_w$, for $j \geq 1$. To get such expressions, we simply evaluate them for the first few values of j, notice the pattern, and prove it by mathematical induction. The results are

$$\left[(i_g d)^j \gamma_w\right]_r = -\frac{ie}{c}\left(\frac{i}{\epsilon}\right)^j (g^r \cdot \mathbf{k})^{j-1} g^r \cdot \tilde{F} \exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}\left(\frac{1}{\epsilon^{j-1}}\right) + \text{c.c.}$$

$$\left[(i_g d)^j \gamma_w\right]_\mu = \frac{-e}{\epsilon c}(j-1)\left(\frac{i g^r \cdot \mathbf{k}}{\epsilon}\right)^{j-2} \frac{\partial g^r}{\partial \mu} \cdot \tilde{F} \cdot g^r \exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}\left(\frac{1}{\epsilon^{j-2}}\right) + \text{c.c.}$$

$$\left[(i_g d)^j \gamma_w\right]_\theta = \frac{-e}{\epsilon c}(j-1)\left(\frac{i g^r \cdot \mathbf{k}}{\epsilon}\right)^{j-2} \frac{\partial g^r}{\partial \theta} \cdot \tilde{F} \cdot g^r \exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}\left(\frac{1}{\epsilon^{j-2}}\right) + \text{c.c.}$$

and

$$i_g (d i_g)^j \gamma_w = \frac{e}{c}\left(\frac{i}{\epsilon}\right)^j (g^r \cdot \mathbf{k})^j g^r \cdot \tilde{A} \exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}\left(\frac{1}{\epsilon^{j-1}}\right) + \text{c.c.} \qquad (4.632)$$

Note that the components of $(i_g d)^j \gamma_w$ are manifestly gauge invariant, as promised. Then $i_g (d i_g)^j \gamma_w$ is not manifestly gauge invariant, but this is the term that will be removed by the Lagrangian gauge transformation. Thus, everything is going as planned.

Now we must plug the above results into Eq. (4.630), and sum the series over j. This is straightforward, and the result is

$$\begin{aligned}
\Gamma_w = \gamma_w &+ \frac{e}{c}\left[-i g^r \cdot \tilde{F}\left(\frac{\exp(-i g^r \cdot \mathbf{k}) - 1}{g^r \cdot \mathbf{k}}\right)\exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}(\epsilon)\right] \cdot dR \\
&+ \frac{e}{c}\left[-\epsilon \frac{\partial g^r}{\partial \mu} \cdot \tilde{F} \cdot g^r \left(\frac{(1 + i g^r \cdot \mathbf{k})\exp(-i g^r \cdot \mathbf{k}) - 1}{(g^r \cdot \mathbf{k})^2}\right)\exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}(\epsilon^2)\right] d\mu \\
&+ \frac{e}{c}\left[-\epsilon \frac{\partial g^r}{\partial \theta} \cdot \tilde{F} \cdot g^r \left(\frac{(1 + i g^r \cdot \mathbf{k})\exp(-i g^r \cdot \mathbf{k}) - 1}{(g^r \cdot \mathbf{k})^2}\right)\exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}(\epsilon^2)\right] d\theta \\
&- d\left[\frac{i\epsilon e}{c} g^r \cdot \tilde{A}\left(\frac{\exp(-i g^r \cdot \mathbf{k}) - 1}{g^r \cdot \mathbf{k}}\right)\exp\left(\frac{i}{\epsilon}\psi\right) + \mathcal{O}(\epsilon^2)\right] + \text{c.c.} \qquad (4.633)
\end{aligned}$$

At this point, we can check the above result by actually applying the exterior derivative to the last term in square brackets. There is extensive cancellation,

and we are left with Eq. (4.624), as expected. We can now make the Lagrangian gauge transformation,

$$\Gamma'_w \equiv \Gamma_w + dS_T \tag{4.634}$$

where

$$S_T \equiv \frac{i\epsilon e}{c} g^r \cdot \tilde{A} \left(\frac{\exp(-ig^r \cdot \mathbf{k}) - 1}{g^r \cdot \mathbf{k}} \right) \exp\left(\frac{i}{\epsilon}\psi\right) + \text{c.c.}, \tag{4.635}$$

thereby removing the last term of Eq. (4.633) to get a *manifestly* gauge invariant one form, as desired.

Now $g^r = -\rho\hat{\mathbf{a}}$, and we can substitute this into Eq. (4.633). Note that the μ component of Γ'_w vanishes because g^r and $\partial g^r/\partial\mu$ are both in the $\hat{\mathbf{a}}$ direction, and they are both dotted into the antisymmetric two form, $\tilde{F}$. The θ component does not vanish, however, because $\partial g^r/\partial\theta$ is in the $\hat{\mathbf{c}}$ direction. We finally have

$$\begin{aligned}\Gamma'_w = \frac{e}{c}\Bigg[\tilde{A} + \rho\hat{\mathbf{a}}\cdot\tilde{F}\left(\frac{\exp(i\rho\hat{\mathbf{a}}\cdot\mathbf{k})-1}{i\rho\hat{\mathbf{a}}\cdot\mathbf{k}}\right) + \mathcal{O}(\epsilon)\Bigg]\cdot dR\exp\left(\frac{i}{\epsilon}\psi\right) \\ - \left[\frac{\epsilon e\rho^2}{2c\lambda_B}F_0 : \tilde{F}\left(\frac{(1-i\rho\hat{\mathbf{a}}\cdot\mathbf{k})\exp(i\rho\hat{\mathbf{a}}\cdot\mathbf{k})-1}{(\rho\hat{\mathbf{a}}\cdot\mathbf{k})^2}\right) + \mathcal{O}(\epsilon^2)\right] d\theta\exp\left(\frac{i}{\epsilon}\psi\right) \\ + \mathcal{O}(\epsilon^2)d\mu + \mathcal{O}(\epsilon^2)\cdot dU + \text{c.c.}\end{aligned} \tag{4.636}$$

To proceed, we must Fourier analyze the above expression in preparation for the oscillation-center Lie transformation.

4.4 Fourier Expansion in Gyroangle

We now write the components of $\mathbf{k}$ in the $\hat{\mathbf{e}}_\mu$ basis, introduced back in Chapter 3, as follows:

$$\mathbf{k} = \mathrm{k}_\parallel - \mathrm{k}_\perp(\hat{\mathbf{e}}_2\sin\alpha + \hat{\mathbf{e}}_3\cos\alpha), \tag{4.637}$$

where $\mathbf{k}_\parallel$ lies entirely within the parallel two-flat. The geometrical situation is illustrated schematically in Fig. 4.1. Then, using Eq. (3.346), we find

$$\hat{\mathbf{a}}\cdot\mathbf{k} = \mathrm{k}_\perp\sin(\theta-\alpha). \tag{4.638}$$

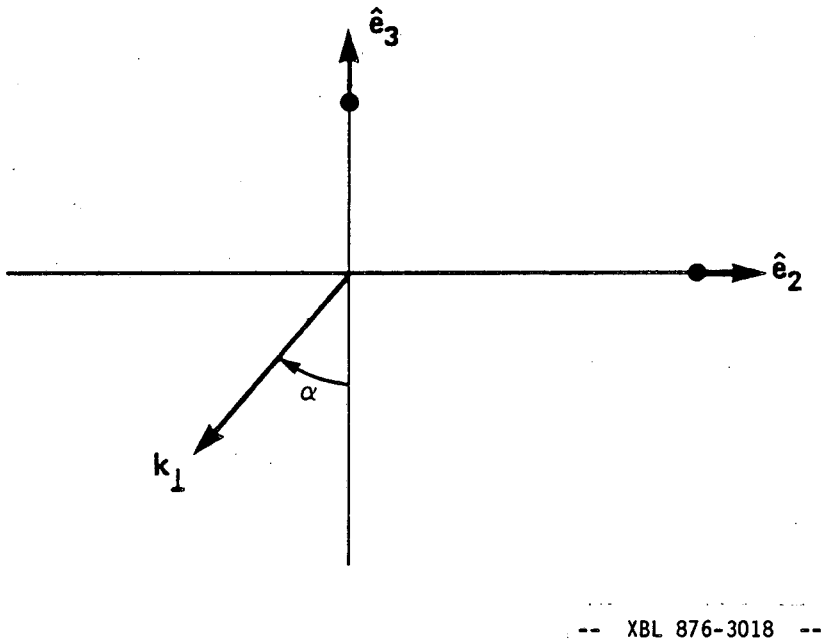


Figure 4.1: Components of the Four Wavevector

Now we may Fourier Expand the quantities

$$\begin{aligned}\exp(i\rho\hat{\mathbf{a}}\cdot\mathbf{k}) &= e^{i\mathrm{k}_\perp\rho\sin(\theta-\alpha)} \\ &= \sum_\ell J_\ell(\mathrm{k}_\perp\rho)e^{i\ell(\theta-\alpha)},\end{aligned} \tag{4.639}$$

$$\begin{aligned}\frac{\exp(i\rho\hat{\mathbf{a}}\cdot\mathbf{k})-1}{i\rho\hat{\mathbf{a}}\cdot\mathbf{k}} &= \frac{e^{i\mathrm{k}_\perp\rho\sin(\theta-\alpha)}-1}{i\mathrm{k}_\perp\rho\sin(\theta-\alpha)} \\ &= \sum_\ell Q_\ell(\mathrm{k}_\perp\rho)e^{i\ell(\theta-\alpha)}\end{aligned} \tag{4.640}$$

and

$$\begin{aligned}\frac{(1-i\rho\hat{\mathbf{a}}\cdot\mathbf{k})\exp(i\rho\hat{\mathbf{a}}\cdot\mathbf{k})-1}{(\rho\hat{\mathbf{a}}\cdot\mathbf{k})^2} &= \frac{(1-i\mathrm{k}_\perp\rho\sin(\theta-\alpha))\,e^{i\mathrm{k}_\perp\rho\sin(\theta-\alpha)}-1}{\mathrm{k}_\perp^2\rho^2\sin^2(\theta-\alpha)} \\ &= \frac{1}{2}\sum_\ell R_\ell(\mathrm{k}_\perp\rho)e^{i\ell(\theta-\alpha)},\end{aligned} \tag{4.641}$$

where the J_ℓ are Bessel functions,

$$J_\ell(x) \equiv \frac{1}{2\pi}\int_0^{2\pi} d\xi e^{ix\sin\xi-i\ell\xi}, \tag{4.642}$$

where we have defined the special functions

$$Q_\ell(x) \equiv \frac{1}{2\pi}\int_0^{2\pi} d\xi\left(\frac{e^{ix\sin\xi}-1}{ix\sin\xi}\right)e^{-i\ell\xi} \tag{4.643}$$

and

$$R_\ell(x) \equiv \frac{1}{\pi}\int_0^{2\pi} d\xi\left(\frac{(1-ix\sin\xi)e^{ix\sin\xi}-1}{x^2\sin^2\xi}\right)e^{-i\ell\xi}, \tag{4.644}$$

and where the summations over ℓ extend from minus infinity to infinity. The properties of the Q and R functions will be explored in detail in Appendix D.

Now, along with the expressions in Eqs. (4.639), (4.640) and (4.641), Γ_w and Γ'_w also contains the θ-dependent (and hence oscillatory) quantities, $\hat{\mathbf{c}}$ and $\hat{\mathbf{a}}$. Thus we need to know how to Fourier expand these as well. Using Eq. (3.346), we may write

$$\hat{\mathbf{c}} = \frac{i}{\sqrt{2}}(\hat{\mathbf{e}}_+e^{i\theta}-\hat{\mathbf{e}}_-e^{-i\theta}) \tag{4.645}$$

and

$$\hat{\mathbf{a}} = \frac{1}{\sqrt{2}}(\hat{\mathbf{e}}_+ e^{i\theta} + \hat{\mathbf{e}}_- e^{-i\theta}), \tag{4.646}$$

where we have defined

$$\hat{\mathbf{e}}_\pm \equiv \frac{1}{\sqrt{2}}(\hat{\mathbf{e}}_2 \pm i\hat{\mathbf{e}}_3). \tag{4.647}$$

Note that these are complex unit vectors that obey $\hat{\mathbf{e}}_\pm^* = \hat{\mathbf{e}}_\mp$, $\hat{\mathbf{e}}_\pm \cdot \hat{\mathbf{e}}_\pm = 0$, and $\hat{\mathbf{e}}_\pm^* \cdot \hat{\mathbf{e}}_\pm = 1$. Because they contain $e^{\pm i\theta}$, when we multiply them by the series in Eqs. (4.639), (4.640) and (4.641), they will generate terms with $e^{i(\ell\pm1)\theta}$. By defining new summation variables we can restore these to the form $e^{i\ell\theta}$, but then these terms will be left with special functions that have indices $\ell \pm 1$.

Now then, we may write Γ_w as follows:

$$\Gamma_w = \sum_\ell (\Gamma_{\ell R} \cdot dR + \epsilon\Gamma_{\ell\mu} d\mu + \epsilon\Gamma_{\ell\theta} d\theta) \exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \mathcal{O}(\epsilon^2) + \text{c.c.}, \tag{4.648}$$

where

$$\Gamma_{\ell R} \equiv \frac{e}{c} J_\ell \tilde{A}, \tag{4.649}$$

$$\Gamma_{\ell\mu} \equiv \frac{1}{\sqrt{2}\rho\lambda_B} \mathcal{J}_\ell^+ \cdot \tilde{A} \tag{4.650}$$

and

$$\Gamma_{\ell\theta} \equiv \frac{ie\rho}{\sqrt{2}c} \mathcal{J}_\ell^- \cdot \tilde{A}, \tag{4.651}$$

and where we have defined

$$\Psi_\ell(R,\theta) \equiv \psi(R) + \epsilon\ell(\theta - \alpha(R)) \tag{4.652}$$

and

$$\mathcal{J}_\ell^\pm \equiv \hat{\mathbf{e}}_+ e^{i\alpha} J_{\ell-1} \pm \hat{\mathbf{e}}_- e^{-i\alpha} J_{\ell+1}. \tag{4.653}$$

Similarly, we may write Γ'_w as follows:

$$\Gamma'_w = \sum_\ell (\Gamma'_{\ell R} \cdot dR + \epsilon\Gamma'_{\ell\theta} d\theta) \exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \mathcal{O}(\epsilon^2) + \text{c.c.}, \tag{4.654}$$

where

$$\Gamma'_{\ell R} \equiv \frac{e}{c}\left[\delta_{\ell 0}\tilde{A} + \frac{\rho}{\sqrt{2}}\mathcal{Q}_\ell^+ \cdot \tilde{F} + \mathcal{O}(\epsilon)\right], \tag{4.655}$$

and

$$\Gamma'_{\ell\theta} \equiv -\frac{e\rho^2}{4c\lambda_B} R_\ell F_0 : \tilde{F} + \mathcal{O}(\epsilon^2), \tag{4.656}$$

and where we have defined

$$\mathcal{Q}_\ell^\pm \equiv \hat{\mathbf{e}}_+ e^{i\alpha} Q_{\ell-1} \pm \hat{\mathbf{e}}_- e^{-i\alpha} Q_{\ell+1}. \tag{4.657}$$

In the above expressions, it is understood that J_ℓ, Q_ℓ and R_ℓ are evaluated at $\mathrm{k}_\perp\rho$.

Finally, note that S_T, as defined by Eq. (4.635), has the Fourier decomposition,

$$S_T = \epsilon \sum_\ell S_{T\ell} \exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.} \tag{4.658}$$

where

$$S_{T\ell} \equiv -\frac{e\rho}{\sqrt{2}c}\mathcal{Q}_\ell^+ \cdot \tilde{A}. \tag{4.659}$$

Using Eqs. (4.648), (4.654) and (4.658), it is possible to check that $\Gamma'_w = \Gamma_w + dS_T$.

4.5 The Oscillation-Center Lie Transform

Our aim is to perform a Lie transform that will remove all the effects of the wave from the Poisson structure, and put them into the Hamiltonian. Thus, when we have completed this task, our Poisson brackets will be identical to those for a guiding center with no wave present (through order λ^2). The effect of the wave will be pushed into a term of order λ^2 in the Hamiltonian. We shall do this both for Γ_w and for Γ'_w, in order to verify that we get the same answer either way.

We now reset our variables, so that Γ_w (as given by Eq. (4.648)) and Γ'_w (as given by Eq. (4.654)) will henceforth be called γ_w and γ'_w, respectively. The oscillation-center transform will take us to Γ_w and Γ'_w, but we want these to

vanish by the above argument. Thus, in Eqs. (2.273) and (2.274) we demand that Γ_1 and Γ_2 vanish. This is the step at which we are neglecting resonant effects. Furthermore, in Eq. (2.274) we have $\gamma_2 = 0$ because our wave perturbation is at first order in λ only, and $\Omega_1 = 0$ because $\Gamma_1 = 0$.

First consider the oscillation-center transform of γ_w. We have

$$0 = \gamma_w - i_1\omega_{gc} + dS_1, \tag{4.660}$$

and

$$0 = -i_2\omega_{gc} - \frac{1}{2}i_1\omega_w + dS_2. \tag{4.661}$$

Meanwhile, the Hamiltonian transforms according to Eqs. (2.221) through (2.224) to give

$$K_0 = H_{gc}, \tag{4.662}$$

$$K_1 = -\mathcal{L}_1 H_{gc} = -i_1 dH_{gc}, \tag{4.663}$$

and

$$K_2 = -\mathcal{L}_2 H_{gc} + \frac{1}{2}\mathcal{L}_1^2 H_{gc}. \tag{4.664}$$

Now we demand that $K_1 = -i_1 dH_{gc} = 0$. Let i_0 denote interior multiplication by $\dot{z}$ (the unperturbed flow), so $i_0\omega_{gc} = -dH_{gc}$ (our unperturbed problem is the guiding-center problem). Then, applying i_0 to Eq. (4.660) gives

$$\dot{S}_1 = -i_0\gamma_w + i_1 dH_{gc} = -i_0\gamma_w, \tag{4.665}$$

where the last step follows as a result of our demand that $K_1 = 0$. We can integrate this last equation along unperturbed orbits to get S_1. Then g_1 is given by Eq. (4.660)

$$g_1 = (\gamma_w + dS_1) \cdot J_{gc}. \tag{4.666}$$

At second order, we can solve Eq. (4.661) for g_2 as follows:

$$g_2 = (-\frac{1}{2}i_1\omega_w + dS_2) \cdot J_{gc}. \tag{4.667}$$

Now then, we can insert these generators into Eq. (4.664) to get K_2, as follows:

$$\begin{aligned} K_2 &= -\mathcal{L}_2 H_{gc} + \frac{1}{2}\mathcal{L}_1^2 H_{gc} \\ &= -\mathcal{L}_2 H_{gc} \\ &= (\frac{1}{2} i_1 \omega_w - dS_2) \cdot J_{gc} \cdot dH_{gc} \\ &= \frac{1}{2} i_0 i_1 \omega_w - \dot{S}_2. \end{aligned} \tag{4.668}$$

Now we can choose S_2 to remove the oscillatory part of the first term. Note that we cannot remove the averaged part of the first term, because that would introduce secular terms in S_2. So the best that we can do is to take

$$K_2 = \left\langle \frac{1}{2} i_0 i_1 \omega_w \right\rangle. \tag{4.669}$$

This is the ponderomotive Hamiltonian.

Now suppose that we had started with $\gamma_w' = \gamma_w + dS_T$ instead of γ_w. Instead of Eqs. (4.660) and (4.661), we would have written

$$0 = \gamma_w' - i_{1'} \omega_{gc} + dS_1', \tag{4.670}$$

and

$$0 = -i_{2'} \omega_{gc} - \frac{1}{2} i_{1'} \omega_w + dS_2', \tag{4.671}$$

where $i_{n'}$ is an obvious shorthand for $i_{g_n'}$, and where we are adhering to the convention of using primes to denote quantities arising from the Lagrangian gauge transformed action one form. Of course, we still would have taken $K_0' = K_0 = H_{gc}$ and we still would have demanded that $K_1' = -i_{1'} dH_{gc} = 0 = K_1$. From this it follows that $K_2' = -i_{2'} dH_{gc}$. Thus, if we could show that $g_2 = g_2'$, it would immediately follow that $K_2' = K_2$; that is, it would follow that the ponderomotive Hamiltonian is invariant under the Lagrangian gauge transformation.

From Eq. (4.670), we have

$$\dot{S}_1' = i_0 dS_1' = -i_0 \gamma_w' + i_{1'} dH_{gc} = -i_o \gamma_w' = -i_0(\gamma_w + dS_T) = \dot{S}_1 - \dot{S}_T, \tag{4.672}$$

so

$$S_1' = S_1 - S_T. \tag{4.673}$$

Then

$$g_1' = (\gamma_w' + dS_1') \cdot J_{gc} = (\gamma_w + dS_T + dS_1 - dS_T) \cdot J_{gc} = (\gamma_w + dS_1) \cdot J_{gc} = g_1, \tag{4.674}$$

So g_1 is invariant under the Lagrangian gauge transformation. Next, from Eq. (4.671) we have

$$g_2' = \left(-\frac{1}{2} i_1 \omega_w + dS_2' \right) \cdot J_{gc}, \tag{4.675}$$

so

$$K_2' = \left(\frac{1}{2} i_1 \omega_w - dS_2' \right) \cdot J_{gc} \cdot dH_{gc} = \frac{1}{2} i_0 i_1 \omega_w - \dot{S}_2' = \frac{1}{2} i_0 i_1 \omega_w - \dot{S}_2'. \tag{4.676}$$

Thus we have

$$\dot{S}_2' = \frac{1}{2} i_0 i_1 \omega_w - \left\langle \frac{1}{2} i_0 i_1 \omega_w \right\rangle = \dot{S}_2, \tag{4.677}$$

so

$$S_2' = S_2, \tag{4.678}$$

and so

$$g_2' = g_2. \tag{4.679}$$

It immediately follows that

$$K_2' = \left\langle \frac{1}{2} i_0 i_1 \omega_w \right\rangle = K_2, \tag{4.680}$$

so the ponderomotive Hamiltonian is indeed invariant under the Lagrangian gauge transformation. Note that g_1, g_2, and S_2 are also thus invariant, but that γ_w and S_1 are *not*. The latter two quantities transform under the Lagrangian gauge transformation as follows:

$$\gamma_w' = \gamma_w + dS_T \tag{4.681}$$

and

$$S_1' = S_1 - S_T, \tag{4.682}$$

so that the combination $\gamma_w + dS_1$ *is* invariant.

Though we have just shown that we would get the same answer for the ponderomotive Hamiltonian either way, it bears repeating that the advantage of starting with γ_1' is its *manifest* gauge invariance. In the next section, we shall further discuss the relative merits of each of the two ways of calculating K_2.

While the above expression, Eq. (4.669), for the ponderomotive Hamiltonian is wonderfully compact, it is also very formal. We need to plug in Eq. (4.648) and/or Eq. (4.654), and work it out in detail. This is done in the next section.

4.6 The Ponderomotive Hamiltonian

Our unperturbed equations of motion are

$$\dot{R} = U + \epsilon U_d$$

$$\dot{U} = \mathcal{O}(1)$$

$$\dot{\mu} = 0$$

and

$$\dot{\theta} = \frac{1}{\epsilon}\Omega_B, \tag{4.683}$$

where U_d denotes the guiding-center drift motion, and where we do not need to know anything about $\dot{U}$ other than the fact that it is order unity in ϵ. Then Eq. (4.665) for S_1 becomes

$$\dot{S}_1 = -\sum_\ell \left[\gamma_{\ell R}\cdot(U+\epsilon U_d) + \gamma_{\ell\theta}\Omega_B\right]\exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.}, \tag{4.684}$$

and Eq. (4.672) for S_1' becomes

$$\dot{S}_1' = -\sum_\ell \left[\gamma_{\ell R}'\cdot(U+\epsilon U_d) + \gamma_{\ell\theta}'\Omega_B\right]\exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.} \tag{4.685}$$

Integrating over unperturbed orbits, we get

$$S_1 = i\epsilon \sum_\ell \left[\gamma_{\ell R} \cdot (U + \epsilon U_d) + \gamma_{\ell\theta}\Omega_B\right] \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.}, \tag{4.686}$$

and

$$S_1' = i\epsilon \sum_\ell \left[\gamma_{\ell R}' \cdot (U + \epsilon U_d) + \gamma_{\ell\theta}'\Omega_B\right] \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.}, \tag{4.687}$$

respectively. Here we have defined the resonant denominator

$$D_\ell \equiv \dot{\Psi}_\ell = \mathbf{k} \cdot (U + \epsilon U_d) + \ell\Omega_B. \tag{4.688}$$

Using Eqs.(4.658), (4.686) and (4.687), it is possible to verify Eq. (4.673); that is, it is possible to show explicitly that $S_1' = S_1 - S_T$.

Now we use Eq. (4.666) to get the components of the generator g_1,

$$g_1^R = -\frac{\epsilon}{m} \sum_\ell \left(H_{1\ell}\mathbf{k} + \frac{e}{c} J_\ell D_\ell \tilde{A}\right) \cdot \left(\frac{F_0 D_\ell}{\lambda_B \Omega_B} + iP_\parallel\right) \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell^2}$$
$$+ \text{c.c.} + \mathcal{O}(\epsilon^2), \tag{4.689}$$

$$g_1^U = \frac{1}{m} \sum_\ell \left(H_{1\ell}\mathbf{k} + \frac{e}{c} J_\ell D_\ell \tilde{A}\right) \cdot P_\parallel \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.} + \mathcal{O}(\epsilon), \tag{4.690}$$

$$g_1^\mu = -\frac{1}{\Omega_B} \sum_\ell \left(H_{1\ell}\mathbf{k} + \frac{e}{c} J_\ell D_\ell \tilde{A}\right) \cdot U \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.} + \mathcal{O}(\epsilon), \tag{4.691}$$

and

$$g_1^\theta = \sum_\ell \Big[-\frac{e}{mc\lambda_B}\mathbf{k} \cdot F_0 \cdot \tilde{A} J_\ell + \frac{i\mathbf{k}_\perp}{2\rho\lambda_B}(J_{\ell+1} - J_{\ell-1}) U \cdot \tilde{A}$$
$$- \frac{1}{\sqrt{2}\rho\lambda_B} \tilde{A} \cdot \mathcal{J}_\ell^+ \mathbf{k} \cdot U\Big] \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.} + \mathcal{O}(\epsilon), \tag{4.692}$$

where we have defined

$$H_{1\ell} \equiv -\frac{e}{c}\left(U J_\ell + \frac{i\rho\Omega_B}{\sqrt{2}} \mathcal{J}_\ell^-\right) \cdot \tilde{A}. \tag{4.693}$$

If we had instead used the the first of Eqs. (4.674), we would have obtained the following results for the components of g_1':

$$g_1'^R = \frac{i\epsilon e}{mc} \sum_\ell \left(U J_\ell + \frac{i\rho\Omega_B}{\sqrt{2}} \mathcal{J}_\ell^-\right) \cdot \tilde{F}$$

$$\cdot \left(\frac{F_0 D_\ell}{\lambda_B \Omega_B} + i P_\parallel \right) \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell^2} + \text{c.c.} + \mathcal{O}(\epsilon^2), \tag{4.694}$$

$$g_1'^U = -\frac{ie}{mc} \sum_\ell \left(U J_\ell + \frac{i\rho\Omega_B}{\sqrt{2}} \mathcal{J}_\ell^- \right) \cdot \tilde{F} \cdot P_\parallel \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.} + \mathcal{O}(\epsilon), \tag{4.695}$$

$$g_1'^\mu = \frac{e\rho}{\sqrt{2}c} \tilde{F} \colon \sum_\ell \left(U \mathcal{J}_\ell^- \right) \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.} + \mathcal{O}(\epsilon), \tag{4.696}$$

and

$$g_1'^\theta = \frac{ie}{2mc\lambda_B} \tilde{F} : \sum_\ell \left[F_0 \cdot \left(\mathbf{1} J_\ell + \frac{\sqrt{2}i}{\rho\Omega_B} \mathcal{J}_\ell^- U \right) \right] \frac{\exp\left(\frac{i}{\epsilon}\Psi_\ell\right)}{D_\ell} + \text{c.c.} + \mathcal{O}(\epsilon). \tag{4.697}$$

By straightforward calculation, it is possible to directly verify that $g_1' = g_1$, as required by Eq. (4.674). To do this, simply substitute $\tilde{F} \equiv i(\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k})$ into Eqs. (4.694) through (4.697); upon simplification, the results will be Eqs. (4.689) through (4.692). If we had not made the Lagrangian gauge transformation, and had instead started with only γ_w and S_1, we might have had difficulty casting Eqs. (4.689) through (4.692) in the manifestly gauge-invariant form of Eqs. (4.694) through (4.697).

Next we compute the components of $\omega_w = d\gamma_w$. Direct calculation gives

$$\omega_w = \sum_\ell \omega_{w\ell} \exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.}, \tag{4.698}$$

where

$$\omega_{w\ell RR} = \frac{ie}{\epsilon c} J_\ell (\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k}), \tag{4.699}$$

$$\omega_{w\ell RU} = \mathcal{O}(\epsilon^2), \tag{4.700}$$

$$\omega_{w\ell R\mu} = \frac{i}{\sqrt{2}\rho\lambda_B} \left[\mathbf{k}\mathcal{J}_\ell^+ - \frac{i\mathbf{k}_\perp}{2} \left(J_{\ell+1} - J_{\ell-1} \right) \mathbf{1} \right] \cdot \tilde{A}, \tag{4.701}$$

$$\omega_{w\ell R\theta} = -\frac{e\rho}{\sqrt{2}c} \left(\mathbf{k}\mathcal{J}_\ell^- + \frac{\sqrt{2}i\ell}{\rho} J_\ell \mathbf{1} \right) \cdot \tilde{A}, \tag{4.702}$$

$$\omega_{w\ell UU} = \mathcal{O}(\epsilon^2), \tag{4.703}$$

$$\omega_{w\ell U\mu} = \mathcal{O}(\epsilon^2), \tag{4.704}$$

$$\omega_{w\ell U\theta} = \mathcal{O}(\epsilon^2), \tag{4.705}$$

and

$$\omega_{w\ell\mu\theta} = -\frac{i\epsilon}{\lambda_B^2}\mathbf{k}\cdot F_0\cdot\tilde{A}J_\ell. \tag{4.706}$$

If we had instead used $\omega'_w = d\gamma'_w$, we would have obtained the following results:

$$\omega'_w = \sum_\ell \omega'_{w\ell}\exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.}, \tag{4.707}$$

where

$$\omega'_{w\ell RR} = \frac{e}{\epsilon c}J_\ell\tilde{F}, \tag{4.708}$$

$$\omega'_{w\ell RU} = \mathcal{O}(\epsilon^2), \tag{4.709}$$

$$\omega'_{w\ell R\mu} = \frac{1}{\sqrt{2}\rho\lambda_B}\tilde{F}\cdot\mathcal{J}_\ell^+, \tag{4.710}$$

$$\omega'_{w\ell R\theta} = \frac{ie\rho}{\sqrt{2}c}\tilde{F}\cdot\mathcal{J}_\ell^-, \tag{4.711}$$

$$\omega'_{w\ell UU} = \mathcal{O}(\epsilon^2), \tag{4.712}$$

$$\omega'_{w\ell U\mu} = \mathcal{O}(\epsilon^2), \tag{4.713}$$

$$\omega'_{w\ell U\theta} = \mathcal{O}(\epsilon^2), \tag{4.714}$$

and

$$\omega'_{w\ell\mu\theta} = -\frac{\epsilon}{2\lambda_B^2}F_0 : \tilde{F}J_\ell. \tag{4.715}$$

By direct calculation, it is once again possible to verify that $\omega'_w = d\gamma'_w = d(\gamma_w + dS_T) = d\gamma_w = \omega_w$ by simply substituting $\tilde{F} \equiv i(\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k})$ into the results for the components of ω'_w and simplifying to get the components of ω_w.

Before using the above results to calculate K_2, we digress for one last discussion about the relative merits of starting with γ_w and γ'_w. First note that all of the components of g_1 and ω_w are indeed manifestly gauge invariant. If we had

started the calculation with γ'_w, this would not be a surprise since γ'_w is itself manifestly gauge invariant; if however we had started the calculation with γ_w, the manifest gauge invariance of the result would seem fortuitous. In the latter event, we would have had results in terms of $\tilde{A}$, and only through some tedious algebraic manipulations would we have discovered that their dependence on $\tilde{A}$ arose only through a dependence on $\tilde{F}$. On the other hand, note that the only special functions that appear in the components of g_1 and ω_w are the Bessel functions, J_ℓ. The Q_ℓ and R_ℓ functions have all disappeared in favor of the J_ℓ. If we had started the calculation with γ_w, this would not be a surprise since γ_w itself depends only on the J_ℓ, and not on the Q_ℓ and R_ℓ; if however we had started the calculation with γ'_w, the disappearance of the Q_ℓ and R_ℓ functions would seem fortuitous. In the latter event, we would have had results in terms of the Q_ℓ and R_ℓ functions, and only through some tedious algebraic manipulations would we have discovered that the recursion relations and derivative formulas could be used to cast them in terms of J_ℓ alone. There is thus a peculiar duality between the presence of special functions and of manifest gauge invariance.

We now insert the above formulas into our expression for K_2. The averaging is carried out as follows:

$$\left\langle \exp\left(\frac{i}{\epsilon}\Psi_\ell\right)^* \exp\left(\frac{i}{\epsilon}\Psi_{\ell'}\right)\right\rangle = \delta_{\ell\ell'}. \tag{4.716}$$

We get

$$K_2 = \frac{e^2}{2mc^2}(\tilde{A}^* \cdot P_\parallel \cdot \tilde{A} + \frac{i}{\lambda_B\Omega_B}\mathbf{k}\cdot U\tilde{A}^* \cdot F_0 \cdot \tilde{A} + \text{c.c.}) + \sum_\ell K_{2\ell}, \tag{4.717}$$

where

$$\begin{aligned} K_{2\ell} = \frac{eH^*_{1\ell}}{mcD_\ell}\Big\{\mathbf{k}\cdot P_\parallel \cdot \tilde{A}J_\ell - \frac{i\mathbf{k}\cdot U}{\rho\Omega_B}\Big[-\frac{i\mathrm{k}_\perp U}{2\Omega_B}(J_{\ell-1} - J_{\ell+1}) \\ - \frac{\mathbf{k}\cdot U}{\sqrt{2}\Omega_B}\mathcal{J}^+_\ell + \frac{\rho}{\lambda_B}F_0\cdot\mathbf{k}J_\ell\Big]\cdot\tilde{A}\Big\} \\ + \frac{|H_{1\ell}|^2}{2mD_\ell^2}\mathbf{k}\cdot P_\parallel\cdot\mathbf{k} + \text{c.c.} + \mathcal{O}(\epsilon), \end{aligned} \tag{4.718}$$

where c.c. denotes the complex conjugate, and where $H_{1\ell}$ is defined in Eq. (4.693). If we had instead computed K_2' according to Eq. (4.676), we would have obtained the result,

$$K_2' = \sum_\ell K_{2\ell}', \tag{4.719}$$

where

$$\begin{aligned} K_{2\ell}' = \frac{ie^2}{2mc^2 D_\ell^2} \left(J_\ell U + \frac{i\rho\Omega_B}{\sqrt{2}} \mathcal{J}_\ell^- \right) \cdot \tilde{F} \cdot \left(\frac{F_0 D_\ell}{\lambda_B \Omega_B} + iP_\parallel \right) \\ \cdot \tilde{F}^* \cdot \left(J_\ell U + \frac{i\rho\Omega_B}{\sqrt{2}} \mathcal{J}_\ell^- \right)^* + \frac{e^2\rho}{2\sqrt{2}mc^2 D_\ell} \tilde{F} : \left(\mathcal{J}_\ell^- U \right) \\ \tilde{F}^* : \left[\frac{1}{\lambda_B} F_0 \cdot \left(\mathbf{1} J_\ell - \frac{\sqrt{2}i}{\rho\Omega_B} \mathcal{J}_\ell^- U \right)^* \right] + \text{c.c.} + \mathcal{O}(\epsilon). \end{aligned} \tag{4.720}$$

Once again, by substituting $\tilde{F} \equiv i(\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k})$ into Eq. (4.720) and simplifying, it is possible to reduce the expression to Eq. (4.718), thus directly verifying that $K_2' = K_2$. In the course of this calculation, some of the sum rules of Appendix E are useful. Henceforth we shall drop the prime in our notation, and refer to the ponderomotive Hamiltonian only as K_2, whether or not it is in manifestly gauge-invariant form.

Note that K_2 is a function of the phase space coordinates, R, U, μ and θ; in particular, it depends on R through its dependence on the background fields, $F_0(R)$ and $F_1(R)$, and through its dependence on the eikonal wave field parameters, $\tilde{F}(R)$ and $\mathbf{k}(R)$. Thus we write $K_2(Z; F_i(R), \tilde{F}(R), \mathbf{k}(R))$, where $i = 0, 1$.

The ponderomotive Hamiltonian will be used extensively in the next chapter where we shall study the self-consistent dynamics of magnetized relativistic plasma in an eikonal wave field.

4.7 Obtaining the Ponderomotive Hamiltonian Using Canonical Lie Transforms

Grebogi and Littlejohn [8] have obtained the ponderomotive Hamiltonian by first performing a single noncanonical coordinate transformation to remove the perturbation from the action one form, and then using *canonical* Lie transforms on the Hamiltonian. We shall use that procedure in this section in order to check our above result for K_2.

Let us return to the point at which the wave perturbation was first added to the single-particle action one form in Eq. (4.620). Recall the definition of the single-particle velocity u in Eq. (3.366). Suppose that we change this definition to absorb the wave perturbation; that is, we adopt the following new definition for u:

$$u = \frac{1}{m}\left(p - \frac{e}{c}A(q)\right) + \frac{\lambda e}{mc}\tilde{A}(q)\exp\left(\frac{i}{\epsilon}\psi\,(q)\right) + \text{c.c.} \tag{4.721}$$

This has the effect of returning the action one form to the functional form that it had before the wave was introduced. Of course, the *definitions* of the quantities that appear in the one form will be different; that is, u and anything that depends on u (e.g. k, β, w, and θ) will be defined differently in terms of the single-particle position and velocity. Nevertheless, the action one form is returned to the form that it had when no wave was present, and now we can apply the usual guiding-center transformation to take it to the guiding-center action one form Γ_w, given implicitly in Eq. (3.582), with no remaining perturbation due to the wave.

Whereas the action one form has thus been simplified by this transformation, the Hamiltonian, Eq. (3.367), now becomes considerably more complicated. Using the new definition of u in Eq. (3.364), we have

$$H'(r,u) = H'_0(r,u) + \lambda H'_1(r,u) + \lambda^2 H'_2(r,u), \tag{4.722}$$

where

$$H_0'(r,u) = \frac{m}{2}u^2 \tag{4.723}$$

$$H_1'(r,u) = -\frac{e}{c}u\cdot\tilde{A}(r)\exp\left(\frac{i}{\epsilon}\psi(r)\right) + \text{c.c.} \tag{4.724}$$

and

$$H_2'(r,u) = \frac{e^2}{2mc^2}\tilde{A}(r)\cdot\tilde{A}^*(r) + \frac{e^2}{2mc^2}\tilde{A}(r)\cdot\tilde{A}(r)\exp\left(\frac{2i}{\epsilon}\psi(r)\right) + \text{c.c.} \tag{4.725}$$

At this point we can apply the guiding-center transformation, $(r,u) \mapsto (R,U,\mu,\Theta)$, which may be taken to be simply $R = r - \epsilon\rho$ to the order to which we are working. The result may be Fourier expanded in the gyroangle using the usual Bessel function identities. The result is

$$H = H_0 + \lambda H_1 + \lambda^2 H_2, \tag{4.726}$$

where

$$H_0 = \frac{m}{2}U^2 + \mu\Omega_B \tag{4.727}$$

is the usual guiding-center Hamiltonian (to lowest order), where

$$H_1 = \sum_\ell H_{1\ell}\exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.} \tag{4.728}$$

with $H_{1\ell}$ given by Eq. (4.693), and where

$$H_2 = \frac{e^2}{mc^2}\tilde{A}\cdot\tilde{A}^* + \text{oscillatory terms.} \tag{4.729}$$

To recap, we have applied a noncanonical transformation to remove the perturbation from the Poisson structure and deposit it in the Hamiltonian. We can now use a *canonical* Lie transform to remove H_1 (neglecting resonances) and average H_2 to get K_2. Note that this method does not preserve manifest gauge invariance; that was lost in the very first step when we redefined u in a gauge-dependent way.

Applying canonical Lie transform perturbation theory, at first order we have from Eq. (2.251)

$$0 = K_1 = H_1 + \{W_1, H_0\}, \tag{4.730}$$

so

$$\{W_1, H_0\} = -H_1 = -\sum_\ell H_{1\ell} \exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.} \tag{4.731}$$

Integrate this along unperturbed orbits to get the scalar generator

$$W_1 = i\epsilon \sum_\ell \frac{H_{1\ell}}{D_\ell} \exp\left(\frac{i}{\epsilon}\Psi_\ell\right) + \text{c.c.} \tag{4.732}$$

Proceeding to second order, we have from Eq. (2.252)

$$K_2 = H_2 + \{W_2, H_0\} + \frac{1}{2}\{W_1, H_1\}. \tag{4.733}$$

Now W_2 is chosen to average the result, so without having to explicitly calculate it, we can write

$$K_2 = \left\langle H_2 + \frac{1}{2}\{W_1, H_1\} \right\rangle. \tag{4.734}$$

After a short calculation, this reduces to the result

$$\begin{aligned} K_2 &= \frac{e^2}{2mc^2}\tilde{A}\cdot\tilde{A}^* - \frac{1}{2}\sum_\ell \{\Psi_\ell, \frac{|H_{1\ell}|^2}{D_\ell}\} + \text{c.c.} \\ &= \frac{e^2}{2mc^2}\tilde{A}\cdot\tilde{A}^* - \frac{1}{2}\sum_\ell \left(\frac{1}{m}\mathbf{k}_\parallel \cdot \frac{\partial}{\partial U} + \ell\frac{\partial}{\partial \mu}\right)\frac{|H_{1\ell}|^2}{D_\ell} + \text{c.c.} \end{aligned} \tag{4.735}$$

That this answer is equal to our previous result for K_2 may be proved by expanding the derivatives in Eq. (4.735), replacing ℓ by $[D_\ell - \mathbf{k}\cdot(U + \epsilon U_d)]/\Omega_B$, and using the sum rules of Appendix E to sum the terms with no resonant denominator. The result is Eqs. (4.717) and (4.718).

Note that this is by far the easiest way to get K_2. Furthermore, it yields the result in a considerably more compact form than the Lagrangian Lie transform approach does. On the other hand, as has already been noted, it does not yield the result in manifestly gauge invariant form.

This result may be compared with that of Grebogi and Littlejohn [8] who used "1 + 3" notation and whose result was gauge invariant but not manifestly so. To make this comparison, use the technique for translating our results into "1 + 3" notation that was introduced back in Section 3.13. It is then a straightforward exercise to show that our ponderomotive Hamiltonian gives rise to the same equations of motion as that of Littlejohn and Grebogi, though the two are *not* numerically equal. The reason that the two results for K_2 are not numerically equal can be traced back to the fact that the corresponding *unperturbed* Hamiltonians are not numerically equal. This is because Littlejohn and Grebogi started with the Hamiltonian (written in terms of three-vector coordinates and velocities),

$$H_{LG} = (\gamma_v - 1)mc^2 + e\phi, \tag{4.736}$$

which is not numerically equal to the Hamiltonian that we started with, though it does yield the same equations of motion.

It is easier to compare our result with that of Achterberg [49] who used a four-vector approach, but who did not worry about manifest gauge invariance and who used essentially the same method outlined in this section. His result is identical to our Eq. (4.735), outside of some minor notational differences.

Chapter 5

The Relativistic Guiding-Center Plasma

5.1 Discussion

The reason that a Vlasov plasma is a nonlinear medium is that the plasma currents generate fields which in turn drive the motion of the plasma. Up until now in this thesis, we have dealt only with single particles (or single guiding centers or single guiding/oscillation centers) moving in fields that are known in advance as fixed functions of spacetime. In this final chapter, we show how to pass from this single particle description to a *self-consistent* description of the dynamics of the guiding-center plasma; this includes the dynamics of the fields as well as that of the particles. We shall do this by imbedding the single particle action in a system action, and coupling it to the Maxwell field.

In Section 5.2, we prove Liouville's theorem, and show how to write the Vlasov equation in any desired coordinate system. In Section 5.3, we sum the guiding-center Lagrangian action over a full distribution of guiding centers and couple to

the Maxwell field in order to obtain the Lagrangian action of the full guiding-center Vlasov plasma. The variation of this with respect to the guiding-center coordinates yields the relativistic kinetic equation for guiding centers, while the variation with respect to the four potential yields the self-consistent field equation including the guiding-center magnetization and current densities.

In Section 5.4, Noether's theorem is applied constructively to obtain covariant conservation laws for the momentum-energy and the angular momentum of a guiding-center plasma. That is, we obtain the stress-energy and angular momentum tensors of the guiding-center plasma, including the contribution to the angular momentum due to guiding-center spin.

Finally, in Section 5.5, we employ the results of Chapter 4 to generalize the results of Sections 5.3 and 5.4 to the case of a guiding-center plasma in an eikonal wave field. We begin by forming a system action, this time including the Maxwell action of the eikonal wave field, and the ponderomotive Hamiltonian of the guiding/oscillation centers. Variation with respect to the coordinates again yields the kinetic equation, which now includes a term due to the ponderomotive effects caused by the wave field. Variation with respect to the four potential of the *background* field again yields the self-consistent field equation, which now includes a modification in the magnetization density due to the presence of the wave. There are then two new additional variations: Variation with respect to the eikonal wave field amplitude yields the linear dispersion relation for the wave, and variation with respect to the eikonal wave phase yields the conservation law for wave action. Constructive application of Noether's theorem to this new system action yields the laws of conservation of energy-momentum and angular momentum for the combined system of plasma, background field, and wave field. Specifically, the modification to the stress-energy and angular momentum tensors due to the presence of the wave field is presented and discussed.

5.2 Liouville's Theorem

5.2.1 Lagrangian and Eulerian Descriptions of Relativistic Plasma

In this section, we present a version of Liouville's theorem that is valid for relativistic Hamiltonian systems with noncanonical coordinates. We begin by examining the difference between the Lagrangian and Eulerian descriptions of relativistic kinetic theory.

Recall that a Lagrangian description keeps track of the trajectory of each particle of the system, whereas an Eulerian description uses a distribution function to specify the phase-space density of particles (we discussed this briefly in Section 2.3.3). Thus, a Lagrangian description for a system of relativistic particles might be the specification of $z(\eta, \tau)$, where z denotes a set of n-dimensional phase space coordinates, η is a continuous particle label, and $\tau(\eta)$ is an orbit parameter along the world line of the particle with label η. Specifying z as a function of η and τ is equivalent to specifying the phase space orbit of every particle in the system. The corresponding Eulerian distribution is

$$f_n(Z) = \int dN(\eta) \int d\tau(\eta) \delta^n(Z - z(\eta, \tau)). \tag{5.737}$$

Here $dN(\eta)$ is some measure describing the number of particles with labels between η and $\eta + d\eta$. This measure appears when we pass from the discrete to the continuum description; that is

$$\sum_{\text{particles}} \rightarrow \int dN(\eta). \tag{5.738}$$

In what follows, we shall frequently not bother to write the explicit η dependence of τ, but it should be kept in mind that each particle has its own proper time.

Note that $f(Z)$ has support only on a space of dimension smaller than that of the full n-dimensional phase space. This is because there are constraints that

must be satisfied by the various coordinates involved. For example, single-particle dynamics must remain on the mass shell, since $u \cdot u = -c^2$. Upon making the guiding-center transformation, this requirement is easily seen to become $H_{gc} = -mc^2/2$ (the guiding-center transformation is a diffeomorphism, so the mass shell is distorted but not topologically altered). So, when using the $(R, K, \mathcal{B}, \mu, \Theta)$ coordinates, f has support on a seven dimensional submanifold in an eight dimensional phase space. When we use the (R, U, μ, Θ) coordinates the phase space is ten dimensional, and when we use the $(R, U, \mu, \hat{\alpha})$ coordinates the phase space is thirteen dimensional; in all cases, however, f has support only on a manifold of seven dimensions thanks to the constraints on these coordinates.

The Lagrangian description keeps track of the dynamics of all the particles in the system as though they were distinguishable, and so it includes more degrees of freedom than the Eulerian description. That is why it is possible to write the Eulerian distribution $f(Z)$ in terms of the Lagrangian description $z(\eta, \tau)$, but it is impossible to do the reverse. There are many different functional forms for $z(\eta, \tau)$ that yield the same $f(Z)$. Nevertheless, for a plasma of indistinguishable particles (we are not going to bother about species labelling in this thesis) it is clear that any *physically relevant* quantity can be expressed in terms of the Eulerian distribution, $f(Z)$. This is because any physically relevant quantity should not depend on the identity of the individual particles in the system.

This is really a gauge invariance issue. The gauge group is the group of identical particle interchanges. The Lagrangian description keeps track of extra nonphysical *gauge degrees of freedom.* A physically relevant quantity can be written in terms of the Eulerian distribution since it is *gauge invariant* in this regard.

Consider for example the value of some phase function, $\Phi(z)$, summed over

all the particles in the system and integrated along world lines

$$N_\Phi = \int dN(\eta) \int d\tau \Phi\big(z(\eta,\tau)\big). \tag{5.739}$$

This object is invariant under the gauge group of identical particle interchanges because it can be written in terms of the Eulerian distribution as follows:

$$\begin{aligned} N_\Phi &= \int d^n Z \int dN(\eta) \int d\tau \delta^n\big(Z - z(\eta,\tau)\big)\Phi(Z) \\ &= \int d^n Z f_n(Z)\Phi(Z). \end{aligned} \tag{5.740}$$

Though we shall frequently work with the Lagrangian description of things, we must be able to show that our results can be expressed in terms of the Eulerian distribution. Fortunately, this will pose no problem.

The Lagrangian description of the *dynamics* of the system is then given by

$$\dot{z}(\eta,\tau) = V\big([z], z(\eta,\tau)\big), \tag{5.741}$$

where the dot denotes differentiation with respect to τ, and where V is the dynamical vector field expressed as a function of $z(\eta,\tau)$ and as a functional of z (since the dynamics of one particle may depend on the phase space positions of all the other particles in the system). The corresponding Eulerian description of the dynamics is then found as follows:

$$\begin{aligned} 0 &= -\int dN(\eta) \int d\tau \frac{d}{d\tau}\delta^n\big(Z - z(\eta,\tau)\big) \\ &= \int dN(\eta) \int d\tau \dot{z}(\eta,\tau) \cdot \frac{\partial}{\partial Z}\delta^n\big(Z - z(\eta,\tau)\big) \\ &= \frac{\partial}{\partial Z} \cdot \left[\int dN(\eta) \int d\tau \dot{z}(\eta,\tau)\delta^n\big(Z - z(\eta,\tau)\big)\right] \\ &= \frac{\partial}{\partial Z} \cdot \left[\int dN(\eta) \int d\tau V\big([z], z(\eta,\tau)\big)\delta^n\big(Z - z(\eta,\tau)\big)\right] \\ &= \frac{\partial}{\partial Z} \cdot \left[V([f_n], Z) \int dN(\eta) \int d\tau \delta^n\big(Z - z(\eta,\tau)\big)\right] \\ &= \frac{\partial}{\partial Z} \cdot \big[V([f_n], Z) f_n(Z)\big]. \end{aligned} \tag{5.742}$$

The first line above follows from the fact that at any finite time τ is finite, so the delta function vanishes at the limits of integration $\tau \to \pm\infty$. Note that we had to assume that the functional dependence of V on z could be replaced by a functional dependence on f_n; this is just a statement of the very reasonable condition that the dynamics cannot depend on particle labels. The resulting kinetic equation for $f_n(Z)$ is called the *continuity equation*, and it expresses conservation of particles. It is true for any relativistic system of particles, regardless of the nature of the forces involved (they could even be dissipative in nature).

5.2.2 Conservation of Phase Space Volume

One thing that distinguishes Hamiltonian systems from other dynamical systems is the property that phase space volume is conserved by a Hamiltonian flow. This means that if we take a volume element in phase space and drag each point of its boundary surface along a Hamiltonian vector field for some parameter increment, the volume enclosed will be unchanged. As we shall now see, this property follows from the Jacobi identity; this fact was used in Section 3.10 as an argument for using brackets that satisfy the Jacobi identity *exactly* (as opposed to satisfying it only to some order in an expansion parameter).

Suppose that we have a set of canonical coordinates Z_c, and that the Eulerian distribution function in these coordinates is $f_c(Z_c)$. Now under a (possibly non-canonical) coordinate transformation, $Z_c \mapsto Z$, a distribution function transforms in such a way as to keep the number of particles in a fixed phase space volume element constant. That is

$$f(Z)d^nZ = f_c(Z_c)d^nZ_c, \tag{5.743}$$

where n is the number of dimensions in phase space. Thus, f transforms like a pseudoscalar,

$$f(Z) = f_c(Z_c)D, \tag{5.744}$$

where we have defined the Jacobian of the transformation

$$D(Z) = \frac{\partial^n Z_c}{\partial^n Z}. \tag{5.745}$$

Alternatively, we can define a scalar distribution function, $\mathsf{f}(Z)$, which transforms as follows:

$$\mathsf{f}(Z) = f_c(Z_c). \tag{5.746}$$

It follows that in *any* coordinate system we have

$$f(Z) = \mathsf{f}(Z)D(Z). \tag{5.747}$$

Note that $f(Z) = \mathsf{f}(Z)$ in any *canonical* coordinate system, since the Jacobian of a canonical transformation is unity. In noncanonical coordinates, however, $f(Z)$ and $\mathsf{f}(Z)$ are different.

The Lagrangian two-form in coordinate system Z is given by

$$\Omega_{\mu\nu} = \frac{\partial Z_c^\alpha}{\partial Z^\mu}\frac{\partial Z_c^\beta}{\partial Z^\nu}\Omega^c_{\alpha\beta}, \tag{5.748}$$

where Ω^c is the canonical Lagrangian two-form. Taking the determinant of both sides, we find

$$\det\Omega = D^2. \tag{5.749}$$

We now no longer need to make reference to the canonical coordinate system, Z_c. Eqs. (5.747) and (5.749) tell us all we need to know, and they are written entirely in the general coordinates, Z.

Take the gradient of both sides of Eq. (5.749) to get

$$\begin{aligned} 2DD_{,\alpha} &= (\det\Omega)_{,\alpha} \\ &= D^2 J^{\beta\gamma}\Omega_{\gamma\beta,\alpha}, \end{aligned} \tag{5.750}$$

where we used the formula for the derivative of a determinant,

$$(\det A)_{,\alpha} = (\det A)(A^{-1})^{\beta\gamma}A_{\gamma\beta,\alpha}. \tag{5.751}$$

We are now ready to prove Liouville's theorem. We have

$$\begin{aligned} D(\dot{Z}^\alpha D)_{,\alpha} &= D(J^{\alpha\beta}H_{,\beta}D)_{,\alpha} \\ &= D^2 H_{,\beta}(J^{\alpha\beta}{}_{,\alpha} + \frac{1}{2}J^{\alpha\beta}J^{\mu\nu}\Omega_{\nu\mu,\alpha}) \\ &= \frac{1}{2}D^2 H_{,\beta}J^{\beta\mu}J^{\alpha\nu}(\Omega_{\nu\mu,\alpha} + \Omega_{\mu\alpha,\nu} + \Omega_{\alpha\nu,\mu}) \\ &= 0, \end{aligned} \tag{5.752}$$

where we used the above formula for $DD_{,\alpha}$, and where we used the Jacobi identity in the last step. Thus, since D is never zero, we have proved Liouville's theorem,

$$\frac{\partial}{\partial Z}\cdot(\dot{Z}D) = 0. \tag{5.753}$$

Now Eq. (5.742) may be written for a Hamiltonian system as follows:

$$\begin{aligned} 0 &= \frac{\partial}{\partial Z}\cdot(\dot{Z}f) \\ &= \frac{\partial}{\partial Z}\cdot(\dot{Z}D\mathsf{f}). \end{aligned} \tag{5.754}$$

Applying Liouville's theorem, we get the Vlasov equation,

$$0 = \dot{Z}\cdot\frac{\partial \mathsf{f}}{\partial Z}. \tag{5.755}$$

Our proof of this result has been quite general, and so in the future we can simply write down the Vlasov equation for any Hamiltonian equations of motion.

The careful reader will have noticed that we assumed invertibility of the Poisson tensor in the above proof, whereas our Poisson tensors in the (R, U, μ, Θ) and $(R, U, \mu, \hat{\alpha})$ coordinate systems are definitely singular. Recall, however, that we showed in Section 3.14 how these constrained coordinate systems could be imbedded in larger unconstrained coordinate systems. That is, we can obtain the (R, U, μ, Θ) coordinates by a smooth coordinate transformation from the $(R, K, \mathcal{B}, \mu, \Theta, C_{1a}, C_{1b})$ coordinates, and we can obtain the $(R, U, \mu, \hat{\alpha})$ coordinates by a smooth coordinate transformation from

the $(R, K, \mathcal{B}, \mu, \Theta, C_{1a}, C_{1b}, C_{2a}, C_{2b}, C_3)$ coordinates. In both cases, the physical motion takes place on the subspace for which $C_{1a} = C_{1b} = C_{2a} = C_{2b} = 0$ and $C_3 = 1$; if the initial conditions are on this subspace, the dynamics will keep them there. From this point of view, there is nothing singular about the transformation that led to these coordinate systems, and the only reason that their Poisson tensors are singular is that we enforced the constraints by setting $C_{1a} = C_{1b} = C_{2a} = C_{2b} = 0$ and $C_3 = 1$ at the very end of the calculation that led to them.

Armed with this insight, it is easy to compute the Jacobian D for these coordinate systems. First we consider the guiding-center transformation that led to the $(R, K, \mathcal{B}, \mu, \Theta)$ coordinates from canonical coordinates. The Jacobian of this transformation is

$$D_1 = \sqrt{\det \Omega_{gc}}, \tag{5.756}$$

where Ω_{gc} is the Lagrangian two-form given in Eqs. (3.490) through (3.494). The result is

$$D_1 = \frac{m^3}{\epsilon} K \Omega_B \Upsilon. \tag{5.757}$$

The coordinates $(C_{1a}, C_{1b}, C_{2a}, C_{2b}, C_3)$, which can be thought of as describing directions transverse to those described by the $(R, K, \mathcal{B}, \mu, \Theta)$ coordinates, are unaffected by the above transformation.

We now transform to either the (R, U, μ, Θ) system or the $(R, U, \mu, \hat{\alpha})$ system. This transformation will involve the coordinates $(C_{1a}, C_{1b}, C_{2a}, C_{2b}, C_3)$. Its Jacobian is given by

$$D_2 = \frac{\partial(R, U, \mu, \Theta)}{\partial(R, K, \mathcal{B}, \mu, \Theta, C_{1a}, C_{1b})} \tag{5.758}$$

or

$$D_2 = \frac{\partial(R, U, \mu, \hat{\alpha})}{\partial(R, K, \mathcal{B}, \mu, \Theta, C_{1a}, C_{1b}, C_{2a}, C_{2b}, C_3)}, \tag{5.759}$$

respectively. We can use the transformation equations, Eqs. (3.562) or (3.591), to calculate the above expressions. The important thing is that we take *all* of the

derivatives involved in calculating the Jacobian *before* enforcing the constraints by setting $C_{1a} = C_{1b} = C_{2a} = C_{2b} = 0$ and $C_3 = 1$. The calculation is straightforward, and we find that for either the (R, U, μ, Θ) or the $(R, U, \mu, \hat{\alpha})$ coordinates we get

$$D_2 = \frac{1}{2K\lambda_B^2}. \tag{5.760}$$

The *overall* Jacobian of the above transformation is thus

$$D = D_1 D_2 = \frac{em^2}{2\epsilon c \lambda_B(R)} \Upsilon'(R), \tag{5.761}$$

where Υ' is given by Eq. (3.575). Note that this same expression may be used for the guiding/oscillation-center problem, since it has exactly the same brackets as the guiding-center problem with no wave present. This is because our oscillation-center Lie transform took the wave perturbation out of the brackets and put it into the Hamiltonian (which is how we got K_2).

Thus by imbedding our singular coordinate systems in larger nonsingular ones, we are able to validate the above derivation of the Vlasov equation for our coordinates. Because we had to introduce the coordinates $(C_{1a}, C_{1b}, C_{2a}, C_{2b}, C_3)$, however, we should ask what the distribution function looks like, and whether or not the kinetic equation that we have started with makes sense. Consider Eq. (5.737), written for the coordinate system $Z = (R, K, \mathcal{B}, \mu, \Theta, C_{1a}, C_{1b}, C_{2a}, C_{2b}, C_3)$. We adopt the shorthand notation $Z = (Y, C)$ where $Y = (R, K, \mathcal{B}, \mu, \Theta)$ and $C = (C_{1a}, C_{1b}, C_{2a}, C_{2b}, C_3)$. Then we have

$$f_{13}(Z) = \int dN(\eta) \int d\tau \delta^8 \left(Y - y\left(\eta, \tau\right)\right) \delta^5 \left(C - c\left(\eta, \tau\right)\right), \tag{5.762}$$

where $y(\eta, \tau)$ and $c(\eta, \tau)$ give the dynamics of Y and C, respectively. Note, however, that since the integral, $\int dN(\eta)$, includes only particles that obey the constraints $C_{1a} = C_{1b} = C_{2a} = C_{2b} = 0$ and $C_3 = 1$, and since the dynamics is known to keep such particles on the constraint surface, it must be that $c(\eta, \tau) = (0, 0, 0, 0, 1)$. Thus the delta functions involving C can be pulled out of the integral

to finally yield

$$f_{13}(Z) = \delta(C_{1a})\delta(C_{1b})\delta(C_{2a})\delta(C_{2b})\delta(C_3 - 1)\int dN(\eta)\int d\tau\delta^8\left(Y - y\left(\eta,\tau\right)\right). \tag{5.763}$$

The proportionality of f_{13} to delta functions in the C is simply a mathematical restatement of our earlier observation that it has support only on a space of dimension less than that coordinatized by Z. In fact, it has support only on a space of seven dimensions (there is another delta function still hiding in the integral on the right hand side of the above equation due to the fact that the Hamiltonian is a constant of the motion). The Vlasov equation written in these coordinates is then

$$0 = \dot{Y}\cdot\frac{\partial \mathsf{f}_{13}}{\partial Y}, \tag{5.764}$$

where $\mathsf{f}_{13} = f_{13}/D_1$ and where the terms $\dot{C}\cdot\partial \mathsf{f}_{13}/\partial C$ are not present because $\dot{C} = 0$. We can now integrate the above Vlasov equation over the C coordinates to get

$$0 = \dot{Y}\cdot\frac{\partial \mathsf{f}_8(Y)}{\partial Y}, \tag{5.765}$$

where

$$\begin{aligned} f_8(Y) &= \int d^5C f_{13}(Z) \\ &= \int dN(\eta)\int d\tau\delta^8\left(Y - y\left(\eta,\tau\right)\right), \end{aligned} \tag{5.766}$$

and $\mathsf{f}_8 = f_8/D_1$. This is obviously the same Vlasov equation that we would have obtained if we had used only the clearly nonpathological $(R, K, \mathcal{B}, \mu, \Theta)$ coordinates from the start.

It turns out to be easier (for reasons that will become clear shortly) to write the Vlasov equation in terms of f and easier to write the field equation in terms of f. Since we know what D is, however, there is clearly no problem involved in writing both equations in terms of either f or f (recall that f and f are related by Eq. (5.747) with D given by Eq. (5.761)).

5.3 Self-Consistent Kinetic and Field Equations

5.3.1 Constructing the System Action

We begin by considering the case in which there is no eikonal wave field present. Our action one-form and Hamiltonian for a single guiding-center are thus given by Eqs. (3.604) and (3.581), respectively. In Section 5.5, we generalize our results to the case in which the plasma is bathed in an eikonal wave field. For now we construct the action for the coupled system of guiding-center plasma and Maxwell field. This has the form

$$S = S_{gc} + S_m, \tag{5.767}$$

where S_{gc} is the total action of the guiding centers, and where S_m is the action of the Maxwell field.

Now the action of the guiding centers is found by simply summing that for a single guiding center over the full distribution. Thus we write

$$\begin{aligned} S_{gc}[Z, A_i] = \int dN(\eta) \int d\tau \Big[& \Gamma_{gc}\big(Z(\eta,\tau); A_i(R(\eta,\tau)), F_i(R(\eta,\tau))\big) \cdot \dot{Z}(\eta,\tau) \\ & - \sum_\nu \lambda_\nu(\eta,\tau) C_\nu\big(Z(\eta,\tau); F_i(R(\eta,\tau))\big) \\ & - H_{gc}\big(Z(\eta,\tau); F_i(R(\eta,\tau))\big) \Big]. \end{aligned} \tag{5.768}$$

Here we have written Z for the full set of boostgauge and gyrogauge invariant guiding-center coordinates, $(R, U, \mu, \hat{\alpha})$. We have enforced the constraints by means of Lagrange multipliers, using λ_ν to denote the multiplier for constraint C_ν, where the index ν runs over all the constraints present as usual. Finally, we have indicated separately the functional dependence of the various terms on the four potential A_i and the background field F_i (here i denotes the ordering of the field as discussed in Section 3.7).

Now Eq. (5.768) may be written in the form

$$S_{gc} = \int d^4x \mathcal{L}_{gc}, \tag{5.769}$$

where x denotes spacetime position, and where we have defined the Lagrangian density for the guiding centers,

$$\mathcal{L}_{gc}(x) = \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \Big[\Gamma_{gc}(Z(\eta,\tau); A_i, F_i) \cdot \dot{Z}(\eta,\tau) - \sum_\nu \lambda_\nu(\eta,\tau) C_\nu(Z(\eta,\tau); F_i) - H_{gc}(Z(\eta,\tau); F_i) \Big]. \quad (5.770)$$

Here we have adopted the convention that A_i and F_i denote $A_i(x)$ and $F_i(x)$, respectively.

The Maxwell action is well known to be (see, for example, Jackson [44])

$$S_m = \int d^4x \mathcal{L}_m(x), \quad (5.771)$$

where the Lagrangian density for the Maxwell field is

$$\mathcal{L}_m = -\frac{1}{16\pi}(F_0 + \epsilon F_1 + \cdots) : (F_0 + \epsilon F_1 + \cdots). \quad (5.772)$$

In this study, we shall retain terms in $\mathcal{L}_m$ only to order ϵ; thus we write

$$\mathcal{L}_m = -\frac{1}{16\pi}(F_0 : F_0 + 2\epsilon F_0 : F_1). \quad (5.773)$$

5.3.2 The Vlasov Equation for Guiding Centers

We first vary the system action with respect to the particle field, $Z(\eta,\tau)$. After a short calculation, we find

$$\begin{aligned} 0 &= \frac{\delta S}{\delta Z(\eta,\tau)} \\ &= \Omega_{gc}(Z(\eta,\tau); A_i(R(\eta,\tau)), F_i(R(\eta,\tau))) \cdot \dot{Z}(\eta,\tau) \\ &\quad - \sum_\nu \lambda_\nu(\eta,\tau) \frac{\partial C_\nu}{\partial Z}(Z(\eta,\tau); F_i(R(\eta,\tau))) \\ &\quad - \frac{\partial H_{gc}}{\partial Z}(Z(\eta,\tau); F_i(R(\eta,\tau))). \end{aligned} \quad (5.774)$$

where $\Omega_{gc} = d\Gamma_{gc}$. This equation, coupled with the constraints

$$C_\nu(Z(\eta,\tau); F_i(R(\eta,\tau))) = \delta_{\nu 3} \quad (5.775)$$

(which are needed to determine the Lagrange multipliers), shows clearly that the fields, $Z(\eta, \tau)$, obey the usual equations of motion for a single guiding center. Knowing this, and using the ideas developed in the previous section, it is now possible to write down the Vlasov equation,

$$0 = \dot{Z} \cdot \frac{\partial \mathsf{f}}{\partial Z}, \tag{5.776}$$

using the equations of motion for a single guiding center.

In particular, if we use the (R, U, μ, Θ) coordinates, this becomes

$$0 = \dot{R} \cdot \frac{\partial \mathsf{f}_{10}}{\partial R} + \dot{U} \cdot \frac{\partial \mathsf{f}_{10}}{\partial U} + \dot{\mu} \frac{\partial \mathsf{f}_{10}}{\partial \mu} + \dot{\Theta} \frac{\partial \mathsf{f}_{10}}{\partial \Theta}. \tag{5.777}$$

We can now define the *guiding-center distribution function*,

$$\bar{\mathsf{f}}_9(R, U, \mu) \equiv \int_0^{2\pi} d\Theta \mathsf{f}_{10}(R, U, \mu, \Theta). \tag{5.778}$$

This is nothing more than 2π times the Θ-average of the full distribution function f_{10}. Now because $\dot{Z}$ is independent of Θ (thanks to our guiding-center transformation) and because $\dot{\mu} = 0$, taking the Θ-average of the above kinetic equation yields

$$0 = \dot{R} \cdot \frac{\partial \bar{\mathsf{f}}_9}{\partial R} + \dot{U} \cdot \frac{\partial \bar{\mathsf{f}}_9}{\partial U} \tag{5.779}$$

This is the *reduced* kinetic equation for the guiding-center distribution function.

5.3.3 The Field Equations

Generally speaking, the idea is now to vary the above action with respect to the four potential to get the dynamical equations for the fields. This must be done carefully, however, as there are two additional constraints that such variation must respect. Recall that in our derivation of the guiding-center action we assumed that the background field scale lengths were large in comparison to the gyroradius, and we assumed that the zero-order fields have $\lambda_E = 0$. We must

make certain that the dynamics of the fields do not evolve them into a configuration for which either of these assumptions are violated. In order to get dynamical equations for the fields that respect these constraints, our variation of the action with respect to the four potential must be a *constrained variation*; that is, arbitrary variations of the four potential are not allowed. Only those variations of the four potential that preserve the vanishing of λ_E to lowest order and the smallness of the ratio of gyroradius to scale length are allowed.

We thus begin our derivation of the field equations by examining the variation of the action due to variations of the A_i, without assuming in any way that the variations of the A_i are arbitrary. Recall that we have indicated separately the functional dependence of the various terms in the action on the four potential A_i and the background field F_i. Of course, $F_i = dA_i$, so when we vary with respect to the A_i we must take into account the F_i dependence. To do this, it is convenient to distinguish between *total* and *partial* functional derivatives with respect to A_i. We use the chain rule to write

$$\left.\frac{\delta S}{\delta A_{i\rho}(x)}\right|_{\text{total}} = \frac{\delta S}{\delta A_{i\rho}(x)} + \int d^4x' \frac{\delta S}{\delta F_{i\mu\nu}(x')} \frac{\delta F_{i\mu\nu}(x')}{\delta A_{i\rho}(x)}. \tag{5.780}$$

To proceed, note that

$$\begin{aligned} F_{i\mu\nu}(x') &= A_{i\nu,\mu}(x') - A_{i\mu,\nu}(x') \\ &= \int d^4x \delta^4(x-x') A_{i\nu,\mu}(x) - A_{i\mu,\nu}(x) \\ &= \int d^4x \left\{ A_{i\mu}[\delta^4(x-x')]_{,\nu} - A_{i\nu}[\delta^4(x-x')]_{,\mu} \right\}, \end{aligned} \tag{5.781}$$

so that

$$\frac{\delta F_{i\mu\nu}(x')}{\delta A_{i\rho}(x)} = \delta_{\mu\rho}[\delta^4(x-x')]_{,\nu} - \delta_{\nu\rho}[\delta^4(x-x')]_{,\mu}. \tag{5.782}$$

Using this in Eq. (5.780), we get

$$\left.\frac{\delta S}{\delta A_i}\right|_{\text{total}} = \frac{\delta S}{\delta A_i} - 2\vec{\nabla} \cdot \left(\frac{\delta S}{\delta F_i}\right). \tag{5.783}$$

This formula is very useful in what follows.

Using Eq. (5.783) to vary the action with respect to the four potential, we arrive straightforwardly at the following result:

$$\delta S = \int d^4x \left[\mathcal{J}_0(x) \cdot \delta A_0(x) + \mathcal{J}_1(x) \cdot \delta A_1(x) \right], \tag{5.784}$$

where we have defined

$$\mathcal{J}_0(x) \equiv \frac{1}{c} J(x) + \frac{1}{4\pi} \vec{\nabla} \cdot G_0(x) \tag{5.785}$$

and

$$\mathcal{J}_1(x) \equiv \frac{\epsilon}{c} J(x) + \frac{\epsilon}{4\pi} \vec{\nabla} \cdot G_1(x), \tag{5.786}$$

where in turn we have defined the *guiding-center current density*

$$\begin{aligned} J(x) &\equiv c \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \frac{\partial \Gamma_{gc}}{\partial A_0}(Z(\eta,\tau); A_i, F_i) \cdot \dot{Z}(\eta,\tau) \\ &= \frac{c}{\epsilon} \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \frac{\partial \Gamma_{gc}}{\partial A_1}(Z(\eta,\tau); A_i, F_i) \cdot \dot{Z}(\eta,\tau) \\ &= \frac{e}{\epsilon} \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \dot{R}(\eta,\tau) \\ &= \frac{e}{\epsilon} \int dR \int dU \int d\mu \int d\Theta f_{10}(R,U,\mu,\Theta) \delta^4(x - R) \dot{R}(R,U,\mu) \\ &= \frac{e}{\epsilon} \int dR \int dU \int d\mu \overline{f}_9(R,U,\mu) \delta^4(x - R) \dot{R}(R,U,\mu), \end{aligned} \tag{5.787}$$

and the *macroscopic field tensors*

$$G_0(x) \equiv F_0(x) + \epsilon F_1(x) - 4\pi M_0(x) \tag{5.788}$$

$$G_1(x) \equiv F_0(x) - 4\pi M_1(x), \tag{5.789}$$

and where in turn we have defined the *guiding-center magnetization densities*

$$\begin{aligned} M_0(x) \equiv 2 \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \Big[&\frac{\partial \Gamma}{\partial F_0}(Z(\eta,\tau); A_i, F_i) \cdot \dot{Z}(\eta,\tau) \\ &- \sum_\nu \lambda_\nu(\eta,\tau) \frac{\partial C_\nu}{\partial F_0}(Z(\eta,\tau); F_i) - \frac{\partial H_{gc}}{\partial F_0}(Z(\eta,\tau); F_i) \Big] \end{aligned}$$

$$= 2\int dR \int dU \int d\mu \int d\Theta f_{10}(R,U,\mu,\Theta)\delta^4(x-R)\Big[\frac{\partial\Gamma}{\partial F_0}(Z;A_i,F_i) \cdot \dot{Z}(R,U,\mu) - \sum_\nu \lambda_\nu \frac{\partial C_\nu}{\partial F_0}(Z;F_i) - \frac{\partial H_{gc}}{\partial F_0}(Z;F_i)\Big]$$

$$= 2\int dR \int dU \int d\mu \bar{f}_9(R,U,\mu)\delta^4(x-R)\Big[\frac{\partial\Gamma}{\partial F_0}(Z;A_i,F_i)\cdot\dot{Z}(R,U,\mu) - \sum_\nu \lambda_\nu \frac{\partial C_\nu}{\partial F_0}(Z;F_i) - \frac{\partial H_{gc}}{\partial F_0}(Z;F_i)\Big] \tag{5.790}$$

$$M_1(x) \equiv \frac{2}{\epsilon}\int dN(\eta)\int d\tau\delta^4(x-R(\eta,\tau))\Big[\frac{\partial\Gamma}{\partial F_1}(Z(\eta,\tau);A_i,F_i)\cdot\dot{Z}(\eta,\tau) - \sum_\nu \lambda_\nu(\eta,\tau)\frac{\partial C_\nu}{\partial F_1}(Z(\eta,\tau);F_i) - \frac{\partial H_{gc}}{\partial F_1}(Z(\eta,\tau);F_i)\Big]$$

$$= 2\int dR \int dU \int d\mu \int d\Theta f_{10}(R,U,\mu,\Theta)\delta^4(x-R)\Big[\frac{\partial\Gamma}{\partial F_1}(Z;A_i,F_i) \cdot \dot{Z}(R,U,\mu) - \sum_\nu \lambda_\nu \frac{\partial C_\nu}{\partial F_1}(Z;F_i) - \frac{\partial H_{gc}}{\partial F_1}(Z;F_i)\Big]$$

$$= 2\int dR \int dU \int d\mu \bar{f}_9(R,U,\mu)\delta^4(x-R)\Big[\frac{\partial\Gamma}{\partial F_1}(Z;A_i,F_i)\cdot\dot{Z}(R,U,\mu) - \sum_\nu \lambda_\nu \frac{\partial C_\nu}{\partial F_1}(Z;F_i) - \frac{\partial H_{gc}}{\partial F_1}(Z;F_i)\Big]. \tag{5.791}$$

Note that the magnetization came from the second term on the right of Eq. (5.783). Also note that the only thing that depends explicitly on F_1 is the first-order piece of the Hamiltonian, so that only the last term in square brackets in the above expression for M_1 survives; of course, F_1 also appears in the brackets due to the A_1 dependence of Γ_{gc}. Finally note that we were able to write the current and the magnetizations in terms of the reduced Eulerian distribution function, $\bar{f}_9$.

Now because the δA_i are not arbitrary, we cannot simply set $\mathcal{J}_0 = \mathcal{J}_1 = 0$. Instead, as discussed above, we must restrict the variation so that it respects the constraints that $\lambda_E = 0$ to lowest order and that the ratio of gyroradius to scale length is small. To deal with the first of these constraints, let us temporarily introduce Clebsch variables for the fields. We define four scalar fields,

$\alpha(x), \beta(x), \kappa(x), \sigma(x)$, such that in terms of these fields the four potential is given by

$$A_0 = \alpha d\beta \tag{5.792}$$

$$A_1 = \kappa d\sigma, \tag{5.793}$$

and consequently the field tensor is given by

$$F_0 = dA_0 = d(\alpha d\beta) = d\alpha \wedge d\beta \tag{5.794}$$

$$F_1 = dA_1 = d(\kappa d\sigma) = d\kappa \wedge d\sigma. \tag{5.795}$$

That such scalar fields exist is guaranteed by the Darboux theorem. That is, because F is a closed two-form, it can be written in the form $F = d\alpha \wedge d\beta + \epsilon d\kappa \wedge d\sigma$, where we are guaranteed enough freedom to choose α and β such that $P_{\|} \cdot (d\alpha \wedge d\beta) = 0$.

It is clear that the above construction insures that

$$P_{\|} \cdot F_0 = 0. \tag{5.796}$$

Note that we are ignoring F_i for $i \geq 2$, and that the parallel electric field must lie entirely within F_1. Thus, the specification of the four functions $\alpha(x), \beta(x), \kappa(x)$, and $\sigma(x)$ is a coordinatization of the function space of all electromagnetic fields that automatically ensures the satisfaction of the constraint that $\lambda_E = 0$ to lowest order.

The variation of the action with respect to the four potentials may now be written

$$\begin{aligned}
\delta S &= \int d^4x \left[\mathcal{J}_0 \cdot \delta(\alpha \vec{\nabla} \beta) + \mathcal{J}_1 \cdot \delta(\kappa \vec{\nabla} \sigma) \right] \\
&= \int d^4x (\delta\alpha \mathcal{J}_0 \cdot \vec{\nabla} \beta + \alpha \mathcal{J}_0 \cdot \vec{\nabla} \delta\beta + \delta\kappa \mathcal{J}_1 \cdot \vec{\nabla} \sigma + \kappa \mathcal{J}_1 \cdot \vec{\nabla} \delta\sigma) \\
&= \int d^4x \left[\delta\alpha \mathcal{J}_0 \cdot \vec{\nabla} \beta - \delta\beta \vec{\nabla} \cdot (\alpha \mathcal{J}_0) + \delta\kappa \mathcal{J}_1 \cdot \vec{\nabla} \sigma - \delta\sigma \vec{\nabla} \cdot (\kappa \mathcal{J}_1) \right]
\end{aligned} \tag{5.797}$$

We still cannot set the coefficients of the variations equal to zero, however, because of the remaining constraint that the fields remain sufficiently slowly varying for the guiding-center approximation to remain valid. This point requires some discussion.

Consider a general Fourier decomposition of the electromagnetic field in and around a plasma. We can divide the Fourier space into three regions. The first consists of slowly varying fields for which the guiding-center approximation is clearly valid; we call these *background fields*. The second consists of rapidly varying fields that are due to collective motion of the plasma; we call these *wave fields*, and their effect on a single guiding center was the subject of Chapter 4. Note that wave fields violate the guiding-center approximation, and the only reason that we were able to treat them perturbatively was our assumption that their amplitudes are small. The third consists of the extremely rapid fluctuations associated with collisions and higher correlations.

Now fields belonging to the third region of Fourier space are clearly outside of the scope of this thesis; our Vlasov kinetic description of the plasma neglects correlations. Wave fields were studied in a single particle context in Chapter 4, and their self-consistent evolution will be studied in Section 5.5. For now we are interested in the dynamics of the background fields. We thus define a projection operator, $\mathcal{P}$, that, when applied to an arbitrary field, projects out the part that is slowly varying. We shall not be specific about the nature of this operator except to say that, since it is a projection operator, we expect it to be idempotent. A moment's thought convinces one that this means that it must be a convolution of the field with a filter function whose Fourier transform is piecewise constant, having a value of either zero or one everywhere in Fourier space. Specifically, it has a value of one in the first of the above-described three regions of Fourier space, and a value of zero in the other two regions. Exactly how one draws these boundaries is what we are leaving unspecified.

Thus, although we cannot set the coefficients of $\delta\alpha(x), \delta\beta(x), \delta\kappa(x)$, and $\delta\sigma(x)$ equal to zero in Eq. (5.797), we can enforce the constraint that the fields are slowly varying by requiring that their *variations* be slowly varying; thus

$$\delta\alpha(x) = \mathcal{P}\delta\alpha(x) \tag{5.798}$$

(and similarly for the other three variations). We can also decompose the coefficients of the variations into slowly varying and rapidly varying parts; thus

$$\mathcal{J}_0 \cdot \vec{\nabla}\,\beta = \mathcal{P}(\mathcal{J}_0 \cdot \vec{\nabla}\,\beta) + (\mathbf{1} - \mathcal{P})(\mathcal{J}_0 \cdot \vec{\nabla}\,\beta) \tag{5.799}$$

(and similarly for the other three coefficients). Thus, upon multiplying $\delta\alpha(x)$ and $\mathcal{J}_0 \cdot \vec{\nabla}\,\beta$, we get the product of the slowly varying terms and a cross term. Now the cross term is clearly oscillatory and vanishes upon integration over x. It is then legal to set the coefficients of the *slowly varying* parts of the variations equal to zero. This essentially means that we can set the projection of the coefficients of the variations in Eq. (5.797) equal to zero.

Thus, we get

$$\mathcal{P}[\mathcal{J}_0 \cdot \vec{\nabla}\,\beta] = 0 \tag{5.800}$$

$$\mathcal{P}[\vec{\nabla} \cdot (\alpha\mathcal{J}_0)] = 0 \tag{5.801}$$

$$\mathcal{P}[\mathcal{J}_1 \cdot \vec{\nabla}\,\sigma] = 0 \tag{5.802}$$

$$\mathcal{P}[\vec{\nabla} \cdot (\kappa\mathcal{J}_1)] = 0. \tag{5.803}$$

Now note that from Eq. (5.787), we have

$$\begin{aligned} J^{\mu}{}_{,\mu} &= \frac{e}{\epsilon}\int dN(\eta)\int d\tau \frac{\partial\delta^4\left(x - R(\eta,\tau)\right)}{\partial x^{\mu}}\dot{R}^{\mu}(\eta,\tau) \\ &= -\frac{e}{\epsilon}\int dN(\eta)\int d\tau \frac{\partial\delta^4\left(x - R(\eta,\tau)\right)}{\partial R^{\mu}}\dot{R}^{\mu}(\eta,\tau) \\ &= -\frac{e}{\epsilon}\int dN(\eta)\int dR^{\mu} \frac{\partial\delta^4\left(x - R(\eta,\tau)\right)}{\partial R^{\mu}} \\ &= 0, \end{aligned} \tag{5.804}$$

where the last step follows from the fact that the delta function vanishes at the limits of integration for finite x. This result expresses conservation of particles. From this it follows that

$$\vec{\nabla} \cdot \mathcal{J}_0 = \frac{1}{\epsilon} \vec{\nabla} \cdot \mathcal{J}_1 = \frac{1}{c} \vec{\nabla} \cdot J = 0. \tag{5.805}$$

So our field equations become

$$\mathcal{P}[\mathcal{J}_0 \cdot \vec{\nabla} \alpha] = \mathcal{P}[\mathcal{J}_0 \cdot \vec{\nabla} \beta] = 0 \tag{5.806}$$

$$\mathcal{P}[\mathcal{J}_1 \cdot \vec{\nabla} \kappa] = \mathcal{P}[\mathcal{J}_1 \cdot \vec{\nabla} \sigma] = 0. \tag{5.807}$$

Thus it follows that

$$\mathcal{P}[\mathcal{J}_0 \cdot (\vec{\nabla} \alpha \vec{\nabla} \beta - \vec{\nabla} \beta \vec{\nabla} \alpha)] = 0 \tag{5.808}$$

$$\mathcal{P}[\mathcal{J}_1 \cdot (\vec{\nabla} \kappa \vec{\nabla} \sigma - \vec{\nabla} \sigma \vec{\nabla} \kappa)] = 0, \tag{5.809}$$

or

$$\mathcal{P}[F_0 \cdot \mathcal{J}_0] = 0 \tag{5.810}$$

$$\mathcal{P}[F_1 \cdot \mathcal{J}_1] = 0. \tag{5.811}$$

Note that the Clebsch potentials have disappeared from our final result; this was essential since they have a gauge freedom and we expect our result to be gauge invariant. We simply used the Clebsch potentials to enforce our constraints, and then we got rid of them.

The final results for the field equations are thus

$$\mathcal{P}[F_0 \cdot (\frac{1}{4\pi} \vec{\nabla} \cdot G_0 + \frac{1}{c} J)] = 0 \tag{5.812}$$

$$\mathcal{P}[F_1 \cdot (\frac{1}{4\pi} \vec{\nabla} \cdot G_1 + \frac{1}{c} J)] = 0. \tag{5.813}$$

Note that the first describes field evolution due to perpendicular four current, while the second describes field evolution due to parallel four current.

5.3.4 Summary of Self-Consistent Kinetic and Field Equations

To summarize the results of this section, we present the complete set of kinetic and field equations for the guiding-center plasma. The kinetic equation is

$$0 = \dot{R} \cdot \frac{\partial \bar{f}_9}{\partial R} + \dot{U} \cdot \frac{\partial \bar{f}_9}{\partial U}, \tag{5.814}$$

where $\dot{R} = \{R, H_{gc}\}$ and $\dot{U} = \{U, H_{gc}\}$, and where in turn the Poisson brackets are given in Eqs. (3.569) through (3.574) and the Hamiltonian is given in Eq. (3.581). The field equations are then

$$\mathcal{P}[F_0 \cdot (\frac{1}{4\pi} \vec{\nabla} \cdot G_0 + \frac{1}{c} J)] = 0 \tag{5.815}$$

$$\mathcal{P}[F_1 \cdot (\frac{1}{4\pi} \vec{\nabla} \cdot G_1 + \frac{1}{c} J)] = 0, \tag{5.816}$$

where the current is given by

$$J(x) = \frac{e}{\epsilon} \int dR \int dU \int d\mu \bar{f}_9(R, U, \mu) \delta^4(x - R) \dot{R}(R, U, \mu) \tag{5.817}$$

and the macroscopic field tensors are given by

$$G_0(x) \equiv F_0(x) + \epsilon F_1(x) - 4\pi M_0(x) \tag{5.818}$$

$$G_1(x) \equiv F_0(x) - 4\pi M_1(x), \tag{5.819}$$

and where in turn the magnetization densities are given by

$$M_0(x) = 2 \int dR \int dU \int d\mu \bar{f}_9(R, U, \mu) \delta^4(x - R) \Big[\frac{\partial \Gamma}{\partial F_0}(Z; A_i, F_i) \cdot \dot{Z}(R, U, \mu) - \sum_\nu \lambda_\nu \frac{\partial C_\nu}{\partial F_0}(Z; F_i) - \frac{\partial H_{gc}}{\partial F_0}(Z; F_i) \Big] \tag{5.820}$$

$$M_1(x) = 2 \int dR \int dU \int d\mu \bar{f}_9(R, U, \mu) \delta^4(x - R) \Big[\frac{\partial \Gamma}{\partial F_1}(Z; A_i, F_i) \cdot \dot{Z}(R, U, \mu) - \sum_\nu \lambda_\nu \frac{\partial C_\nu}{\partial F_1}(Z; F_i) - \frac{\partial H_{gc}}{\partial F_1}(Z; F_i) \Big]. \tag{5.821}$$

Of course, these must be supplemented by the homogeneous field equations,

$$\vec{\nabla} \cdot \mathcal{F}_0 = 0 \tag{5.822}$$

$$\vec{\nabla} \cdot \mathcal{F}_1 = 0. \tag{5.823}$$

Note that $\overline{f}_9(R, U, \mu)$ and $\bar{\mathrm{f}}_9(R, U, \mu)$ are related by

$$\overline{f}_9(R, U, \mu) = D(R)\bar{\mathrm{f}}_9(R, U, \mu), \tag{5.824}$$

where the Jacobian D is given by

$$D = \frac{em^2}{2\epsilon c \lambda_B(R)} \Upsilon'(R), \tag{5.825}$$

and where in turn $\Upsilon'(R)$ is given by Eq. (3.575).

5.4 Conservation Laws for the Guiding-Center Plasma

5.4.1 The Noether Method

We now employ Noether's theorem to deduce conservation laws for the energy-momentum and the angular momentum of the guiding-center plasma. The technique has been described by Similon [12], and we shall compare our results to his. We begin by considering the variation in the Lagrangian density due to the variation of all the fields. We start with $\mathcal{L} \equiv \mathcal{L}_{gc} + \mathcal{L}_m$, and apply the variation. Whenever terms involving the derivative of a variation appear, we replace them by a pure divergence minus a term for which the variation is not differentiated; this is almost like integration by parts, but since there is no integral sign, we must keep the pure divergence terms. When we are done, we shall find that $\delta\mathcal{L}$ is equal to a pure divergence minus terms, for each field present, that consist of the variation of that field times the corresponding equation of motion. Thus, if we

then use the equations of motion, we can reduce $\delta\mathcal{L}$ to a pure divergence. The algebra is tedious but very straightforward, and we get

$$\begin{aligned}\delta\mathcal{L}(x) =& \vec{\nabla}\cdot\Big\{\frac{1}{4\pi}\delta A_0\cdot G_0+\frac{\epsilon}{4\pi}\delta A_1\cdot G_1+\mathcal{J}_0\alpha\delta\beta+\mathcal{J}_1\kappa\delta\sigma\\ &+\int dN(\eta)\int d\tau\delta^4(x-R(\eta,\tau))\big[\dot{R}(\eta,\tau)\Gamma_{gc}(Z(\eta,\tau);A_i,F_i)\cdot\delta Z(\eta,\tau)\\ &-\delta R(\eta,\tau)\big(\Gamma_{gc}(Z(\eta,\tau);A_i,F_i)\cdot\dot{Z}(\eta,\tau)\\ &-\sum_\nu\lambda_\nu(\eta,\tau)C_\nu(Z(\eta,\tau);A_i,F_i)-H(Z(\eta,\tau);F_i)\big)\big]\Big\}. \qquad (5.826)\end{aligned}$$

5.4.2 Conservation of Energy-Momentum

To derive the conservation law for energy-momentum, we consider variations in the coordinates that effectively translate in spacetime all the particles of the plasma, the fields in the plasma, the external coils that generate the fields, etc. Following Similon [12], we write these as follows:

$$\delta R=\xi \qquad (5.827)$$

$$\delta U=0 \qquad (5.828)$$

$$\delta\mu=0 \qquad (5.829)$$

$$\delta\hat{\alpha}=0, \qquad (5.830)$$

where ξ is a constant vector. Thus, the particles' position coordinates are pushed forward without altering any of their other phase space coordinates. The fields translate according to the prescription

$$\delta\alpha=-\xi\cdot\vec{\nabla}\,\alpha \qquad (5.831)$$

$$\delta\beta=-\xi\cdot\vec{\nabla}\,\beta \qquad (5.832)$$

$$\delta\kappa=-\xi\cdot\vec{\nabla}\,\kappa \qquad (5.833)$$

$$\delta\sigma=-\xi\cdot\vec{\nabla}\,\sigma, \qquad (5.834)$$

so

$$\begin{aligned}\delta A_0 &= \delta(\alpha \vec{\nabla} \beta) \\ &= \delta\alpha \vec{\nabla} \beta + \alpha \vec{\nabla} \delta\beta \\ &= -\xi \cdot \vec{\nabla} \alpha \vec{\nabla} \beta - \alpha\xi \cdot \vec{\nabla}\vec{\nabla} \beta \\ &= -\xi \cdot (\vec{\nabla} \alpha \vec{\nabla} \beta + \alpha \vec{\nabla}\vec{\nabla} \beta) \\ &= -\xi \cdot \vec{\nabla} (\alpha \vec{\nabla} \beta) \\ &= -\xi \cdot \vec{\nabla} A_0, \end{aligned} \tag{5.835}$$

and similarly

$$\delta A_1 = -\xi \cdot \vec{\nabla} A_1. \tag{5.836}$$

Finally note that the Lagrangian densities transform like scalar fields so

$$\delta \mathcal{L}_{gc} = -\xi \cdot \vec{\nabla} \mathcal{L}_{gc} \tag{5.837}$$

$$\delta \mathcal{L}_m = -\xi \cdot \vec{\nabla} \mathcal{L}_m. \tag{5.838}$$

Inserting these into Eq. (5.826), a short manipulation yields

$$\vec{\nabla} \cdot T = 0, \tag{5.839}$$

where we have introduced the *stress-energy tensor*

$$\begin{aligned} T(x) \equiv & -\frac{1}{4\pi} G_0(x) \cdot F_0(x) - \frac{\epsilon}{4\pi} G_1(x) \cdot F_1(x) + \mathcal{L}_m \mathbf{1} \\ & + \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \dot{R}(\eta,\tau) m U(\eta,\tau) \\ = & -\frac{1}{4\pi} G_0(x) \cdot F_0(x) - \frac{\epsilon}{4\pi} G_1(x) \cdot F_1(x) + \mathcal{L}_m \mathbf{1} \\ & + \int dR \int dU \int d\mu \int d\Theta f_{10}(R,U,\mu,\Theta) \delta^4(x - R) \dot{R}(mU) \\ = & -\frac{1}{4\pi} G_0(x) \cdot F_0(x) - \frac{\epsilon}{4\pi} G_1(x) \cdot F_1(x) + \mathcal{L}_m \mathbf{1} \\ & + \int dR \int dU \int d\mu \bar{f}_9(R,U,\mu) \delta^4(x - R) \dot{R}(mU) \end{aligned} \tag{5.840}$$

Eq. (5.839) expresses conservation of energy-momentum in the guiding-center plasma. Note that the last form for the stress-energy tensor given in Eq. (5.840) expresses the result in terms of the reduced Eulerian distribution function, $\overline{f}_9$.

5.4.3 Conservation of Angular Momentum

To derive the conservation law for angular momentum, we consider variations in the coordinates that effectively rotate about the origin of spacetime all the particles of the plasma, the fields in the plasma, the external coils that generate the fields, etc. Following Similon [12], we write these as follows:

$$\delta R = \Omega \cdot R \tag{5.841}$$

$$\delta U = \Omega \cdot U \tag{5.842}$$

$$\delta \mu = 0 \tag{5.843}$$

$$\delta \hat{\alpha} = \Omega \cdot \hat{\alpha}, \tag{5.844}$$

where Ω is a constant antisymmetric second rank tensor. Thus, the particles' coordinates, R, U, and $\hat{\alpha}$, transform like vectors undergoing an infinitesimal rotation. The fields rotate according to the prescription

$$\delta \alpha = -(\Omega \cdot x) \cdot \vec{\nabla}\, \alpha \tag{5.845}$$

$$\delta \beta = -(\Omega \cdot x) \cdot \vec{\nabla}\, \beta \tag{5.846}$$

$$\delta \kappa = -(\Omega \cdot x) \cdot \vec{\nabla}\, \kappa \tag{5.847}$$

$$\delta \sigma = -(\Omega \cdot x) \cdot \vec{\nabla}\, \sigma, \tag{5.848}$$

so

$$\begin{aligned} \delta A_0 &= \delta(\alpha\, \vec{\nabla}\, \beta) \\ &= \delta\alpha\, \vec{\nabla}\, \beta + \alpha\, \vec{\nabla}\, \delta\beta \end{aligned}$$

$$
\begin{aligned}
&= -(\Omega \cdot x)\cdot \vec{\nabla}\, \alpha\, \vec{\nabla}\, \beta + \alpha\, \vec{\nabla}\, [-(\Omega \cdot x)\cdot \vec{\nabla}\, \beta] \\
&= -(\Omega \cdot x)\cdot (\vec{\nabla}\, \alpha\, \vec{\nabla}\, \beta + \alpha\, \vec{\nabla}\vec{\nabla}\, \beta) + \alpha\Omega \cdot \vec{\nabla}\, \beta \\
&= -(\Omega \cdot x)\cdot \vec{\nabla}\, (\alpha\, \vec{\nabla}\, \beta) + \Omega \cdot (\alpha\, \vec{\nabla}\, \beta) \\
&= -(\Omega \cdot x)\cdot \vec{\nabla}\, A_0 + \Omega \cdot A_0, \qquad (5.849)
\end{aligned}
$$

and similarly

$$\delta A_1 = -(\Omega \cdot x)\cdot \vec{\nabla}\, A_1 + \Omega \cdot A_1. \qquad (5.850)$$

Finally note that the Lagrangian densities transform like scalar fields so

$$\delta \mathcal{L}_{gc} = -(\Omega \cdot x)\cdot \vec{\nabla}\, \mathcal{L}_{gc} \qquad (5.851)$$

$$\delta \mathcal{L}_m = -(\Omega \cdot x)\cdot \vec{\nabla}\, \mathcal{L}_m. \qquad (5.852)$$

Inserting these into Eq. (5.826), a short manipulation yields

$$
\begin{aligned}
\vec{\nabla} \cdot \Big[T \cdot \Omega \cdot x + \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \dot{R}(\eta,\tau) \Gamma_{gc\hat{\alpha}}(Z(\eta,\tau); A_i, F_i) \cdot \Omega \cdot \hat{\alpha} \Big] \\
= 0, \qquad (5.853)
\end{aligned}
$$

where T is the stress-energy tensor given by Eq. (5.840). Since Ω is the generator of an *arbitrary* rotation, this becomes

$$\vec{\nabla} \cdot (L + S) = 0. \qquad (5.854)$$

Here we have defined the third rank *orbital angular momentum tensor*

$$L^{\alpha\beta\gamma} \equiv T^{\alpha\beta} x^\gamma - T^{\alpha\gamma} x^\beta, \qquad (5.855)$$

and the third rank *spin angular momentum tensor*

$$
\begin{aligned}
S^{\alpha\beta\gamma} &\equiv \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \\
&\qquad \dot{R}^\alpha(\eta,\tau)[\Gamma^\beta_{\hat{\alpha}}(Z(\eta,\tau); A_i, F_i)\hat{\alpha}^\gamma - \Gamma^\gamma_{\hat{\alpha}}(Z(\eta,\tau); A_i, F_i)\hat{\alpha}^\beta] \\
&= \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \frac{\epsilon\mu(\eta,\tau)}{\lambda_B(R(\eta,\tau))} \dot{R}^\alpha(\eta,\tau) F_0^{\beta\gamma}(R(\eta,\tau))
\end{aligned}
$$

$$= \int dR \int dU \int d\mu \int d\Theta f_{10}(R,U,\mu,\Theta)\delta^4(x-R)\frac{\epsilon\mu}{\lambda_B}\dot{R}^\alpha(R,U,\mu)F_0^{\beta\gamma}$$
$$= \int dR \int dU \int d\mu \overline{f}_9(R,U,\mu)\delta^4(x-R)\frac{\epsilon\mu}{\lambda_B}\dot{R}^\alpha(R,U,\mu)F_0^{\beta\gamma}. \quad (5.856)$$

Eq. (5.854) expresses conservation of angular momentum in the guiding-center plasma.

We pause to interpret our result for the guiding-center spin, Eq. (5.856). In a preferred frame, $F_0^{\beta\gamma} = 0$ if either $\beta = 0$ or $\gamma = 0$, so we need consider only those components of $S^{\alpha\beta\gamma}$ for which neither β nor γ is zero, as all the rest vanish. Using Eq. (3.314) for F_0 in a preferred frame, we quickly find that

$$S^{\alpha ij} = \int dR \int dU \int d\mu \overline{f}_9(R,U,\mu)\delta^4(x-R)\epsilon\mu\dot{R}^\alpha(R,U,\mu)\epsilon^{ijk}\mathbf{b}_k, \quad (5.857)$$

where Latin indices run from one to three, as usual. Now in three dimensions one must take the three-dual of the angular momentum tensor to get the angular momentum vector. We can now do this for the last two indices of $S^{\alpha ij}$. The first index is present because the relativistically covariant object is not the angular momentum itself, but rather its four flux. Taking the three dual, we find

$$\frac{1}{2}\epsilon_{kij}S^{\alpha ij} = \epsilon \int dR \int dU \int d\mu \overline{f}_9(R,U,\mu)\delta^4(x-R)\dot{R}^\alpha(R,U,\mu)\mu\mathbf{b}_k. \quad (5.858)$$

Thus, to lowest order in ϵ, when $\alpha = 0$ we get c times the spin density, which is the sum over the distribution of guiding centers of the vector with magnitude $\gamma_v\mu$ that points in the direction of $\mathbf{b}$. Thus the spin angular momentum for a single guiding center in a preferred frame may be thought of as having magnitude $\gamma_v\mu$ and pointing in the direction of the magnetic field. For $\alpha = l \neq 0$, it is clear that we get the flux of this quantity, as the integrand has an additonal factor of $v_\parallel^l$ (to lowest order). This makes plausible our interpretation of S as the spin.

Note that

$$L^{\alpha\beta\gamma}{}_{,\alpha} = (T^{\alpha\beta}x^\gamma - T^{\alpha\gamma}x^\beta)_{,\alpha}$$
$$= T^{\alpha\beta}{}_{,\alpha}x^\gamma - T^{\alpha\gamma}{}_{,\alpha}x^\beta + T^{\gamma\beta} - T^{\beta\gamma}$$

$$= T^{\gamma\beta} - T^{\beta\gamma}, \tag{5.859}$$

where we have used Eq. (5.839). Using this result, we can write the angular momentum conservation law in the following form:

$$T - T^T + \vec{\nabla} \cdot S = 0, \tag{5.860}$$

where the superscripted T means "transpose." Note that the antisymmetric part of the stress-energy tensor is equal to the divergence of the spin tensor.

5.5 The Guiding-center Plasma in the Presence of an Eikonal Wave Field

5.5.1 Constructing the System Action

We are now ready to extend the above analysis to the situation for which the plasma is bathed in an eikonal wave field. The full four potential is now

$$A(x) = A_0(x) + \epsilon A_1(x) + \lambda A_w(x), \tag{5.861}$$

where the eikonal wave four potential

$$A_w(x) = \tilde{A}(x) \exp\left(\frac{i}{\epsilon}\psi(x)\right) + \text{c.c.} \tag{5.862}$$

was introduced back in Eq. (4.618) of Section 4.2. The corresponding field is then

$$F(x) = F_0(x) + \epsilon F_1(x) + \lambda F_w(x), \tag{5.863}$$

where

$$F_w(x) = \frac{1}{\epsilon}\tilde{F}(x) \exp\left(\frac{i}{\epsilon}\psi(x)\right) + \text{c.c.} \tag{5.864}$$

and

$$\tilde{F}(x) = i(\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k}) + \epsilon(\vec{\nabla}\,\tilde{A} - \tilde{A}\,\overleftarrow{\nabla}) \tag{5.865}$$

(the $\mathcal{O}(\epsilon)$ term in $\tilde{F}$ is usually neglected in the eikonal approximation). Note that F_0 and F_1 are slowly varying background fields, while F_w is the rapidly varying wave field. We must now construct the system action for a plasma of guiding/oscillation centers immersed in this field. The presence of the wave field has two effects on the system action: It means that the Hamiltonian must now include the ponderomotive contribution, K_2, and it means that the Maxwell action must now include the wave field.

We first consider the effect on the Maxwell action. We form $-F : F/16\pi$, and note that it contains the product of the slowly varying terms, the product of the rapidly varying terms, and cross terms. The cross terms are oscillatory and vanish upon integration over x. The remaining Maxwell action is then

$$S_m = (S_m)_0 + \lambda^2 \tilde{S}_m, \tag{5.866}$$

where $(S_m)_0$ is the functional form of the Maxwell action with no wave present (given by Eqs. (5.771) and (5.773)), and

$$\tilde{S}_m = -\frac{1}{8\pi} \int d^4x \tilde{F}^* : \tilde{F} \tag{5.867}$$

is the contribution due to the wave. Thus the *effective* (averaged) Lagrangian density is

$$\mathcal{L}_m = (\mathcal{L}_m)_0 + \lambda^2 \tilde{\mathcal{L}}_m, \tag{5.868}$$

where $(\mathcal{L}_m)_0$ is the functional form of the Lagrangian density with no wave present (given by Eq. (5.773)), and

$$\tilde{\mathcal{L}}_m = -\frac{1}{8\pi} \tilde{F}^* : \tilde{F} \tag{5.869}$$

is the contribution due to the wave. Note that $\tilde{\mathcal{L}}_m$ is quadratic in the field amplitude.

We now consider the modification of the action due to the presence of the ponderomotive Hamiltonian. Replacing H by $H + \lambda^2 K_2$ in Eq. (5.768), we see

that

$$S_{gc} = (S_{gc})_0 + \lambda^2 \tilde{S}_{gc}, \tag{5.870}$$

where $(S_{gc})_0$ is the functional form of the guiding-center action with no wave present, and

$$\tilde{S}_{gc} = -\int d^4x \int dN(\eta) \int d\tau \delta^4(x - R(\eta, \tau)) K_2 \tag{5.871}$$

is the contribution due to the wave. Also note that the Lagrange multipliers are altered by the introduction of K_2 (recall that the Lagrange multipliers depend on the Hamiltonian). Thus $\lambda_\nu = (\lambda_\nu)_0 + \tilde{\lambda}_\nu$, where

$$\tilde{\lambda}_\nu = \xi_\nu \cdot \frac{\partial K_2}{\partial Z}, \tag{5.872}$$

and where the vectors ξ_ν were given in Eq. (3.613) at the end of Chapter 3.

Now K_2 can be expressed as a real function of the wave field amplitude, $\tilde{F}$, thanks to its manifest gauge invariance. Specifically, examination of Eq. (4.719) shows that it is a real quadratic form in the wave field amplitude. Thus it can be written

$$K_2(Z; F_i, \tilde{F}, \mathbf{k}) = \frac{1}{2} \tilde{F}^*_{\alpha\beta} \mathcal{K}^{\alpha\beta\xi\eta}(Z; F_0, \mathbf{k}) \tilde{F}_{\xi\eta}, \tag{5.873}$$

where the antisymmetry of the field tensor imparts the following symmetry properties to $\mathcal{K}$:

$$\mathcal{K}^{\alpha\beta\xi\eta} = -\mathcal{K}^{\beta\alpha\xi\eta} = \mathcal{K}^{\beta\alpha\eta\xi} = -\mathcal{K}^{\alpha\beta\eta\xi}, \tag{5.874}$$

and the reality of K_2 implies

$$\mathcal{K}^{\alpha\beta\xi\eta} = (\mathcal{K}^{\xi\eta\alpha\beta})^*. \tag{5.875}$$

It is clear that a kernel, $\mathcal{K}$, with the above properties is defined implicitly by Eq. (4.719). Thus we can write

$$\tilde{S}_{gc} = -\frac{1}{2} \int d^4x \int dN(\eta) \int d\tau \delta^4(x - R(\eta, \tau)) \tilde{F}^*_{\alpha\beta} \mathcal{K}^{\alpha\beta\xi\eta}(Z; F_i, \mathbf{k}) \tilde{F}_{\xi\eta}. \tag{5.876}$$

If we now define the fourth rank *generalized susceptibility tensor*

$$\chi^{\alpha\beta\xi\eta}(x,[Z,F_i,\mathbf{k}]) \equiv \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \mathcal{K}^{\alpha\beta\xi\eta}(Z;F_i,\mathbf{k}), \quad (5.877)$$

(note that this differs from the more conventional definition of susceptibility by a minus sign) then we can put this in still more compact form,

$$\tilde{S}_{gc} = -\frac{1}{2} \int d^4x \tilde{F}^* : \chi(x,[Z,F_i,\mathbf{k}]) : \tilde{F}. \quad (5.878)$$

Alternatively, we could write K_2 as a quadratic form in the wave potential amplitude. Using $\tilde{F} = i(\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k})$, we find

$$K_2(Z;F_i,\tilde{A},\mathbf{k}) = 2\tilde{A}^*_\alpha \mathcal{K}^{\alpha\xi} \tilde{A}_\xi, \quad (5.879)$$

where the kernel

$$\mathcal{K}^{\alpha\xi} \equiv \mathrm{k}_\beta \mathrm{k}_\eta \mathcal{K}^{\alpha\beta\xi\eta} \quad (5.880)$$

is a second rank tensor. Note that we denote it by the same symbol ($\mathcal{K}$) that we use for the fourth rank kernel; which is meant should be clear from either the context or the number of indices adorning it. The guiding-center action is then

$$\tilde{S}_{gc} = -2 \int d^4x \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \tilde{A}^*_\alpha \mathcal{K}^{\alpha\xi}(Z;F_i,\mathbf{k}) \tilde{A}_\xi. \quad (5.881)$$

We can then define the second rank *susceptibility tensor*

$$\begin{aligned} \chi^{\alpha\xi}(x,[Z,F_i,\mathbf{k}]) &\equiv 2 \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \mathcal{K}^{\alpha\xi}(Z;F_i,\mathbf{k}) \\ &= 2\mathrm{k}_\beta \mathrm{k}_\eta \chi^{\alpha\beta\xi\eta}, \end{aligned} \quad (5.882)$$

so that we may write

$$\tilde{S}_{gc} = -\int d^4x \tilde{A}^* \cdot \chi \cdot \tilde{A}. \quad (5.883)$$

Once again note that we have used the same symbol to denote the fourth order and second order versions of the susceptibility.

The guiding-center Lagrangian density is then clearly

$$\mathcal{L}_{gc}(x) = (\mathcal{L}_{gc})_0(x) + \lambda^2 \tilde{\mathcal{L}}_{gc}(x), \quad (5.884)$$

where $(\mathcal{L}_{gc})_0(x)$ is the functional form of the Lagrangian density when no wave is present, and

$$\tilde{\mathcal{L}}_{gc}(x) = -\frac{1}{2}\tilde{F}^* : \chi(x, [Z, F_i, \mathbf{k}]) : \tilde{F} \tag{5.885}$$

is the contribution due to the wave.

The *total* action is thus

$$S = (S)_0 + \lambda^2 \tilde{S}, \tag{5.886}$$

where

$$(S)_0 \equiv (S_m)_0 + (S_{gc})_0 \tag{5.887}$$

and

$$\tilde{S} = \tilde{S}_m + \tilde{S}_{gc} = -\frac{1}{8\pi}\int d^4x \tilde{F}^* : \varepsilon(x, [Z, F_i, \mathbf{k}]) : \tilde{F}, \tag{5.888}$$

and where in turn we have defined the fourth rank *generalized dielectric tensor*

$$\varepsilon^{\alpha\beta}{}_{\gamma\xi} \equiv \delta^{\alpha}{}_{\gamma}\delta^{\beta}{}_{\xi} + 4\pi\chi^{\alpha\beta}{}_{\gamma\xi}. \tag{5.889}$$

Alternatively, in terms of the wave potential amplitude, we have

$$\tilde{S} = -\frac{1}{4\pi}\int d^4x \tilde{A}^* \cdot \mathcal{D}(x, [Z, F_i, \mathbf{k}]) \cdot \tilde{A}, \tag{5.890}$$

where we have defined the second rank *dispersion tensor*

$$\mathcal{D}^{\alpha}{}_{\xi} \equiv \mathrm{k}^2\delta^{\alpha}{}_{\xi} - \mathrm{k}^{\alpha}\mathrm{k}_{\xi} + 4\pi\chi^{\alpha}{}_{\xi}. \tag{5.891}$$

Similarly, the *total* Lagrangian density is thus

$$\mathcal{L} = (\mathcal{L})_0 + \lambda^2\tilde{\mathcal{L}}, \tag{5.892}$$

where

$$(\mathcal{L})_0 \equiv (\mathcal{L}_m)_0 + (\mathcal{L}_{gc})_0 \tag{5.893}$$

and

$$\begin{aligned}\tilde{\mathcal{L}} &= \tilde{\mathcal{L}}_m + \tilde{\mathcal{L}}_{gc} \\ &= -\frac{1}{8\pi}\tilde{F}^* : \varepsilon(x, [Z, F_i, \mathbf{k}]) : \tilde{F} \\ &= -\frac{1}{4\pi}\tilde{A}^* : \mathcal{D}(x, [Z, F_i, \mathbf{k}]) : \tilde{A}\end{aligned} \tag{5.894}$$

The above action must be varied with respect to the particle coordinates and the fields as before, but now we must also vary it with respect to the wave fields, $\tilde{A}(x)$ and $\psi(x)$. Note that the action depends on $\tilde{A}$ only through its dependence on $\tilde{F}$, thanks to the manifest gauge invariance of K_2; variation with respect to $\tilde{A}$ will yield the dispersion relation for linear plasma waves. Note also that the action depends on ψ only through its dependence on $\mathbf{k} = \vec{\nabla}\,\psi$, thanks to the averaging out of oscillating terms; thus ψ is an ignorable field coordinate, and variation with respect to it will yield the conservation law for wave action.

Just as we found it useful to denote the dependence of a functional on A_i and F_i separately, we shall also find it useful to denote dependence on $\tilde{A}$ and $\tilde{F}$ separately. Using Eq. (5.865), the analog of Eq. (5.783) is easily found to be

$$\left.\frac{\delta S}{\delta \tilde{A}}\right|_{\text{total}} = \frac{\delta S}{\delta \tilde{A}} - \frac{2i}{\epsilon}\mathbf{k}\cdot\left(\frac{\delta S}{\delta \tilde{F}}\right) - 2\,\vec{\nabla}\cdot\left(\frac{\delta S}{\delta \tilde{F}}\right) \tag{5.895}$$

(in the eikonal approximation, the third term on the right hand side is usually neglected). Similarly, we shall also find it useful to denote dependence on ψ and $\tilde{F}$ separately (note that $\tilde{F}$ contains $\mathbf{k}$ which is the gradient of ψ). Once again, we use Eq. (5.865) to write

$$\left.\frac{\delta S}{\delta \psi}\right|_{\text{total}} = \frac{\delta S}{\delta \psi} - \frac{2i}{\epsilon}\,\vec{\nabla}\cdot\left(\tilde{A}^*\cdot\frac{\delta S}{\delta \tilde{F}}\right). \tag{5.896}$$

These results are very helpful in deriving what follows.

5.5.2 The Vlasov Equation for Guiding/Oscillation Centers

It is straightforward to see that

$$\begin{aligned}\frac{\delta S}{\delta Z(\eta,\tau)} = &\left(\frac{\delta S}{\delta Z(\eta,\tau)}\right)_0 \\ &- \lambda^2\frac{\partial(\tilde{\lambda}_\nu C_\nu + K_2)}{\partial Z}(Z(\eta,\tau); F_i(R(\eta,\tau)), \\ &\tilde{F}(R(\eta,\tau)), \mathbf{k}(R(\eta,\tau))),\end{aligned} \tag{5.897}$$

where, as usual, we have used a subscripted 0 to denote the functional form of a quantity when no wave is present. The above result yields the correction in the equations of motion due to the presence of the ponderomotive Hamiltonian. Thus, the only modification to the kinetic equation due to the wave field is the inclusion of the ponderomotive effects of the wave field on the guiding/oscillation centers of the plasma.

5.5.3 The Field Equations

Next, we use Eq. (5.783) to take the functional derivative of S with respect to the A_i to get

$$\frac{\delta S}{\delta A_i} = \left(\frac{\delta S}{\delta A_i}\right)_0 + 2\lambda^2 \vec{\nabla} \cdot \Big[\int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \frac{1}{2} \tilde{F}^* : \frac{\partial \mathcal{K}(Z(\eta,\tau), F_i, \mathbf{k}) :}{\partial F_i} \tilde{F} \Big]. \tag{5.898}$$

Thus our field equation still follows from

$$\int d^4x (\mathcal{J}_0 \cdot \delta A_0 + \mathcal{J}_1 \cdot \delta A_1) = 0, \tag{5.899}$$

but now:

$$\mathcal{J}_i = (\mathcal{J}_i)_0 + 2\lambda^2 \vec{\nabla} \cdot \Big[\int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \frac{1}{2} \tilde{F}^* : \frac{\partial \mathcal{K}(Z(\eta,\tau), F_i, \mathbf{k}) :}{\partial F_i} \tilde{F} \Big]. \tag{5.900}$$

Note that $\mathcal{K}$ has no explicit dependence on F_1 (the only effect of F_1 is to alter the Poisson brackets), so only $\mathcal{J}_0$ is modified. This may be interpreted as a modification to the guiding-center magnetization density due to the presence of the wave field. That is, our field equations are still given by Eqs. (5.812) and (5.813), but now

$$M = (M)_0 - 2\lambda^2 \int dN(\eta) \int d\tau \delta^4(x - R(\eta,\tau)) \frac{1}{2} \tilde{F}^* : \frac{\partial \mathcal{K}(Z(\eta,\tau), F_i, \mathbf{k}) :}{\partial F_i} \tilde{F}. \tag{5.901}$$

Note that the guiding-center current density is unaffected by the presence of the wave; this is due to our neglect of resonant effects.

5.5.4 The Linear Susceptibility

We now have two additional equations of motion due to the variations with respect to $\tilde{A}$ and ψ. First we consider the variation with respect to $\tilde{A}$. We use Eq. (5.895), and in keeping with the eikonal approximation, we neglect the third term on the right. We immediately get

$$0 = \frac{\delta S}{\delta \tilde{A}} = \mathcal{P}\left(\frac{i\lambda^2}{4\pi\epsilon}\mathbf{k}\cdot\varepsilon(x,[Z,F_i,\mathbf{k}]) : \tilde{F}\right), \tag{5.902}$$

so

$$\mathcal{P}\left(\mathbf{k}\cdot\varepsilon(x,[Z,F_i,\vec{\nabla}\,\psi(x)]) : \tilde{F}(x)\right) = 0. \tag{5.903}$$

This is the eikonal equation for linear plasma waves. To see it in a somewhat more familiar form, write $\tilde{F} = i(\mathbf{k}\tilde{A} - \tilde{A}\mathbf{k})$, so after some straightforward manipulation we arrive at

$$\mathcal{P}(\mathcal{D}\cdot\tilde{A}) = 0, \tag{5.904}$$

where we have used the *dispersion tensor* defined back in Eq. (5.891),

$$\begin{aligned}\mathcal{D}^{\beta}{}_{\xi} &\equiv \mathrm{k}_\alpha \mathrm{k}^\gamma(\varepsilon^{\alpha\beta}{}_{\gamma\xi} - \varepsilon^{\alpha\beta}{}_{\xi\gamma}) \\ &= \mathrm{k}^2\delta^{\beta}{}_{\xi} - \mathrm{k}^\beta\mathrm{k}_\xi + 8\pi\mathrm{k}_\alpha\mathrm{k}^\gamma\chi^{\alpha\beta}{}_{\gamma\xi} \\ &= \mathrm{k}^2\delta^{\beta}{}_{\xi} - \mathrm{k}^\beta\mathrm{k}_\xi + 4\pi\chi^{\beta}{}_{\xi}\end{aligned} \tag{5.905}$$

The dispersion relation for linear plasma waves is found by setting the eigenvalues of the dispersion tensor equal to zero. In "three-plus-one" notation, the dispersion tensor is three by three and so it has only three eigenvalues that can be set to zero. It seems that we are finding an extra branch to the dispersion relation, and one might wonder why this should be so. By multiplying Eq. (5.905) by k_β, however, it is easy to see that $\mathbf{k}$ is a null eigenvector of $\mathcal{D}$. Thus, the extra

eigenvalue is null, so setting it equal to zero does not yield any new information. The other three roots yield the more interesting information about plasma waves.

5.5.5 Conservation of Wave Action

We next consider the equation of motion obtained by varying ψ. Using Eq. (5.896), we immediately find

$$0 = \frac{\delta S}{\delta \psi} = \vec{\nabla} \cdot \mathcal{J}, \tag{5.906}$$

where we have defined the *wave action four flux*

$$\mathcal{J} \equiv \mathcal{P}\left(\frac{i\lambda^2}{2\pi\epsilon}\tilde{A}^* \cdot \varepsilon : \tilde{F} + \frac{\lambda^2}{8\pi}\tilde{F}^* : \frac{\partial \varepsilon}{\partial \mathbf{k}} : \tilde{F}\right). \tag{5.907}$$

Our equation of motion thus expresses the conservation of this wave action.

The wave action takes on a much simpler form when written in terms of the *dispersion tensor*, defined in Eq. (5.905). We find

$$\mathcal{J} \equiv \mathcal{P}\left(\frac{\lambda^2}{8\pi}\tilde{A}^* \cdot \frac{\partial \mathcal{D}}{\partial \mathbf{k}} \cdot \tilde{A}\right). \tag{5.908}$$

Finally note that the wave action is gauge invariant, although this is not manifest in either of the two forms presented above. To prove this, we replace $\tilde{A}^*$ by $\tilde{A}^* - i\mathbf{k}\Lambda^*$ in Eq. (5.907). Using the dispersion relation, Eq. (5.903), it is easy to see that the term involving Λ vanishes, leaving $\mathcal{J}$ unchanged.

5.5.6 Applying the Noether Method

We now consider what happens to the conservation laws obtained by the Noether method when we include the effects of the wave field. In this case, Eq. (5.826) is altered in the following way:

$$\delta\mathcal{L} = (\delta\mathcal{L})_0 - \vec{\nabla} \cdot (\delta\psi\mathcal{J}) - \frac{\lambda^2}{2\pi}\vec{\nabla} \cdot [(\varepsilon : \tilde{F}) \cdot \tilde{A}^*] + \lambda^2 \vec{\nabla} \cdot (\tilde{M} \cdot \delta A_0)$$
$$+ \vec{\nabla} \cdot \left\{ \int dN(\eta) \int d\tau \delta R(\eta, \tau) \delta^4(x - R(\eta, \tau)) \right.$$

$$\frac{\lambda^2}{2}\tilde{F}^* : \mathcal{K}(Z(\eta,\tau);F_i,\mathrm{k}) : \tilde{F}\Big\}. \tag{5.909}$$

To derive this equation, we applied the variation to the full Lagrangian density for the guiding/oscillation-center plasma in the presence of the wave field. We noted that

$$\tilde{F} = \frac{i}{\epsilon}\mathbf{k}\tilde{A} + \vec{\nabla}\,\tilde{A} - (\text{transpose}), \tag{5.910}$$

so

$$\delta\tilde{F} = \frac{i}{\epsilon}\mathbf{k}\delta\tilde{A} + \frac{i}{\epsilon}(\vec{\nabla}\,\delta\psi)\tilde{A} + (\vec{\nabla}\,\delta\tilde{A}) - (\text{transpose}). \tag{5.911}$$

Finally, we used the equations of motion to simplify the result, just as we did for the case in which there was no wave field present.

Note that the second term on the right hand side of Eq. (5.910) and the third term on the right of Eq. (5.911) are usually neglected in the eikonal approximation. They are similar in this respect to the third term on the right of Eq. (5.895), and the $\mathcal{O}(\epsilon)$ terms of Eq. (5.865) (which also must be included in the analysis leading to Eq. (5.909)). Up until now, we have consistently neglected these terms in our analysis. It will turn out that they are also unneccessary in deriving the conservation law for energy-momentum, but they *are* necessary in the derivation of the conservation law for angular momentum in order to obtain the correct expression for the modification of the guiding-center spin due to the presence of the wave.

5.5.7 Conservation of Energy-Momentum

We now use the same translational variation of the system that we did in the case for which no wave was present, but now we add the variations of the wave quantities,

$$\delta\psi = -\xi\cdot\vec{\nabla}\,\psi = -\xi\cdot\mathbf{k} \tag{5.912}$$

and

$$\delta\tilde{A} = -\xi\cdot\vec{\nabla}\,\tilde{A}. \tag{5.913}$$

There are five terms on the right hand side of Eq. (5.909). The fifth term cancels the portion of $\delta\mathcal{L}_{gc} = -\xi\cdot\vec{\nabla}\,\mathcal{L}_{gc}$ (on the left hand side) that is due to K_2. The fourth term is the correction to the magnetization density due to the wave, as defined in Eq. (5.790). It will simply cause the magnetization density that appears in the conservation laws to be corrected for the presence of the wave. The third term is of the sort discussed above that may be neglected in the usual eikonal approximation. The new stuff comes from the second term, $\vec{\nabla}\cdot(\mathcal{J}\mathrm{k}\cdot\xi)$, and from the portion of $\delta\mathcal{L}_m = -\xi\cdot\vec{\nabla}\,\mathcal{L}_m$ (on the left hand side) that is due to the wave.

The new stress-energy tensor is then

$$T = (T)_0 + \lambda^2\tilde{T}, \tag{5.914}$$

where $(T)_0$ is the result with no wave field present (see Eq. (5.840)), and $\tilde{T}$ is the modification due to the wave,

$$\tilde{T} = \tilde{M}\cdot F_0 + \tilde{\mathcal{L}}_m\mathbf{1} + \mathcal{J}\mathbf{k}. \tag{5.915}$$

To recap, the first term on the right hand side above simply insures that the magnetization that appears in the stress-energy tensor is that corrected for the presence of the wave. The second term on the right hand side above similarly insures that the term $\mathcal{L}_m\mathbf{1}$ that appears in the stress-energy tensor is also corrected for the presence of the wave. The third term is the stress-energy due to the wave itself. Note that it is the tensor product of the wave action with the four wavevector. This is sensible since the wave action may be interpreted as the number flux of wave quanta times some unit of action, and the unit of action times the four wavevector is the energy-momentum per quantum.

5.5.8 Conservation of Angular Momentum

Finally, we examine the law of conservation of angular momentum. We use the same rotational variation of the system that we did in the case for which no wave

was present, but now we add the variations of the wave quantities,

$$\delta\psi = -(\Omega \cdot x)\cdot \vec{\nabla}\, \psi = -(\Omega \cdot x) \cdot \mathbf{k} \tag{5.916}$$

and

$$\delta\tilde{A} = -(\Omega \cdot x)\cdot \vec{\nabla}\, \tilde{A} + \Omega \cdot \tilde{A}. \tag{5.917}$$

Once again, we examine the five terms on the right hand side of Eq. (5.909). Now $\delta\mathcal{L} = -(\Omega\cdot x)\cdot \vec{\nabla}\, \mathcal{L} = -\vec{\nabla}\cdot[(\Omega\cdot x)\mathcal{L}]$, so once again the fifth term will cancel with the portion of $\delta\mathcal{L}_{gc}$ (on the left hand side) that is due to K_2. Similarly, it is straightforwardly shown that the fourth term causes the magnetization density that appears in the angular momentum tensor to be corrected for the presence of the wave, just as it did in the stress-energy tensor. The second term is $\vec{\nabla}$ $\cdot[\mathcal{J}\mathbf{k}\cdot\Omega\cdot x]$, and this contributes a new term in the orbital angular momentum tensor; so

$$L = (L)_0 + \lambda^2\tilde{L}, \tag{5.918}$$

where

$$\tilde{L}^{\alpha\beta\gamma} = \tilde{T}^{\alpha\beta}x^\gamma - \tilde{T}^{\alpha\gamma}x^\beta. \tag{5.919}$$

Clearly, this is the orbital angular momentum due to the wave.

This time we retain the third term on the right hand side of Eq. (5.909). It is

$$\frac{\lambda^2}{2\pi}\,\vec{\nabla}\cdot\left\{[-(\Omega\cdot x)\cdot\vec{\nabla}\,\tilde{A}^* + \Omega\cdot\tilde{A}^*]\cdot\,\varepsilon : \tilde{F}\right\}. \tag{5.920}$$

We shall still ignore the first term in square brackets, as it contains a gradient of the wave field amplitude, but we retain the second term. After some manipulation, it becomes

$$\frac{\lambda^2}{4\pi}[\Omega_{\beta\gamma}(\tilde{A}^{\gamma *}\varepsilon^{\beta\alpha}{}_{\mu\nu}\tilde{F}^{\mu\nu} - \tilde{A}^{\beta *}\varepsilon^{\gamma\alpha}{}_{\mu\nu}\tilde{F}^{\mu\nu})]_{,\alpha}. \tag{5.921}$$

From this we can identify a correction to the spin angular momentum tensor. We write

$$S = (S)_0 + \lambda^2\tilde{S}, \tag{5.922}$$

where

$$\tilde{S}^{\alpha\beta\gamma} = \tilde{A}^{\gamma *}\epsilon^{\beta\alpha}{}_{\mu\nu}\tilde{F}^{\mu\nu} - \tilde{A}^{\beta *}\epsilon^{\gamma\alpha}{}_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{5.923}$$

This is the correction to the spin angular momentum tensor of a guiding/oscillation-center plasma due to the presence of an eikonal wave field. This quantity is given by Soper [50] for oscillations in an electromagnetic field in a vacuum. He writes

$$\tilde{S}^{\alpha\beta\gamma}_{\text{vac}} = \tilde{A}^{\gamma *}\tilde{F}^{\beta\alpha} - \tilde{A}^{\beta *}\tilde{F}^{\gamma\alpha}. \tag{5.924}$$

(see his Equation (9.3.14)). If we set the susceptibility in Eq. (5.889) equal to zero, and plug the resulting vacuum dielectric into Eq. (5.923), it is clear that our result will reduce to Soper's. Thus, our result may be considered to be an extension of his result to the case of dielectric media.

The lack of gauge invariance of our result for $\tilde{S}$ is disturbing and will be discussed further in Chapter 6.

Chapter 6

Questions for Future Study

In this chapter, we discuss some questions raised by this study that could be topics for future research. These are in no particular order.

- The neglect of resonant effects is probably the most glaring omission of this thesis, and probably that most likely to limit its utility. There are several schools of thought on how to deal with resonant effects, but they break down into two major categories:

  First, there are attempts to simply "patch up" the nonresonant treatment: For example, since our nonresonant treatment has successfully given us the hermitian part of the susceptibility tensor, we could use the Kramers-Kronig relations to get the antihermitian part. Alternatively, we could simply dictate that all resonant denominators are to be treated according to the Landau prescription. These methods, while successful in describing resonant particle effects on plasma waves, fall far short of a unified description of the effects of resonant particles. Furthermore, there is something aesthetically displeasing about tricks of this sort.

Second, there are attempts to go back and redo the single particle analyses to include resonant effects. The general idea is that we first went astray when we said that we could transform away the first order part of the action due to the eikonal wave field. While we can certainly do this far away from the resonant regions of phase space, we certainly cannot do this at (or even near) the resonance itself. So we should go back and retain the first order part of the action in the region of phase space near the resonance. Like the first technique, this approach explains certain things nicely, but falls short of a unified description of resonant particles. For example, the first order action that we retain will depend on the four potential of the wave, and this will yield a modification to the current density of a guiding-center plasma that is immersed in a wave field; this is the *current drive* due to a wave field that tokamak researchers study. On the other hand, a good description of how this residual piece of the first order action gives rise to Landau damping does not seem to exist. Furthermore, there is a great deal of arbitrariness connected with how to decide just how much of this first order action to keep. One approach uses "window functions" of some characteristic width, but there is a great deal of freedom in just how these window functions should look (square windows, gaussian windows, etc.); Dewar [51] gives a variational principle for determining optimal window shape, but then we have to worry about just what we mean by "optimal." There is also a great deal of freedom in choosing the width of such windows. If we try to transform away the first order action too close to the resonance, problems develop due to the presence of the trapped particles, and the transformation ceases to be a near-identity diffeomorphism. Unfortunately, it is hard to quantify what we mean by "too close" in this regard. Perhaps the window width should itself be treated as a dynamical variable whose dynamics are given by some variational principle (like that of Dewar); this might be a

useful tool for the study of "resonance broadening" effects, where the width of the resonant region varies in time.

- Pursuing the oscillation-center Lie transforms to higher order is a natural and obvious extension of this thesis. In this way, one could study induced scattering and three-wave phenomena. Past attempts to study these have either not used systematic perturbation theory (e.g. Lie transforms), or have used Hamiltonian methods without manifest gauge invariance. This thesis should provide the tools needed to combine the desiderata of systematic perturbation theory and manifest gauge invariance. Central to this effort has been the use of the homotopy formula, and the introduction of the pair of special functions, Q_ℓ and R_ℓ.

  It is interesting to note that this same program could have been carried out for the nonrelativistic problem. One must simply take the perturbation to the action due to the wave (for which there now would be both a vector and a scalar potential), and apply to it the guiding-center Lie transform, using the homotopy formula in the same way that we did here.

- The inclusion of dissipative effects (collisions, correlations, etc.) would be an important generalization of the work presented here. This is undoubtedly related to the problems associated with the inclusion of resonant effects. A unified treatment of correlations would yield the appropriate collision operator in the kinetic equation, and modify the energy-momentum conservation law to describe the flow of energy into heat.

  One way to approach this subject might be through the extended use of projection operators. We employed this technique in Chapter 5 to show that it was possible for energy-momentum and angular momentum to flow from one relevent region of Fourier space to another irrelevent one, and thereby to effectively appear as a source term in the conservation laws. We did not

pursue this idea of partitioning Fourier space into one zone for background fields, one zone for wave fields, and one zone for effects of collisions (for example, we never introduced a second projection operator for the wave fields, or a third one for fields arising in collisions). This approach may prove useful, but it quickly leads to great complication in the procedure, and it is not clear how it might give rise to collision operators, etc.

- When we applied the Noether method to the action to obtain the guiding-center spin angular momentum, we used the version of the action that was both boostgauge and gyrogauge invariant. There is a good reason why we did this. Other versions contain the quantity $\mathcal{R}$ that was introduced back in Chapter 3. If we had tried to apply Noether's theorem to an action containing $\mathcal{R}$, we would at some point have been faced with the question of how to vary $\mathcal{R}$ with respect to the four potential. It seems that $\mathcal{R}$ is not independent of the four potential since it was defined in terms of the unit vectors, $\hat{e}_\alpha$, and these, in turn, depend upon the background field.

  We dodged the issue by going to the boostgauge and gyrogauge invariant coordinates for which $\mathcal{R}$ does not appear in the action, but it is interesting to contemplate the alternatives. If we were to simply ignore this term, we would not get guiding-center spin, and that would be unacceptable. Though we had to go to higher order to find this term in our first derivation of the guiding-center action, it has the same order as the $\mu d\theta$ term which is obviously critically important. Indeed, now that we have the benefit of hindsight, we see that we could have avoided the higher order guiding-center Lie transform altogether by examining the action at classical order and asking what we would have to add to it to make the $\mu d\theta$ term gyrogauge invariant. The answer would have been $-\mu\mathcal{R}\cdot dR$, and this was really the only important term we found at higher order. Thus, the clever application

of a gauge invariance requirement can save one from going to higher order in a perturbation calculation!

So, since we can't ignore this term, how else could we have dealt with it? There are a couple of possible avenues of approach. First, recall the well known result that the stress-energy tensor is given by the derivative of the Lagrangian density with respect to the metric tensor (this is true at least for spinless systems). There seems to be an analogous theorem (or, at least, a conjecture) enunciated by Hehl [52], that the spin angular momentum tensor is the derivative of the Lagrangian density with respect to torsion. Torsion is the result of an asymmetric affine connection, and the affine connection that we had to introduce in Section 3.11 to explain the $\mathcal{R} \cdot \dot{R}$ term in $\dot{\Theta}$ is indeed asymmetric. Now it is not clear to me that $\mathcal{R}$ is a torsion, but these remarks do make it clear that $\mathcal{R}$ has at least something to do with torsion. In any event, $\mathcal{R}$ appears in our guiding-center action with a μ in front of it, so it is possible that we could apply the above theorem (conjecture?) and derive guiding-center spin directly (without recourse to Noether's theorem). I suspect that, if this were possible, it would be of more interest to researchers in quantum gravity (which is the community to whom reference [52] was aimed) than it would be to researchers in plasma physics. It may be that guiding-center motion provides a unique classical forum within which this topic of current research in the field of quantum gravity may be applied, tested, and better understood.

Another possible approach to the spin problem is yet more speculative. It is suggested by the *minimal coupling* idea of gauge field theory. Recall that $\mathcal{R}$ is the gauge potential associated with the gyrogauge group. In Section 3.11, we even went one step further and derived the corresponding gauge field, $\mathcal{N}$. Using the techniques of gauge field theory, it might be possible to use $\mathcal{R}$ to define a gauge covariant derivative. We could then add something like

$\mathcal{N} : \mathcal{N}$ to the Lagrangian density, and treat A and $\mathcal{R}$ as *independent* gauge fields. Though these ideas are suggested by the analogy with gauge field theories, they would all have to be rigorously justified. Furthermore, it is not obvious how guiding-center spin would arise from these considerations.

- Another mystery that should be mentioned is the apparant lack of gauge invariance of the wave modification to guiding-center spin. Our result is clearly the extension to dielectric media of Soper's result for the vacuum [50]. The lack of gauge invariance did not seem to bother him, except for a cryptic footnote that indicates that the result *is* invariant with respect to a certain subgroup of the full gauge group. One possible explanation might be that the *division* of angular momentum into orbital and spin contributions is not a gauge-invariant division. If this were the case, however, one would expect that *neither* the orbital nor the spin angular momentum should be gauge invariant by itself, but that their sum should be gauge invariant. Alas, the orbital angular momentum seems to be gauge invariant all by itself, so the issue remains a mystery.

- It would be nice to find a Hamiltonian field theoretical formulation of the kinetic and field equations for the guiding-center and the guiding/oscillation-center plasma. Manifestly covariant Hamiltonian field theories are, however, tricky to formulate. We cannot give preference to the time variable, and the proper time is not uniquely defined (every particle in the system has its own proper time). There may be ways of getting around this difficulty by generalizing the form of Hamiltonian equations of motion for such systems. If this could be done, it might be possible to use the energy-casimir method to study plasma stability to nonlinear perturbations.

- We have developed conservation laws for energy-momentum and angular momentum for the guiding/oscillation-center plasma. In most studies of

plasma dynamics, use is made of energy conservation, but not of momentum or angular momentum conservation (of course, in a covariant relativistic treatment energy and momentum are inseparable). It is possible that these conserved quantities could play a far greater role in the study of, say, plasma stability theory than they have until now. For example, the Lyapunov method for assessing stability rests heavily on the discovery of conserved quantities. Just how to go about doing this is not immediately clear.

Bibliography

[1] Dewar, R. L., *J. Phys. A: Math. Gen.*, **9**:2043 (1976).

[2] Johnston, S., Kaufman, A. N., *J. Plasma Physics*, **22**:105 (1979).

[3] Johnston, S., *Phys. Fluids*, **19** (1976).

[4] Cary, J. R., "Nonlinear Wave Evolution in Vlasov Plasma: A Lie-Transform Analysis," PhD Dissertation, University of California at Berkeley, University Microfilms International number 80-14628, or Lawrence Berkeley Laboratory Report LBL-8185 (August, 1979).

[5] Cary, J. R., Kaufman, A. N., *Phys. Fluids*, **24** (1981).

[6] Littlejohn, R. G., "Hamiltonian Theory of Guiding-Center Motion," PhD Dissertation, University of California at Berkeley, University Microfilms International number 80-29478, or Lawrence Berkeley Laboratory Report LBL-12942 (May, 1980).

[7] Littlejohn, R. G., *J. Plasma Physics*, **29**:111-125 (1983).

[8] Grebogi, C., Littlejohn, R. G., *Phys. Fluids*, **27**:1996 (1984).

[9] Cary, J. R., Newberger, B. S., *Phys. Fluids*, **28**:423 (1985).

[10] Dubin, D. H. E., Krommes, J. A., Oberman, C., Lee, W. W., *Phys. Fluids*, **26**:3524 (1983).

[11] Kaufman, A. N., Boghosian, B. M., in *Contemporary Mathematics,* American Mathematical Society, Providence, R.I., **28**, *Fluids and Plasmas: Geometry and Dynamics,* edited by J. E. Marsden (1984).

[12] Similon, P. L., *Phys. Lett.*, **112**A:33 (1985).

[13] Fradkin, D. M., *J. Phys. A: Math. Gen.*, **11**:1069-1086 (1978).

[14] Dumais, J.-F., *Am. J. Phys.*, **53**:264 (1985).

[15] Schutz, B. F., *Geometrical Methods of Mathematical Physics*, Cambridge University Press (1980).

[16] Edelen, D. G. B., *Applied Exterior Calculus*, John Wiley and Sons (1985).

[17] Singer, I. M. and Thorpe J. A., *Lecture Notes on Elementary Topology and Geometry*, Springer-Verlag (1967).

[18] Burke, W. L., *Applied Differential Geometry*, Cambridge University Press (1985).

[19] Flanders, H., *Differential Forms*, Academic Press, New York (1963).

[20] Misner, C. W., Thorne, K. S., Wheeler, J. A., *Gravitation*, W. H. Freeman and Company (1973).

[21] Chandrasekhar, S., *The Mathematical Theory of Black Holes*, Oxford University Press, New York (1983).

[22] Littlejohn, R. G., *J. Math. Phys.* **20**:2445 (1979).

[23] Marsden, J. E., Weinstein, A., Ratiu, T., Schmid R. and Spencer, R. G., in the *Proceedings of the IUTAM Symposium on Modern Developments in Analytical Mechanics,* Torino, Italy (June 7-11, 1982).

[24] Goldstein, H., *Classical Mechanics*, second edition, Addison-Wesley, Chapters 4 and 5 (1980).

[25] Arnold, V. I., *Ann. Inst. Fourier, Grenoble*, **16**:319 (1966).

[26] Low, F. E., *Proc. Roy. Soc. A*, **248**:282 (1958).

[27] Iwinski, Z. R. and Turski, L. A., *Lett. in Appl. and Eng. Sci.*, **4**:179 (1976).

[28] Morrison P. J., *Phys. Lett.*, **80A**:383 (1980). Footnote 1 credits Kaufman.

[29] Gibbons, J., *Physica*, **3D**:503 (1981).

[30] Kaufman, A. N. and Dewar, R. L., in *Contemporary Mathematics,* American Mathematical Society, Providence, R.I., **28**, *Fluids and Plasmas: Geometry and Dynamics*, edited by J. E. Marsden (1984).

[31] Littlejohn, R. G., in *Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems,* edited by M. Tabor and Y. M. Treve, American Institute of Physics Conference Proceedings Number 88 (New York, 1982).

[32] Weinstein, A., in *Contemporary Mathematics,* American Mathematical Society, Providence, R.I., **28**, *Fluids and Plasmas: Geometry and Dynamics,* edited by J. E. Marsden (1984).

[33] Littlejohn, R. G., *J. Math. Phys.* **23**:742 (1982).

[34] Littlejohn, R. G. and Cary, J. R., *Annals of Physics*, **151**:1-34 (1983).

[35] Dirac, P. A. M., *Lectures on Quantum Mechanics*, Belfer Graduate School of Science, Yeshiva University, New York (1964).

[36] Deprit, A., *Celestial Mechanics*, **1**:12-30 (1969).

[37] Dragt, A. J. and Finn, J. M., *J. Math. Phys.*, **17**:2215-2227 (1976).

[38] Cary, J. R.,*Phys. Rep.*, **79**:131 (1981).

[39] Kolmogorov, A. N., *Dokl. Akad. Nauk. SSSR*, **98**:527 (1954).

[40] Hori, G., *Pub. Astron. Soc. Japan*, **18**:287 (1966).

[41] Littlejohn, R. G., in *Contemporary Mathematics*, American Mathematical Society, Providence, R. I., **28**:151-167, *Fluids and Plasmas: Geometry and Dynamics*, edited by J. E. Marsden (1984).

[42] Hagan, W. K. and Frieman E. A., *Phys. Fluids*, **28**:2641-2643 (1985).

[43] Northrop, T. G., *The Adiabatic Motion of Charged Particles*, Interscience, New York (1963).

[44] Jackson, J. D., *Classical Electrodynamics*, second edition, John Wiley and Sons, pp. 503-578 (1975).

[45] Kruskal, M., in *Plasma Physics*, I.A.E.A., Vienna, pp. 67-102 (1965).

[46] Kruskal, M., in *Mathematical Models in the Physical Sciences*, Eaglewood Cliffs, N.J., edited by S. Drobot, pp. 17-48 (1963).

[47] Kaufman, A. N., *Phys. Fluids*, **3**:610 (1960).

[48] Vandervoort, P., *Ann. Phys.*, **10**:401 (1960).

[49] Achterberg, A., *J. Plasma Physics*, **35**:257 (1986).

[50] Soper, D. E., *Classical Field Theory*, John Wiley and Sons (1978).

[51] Dewar, R. L., *Physica*, **17**D:37 (1985).

[52] Hehl, F. W., *Spin and Torsion in General Relativity: I. Foundations*, *Gen. Rel. and Grav.*, 4:333 (1973).

Appendix A

Glossary of Notation

In this appendix, we list all the important symbols used in this thesis, giving the number of the equation where they were first used (if appropriate) and a brief description (if appropriate).

SYMBOL	EQUATION	DESCRIPTION
$\alpha(x)$	(5.794)	Clebsch potential for field
$\hat{\alpha}$	(3.591)	Gyrogauge-invariant coordinatization of gyroangle
β	(3.341)	Angular hyperbolic polar coordinate for parallel part of particle four velocity
$\beta(x)$	(5.794)	Clebsch potential for field
$\boldsymbol{\beta}_E$	(3.330)	$\mathbf{E} \times \mathbf{B}/B^2$
$\boldsymbol{\beta}_v$	(3.357)	$\mathbf{v}/c$
Γ_{gc}	(3.489)	Guiding-center action one form
γ		Action one-form
γ_v		Relativistic gamma factor: $\gamma_v = 1/\sqrt{1-\beta_v^2}$.
$\delta^{\alpha}{}_{\beta}$		Kronecker delta

ϵ		Guiding-center expansion parameter
$\epsilon^{\nu_1 \cdots \nu_n}$		Levi-Civita tensor in n dimensions
$\varepsilon^{\alpha\beta}{}_{\gamma\xi}$	(5.889)	Generalized dielectric tensor
η	(5.737)	Continuous particle label
Θ	(3.485)	Angular polar coordinate for perpendicular part of guiding-center four velocity
θ	(3.341)	Angular polar coordinate for perpendicular part of particle four velocity
$\kappa(x)$	(5.795)	Clebsch potential for field
λ	(4.620)	Oscillation-center expansion parameter
λ_1	(3.287)	Lorentz scalar for electromagnetic field
λ_2	(3.288)	Lorentz pseudoscalar for electromagnetic field
λ_ν	(3.604)	Lagrange multiplier
λ_B	(3.293)	Related to eigenvalues of F
λ_E	(3.292)	Related to eigenvalues of F
μ	(3.488)	Gyromomentum
ν		Constraint label
Ξ	(3.506)	
Ξ'	(3.576)	
ξ	(5.827)	Generator of infinitesimal translation in space-time
ξ^α_ν	(3.613)	
$\sigma(x)$	(5.795)	Clebsch potential for field
τ		Proper time
Υ	(3.505)	
Υ'	(3.575)	
$\chi^{\alpha\beta}{}_{\gamma\xi}$	(5.877)	Generalized susceptibility tensor
Ψ_ℓ	(4.652)	

ψ	(4.618)	Phase of eikonal wave
Ω	(5.841)	Generator of infinitesimal rotation in spacetime
Ω_B	(3.297)	Gyrofrequency with respect to proper time
Ω_{gc}	(3.490)	Guiding-center Lagrangian two form
ω		Lagrangian two-form
A	(3.281)	Four-vector potential
A_0		Zero-order four-vector potential
A_1		First-order four-vector potential
A_w	(4.618)	Eikonal wave potential
$\bar{A}$	(4.618)	Amplitude of eikonal wave potential
$\mathbf{A}$	(3.281)	Three-vector potential
$\hat{\mathbf{a}}$	(3.346)	Member of orthonormal basis tetrad
$\mathbf{B}$	(3.281)	Magnetic field pseudovector
$\mathcal{B}$	(3.485)	Angular hyperbolic polar coordinate for parallel part of guiding-center four velocity
$\mathbf{b}$		Unit three-vector in direction of magnetic field
$\hat{\mathbf{b}}$	(3.345)	Member of orthonormal basis tetrad
C_ν	(3.592)	Constraints
c		Speed of light
$\hat{\mathbf{c}}$	(3.344)	Member of orthonormal basis tetrad
D	(5.749)	Jacobian
$\mathcal{D}^\alpha{}_\beta$	(5.905)	Dispersion Tensor
D_ℓ	(4.688)	Resonant denominator
e		Charge
$\mathbf{E}$	(3.281)	Electric field vector
F	(3.281)	Field tensor
F'	(3.495)	
F''	(3.507)	

F'''	(3.577)	
F_0		Zero-order field tensor
F_1		First-order field tensor
F_w	(5.865)	Eikonal wave field
$\tilde{F}$	(5.865)	Amplitude of eikonal wave field
$\mathcal{F}$	(3.282)	Dual field tensor
f_n	(5.737)	Pseudoscalar Eulerian particle distribution function
f_n	(5.747)	Scalar Eulerian particle distribution function
$\overline{f}_n$	(5.737)	Pseudoscalar Eulerian guiding-center distribution function
$\overline{\mathsf{f}}_n$	(5.747)	Scalar Eulerian guiding-center distribution function
$g_{\mu\nu}$		Metric tensor
G_0	(5.788)	Macroscopic field tensor for perpendicular current
G_1	(5.789)	Macroscopic field tensor for parallel current
H		Hamiltonian
H_{gc}	(3.488)	Guiding-center Hamiltonian
i		$\sqrt{-1}$
i_g	(2.60)	Interior product with respect to vector field g
J	(5.787)	Four-current density
J_{gc}		Guiding-center poisson tensor
$\mathcal{J}_0$	(5.785)	
$\mathcal{J}_1$	(5.786)	
$\mathcal{J}_\ell^\pm$	(4.653)	
$\mathbf{k}$		Wave four vector
K	(3.341)	Radial hyperbolic polar coordinate for parallel part of guiding-center four velocity

K_2	(4.669)	Ponderomotive Hamiltonian
$\mathcal{K}$	(5.873)	Kernel of ponderomotive Hamiltonian
k	(3.485)	Radial hyperbolic polar coordinate for parallel part of particle four velocity
L	(5.855)	Guiding-center orbital angular momentum tensor
$\tilde{L}$	(5.919)	Wave contribution to guiding-center orbital angular momentum tensor
L_{gc}	(3.582)	Guiding-center Lagrangian
$\mathcal{L}_g$	(2.43)	Lie derivative with respect to vector field g
$\mathcal{L}_m$	(5.773)	Lagrangian density of Maxwell field
$\tilde{\mathcal{L}}_m$	(5.869)	Lagrangian density of eikonal wave field
ℓ	(4.639)	Index for Fourier expansion in gyroangle
M_0	(5.790)	Magnetization density tensor for perpendicular current
M_1	(5.791)	Magnetization density tensor for parallel current
$\mathcal{M}$	(3.517)	Boostgauge field
m		Mass
$dN(\eta)$	(5.737)	Measure of particles with labels between η and $\eta + d\eta$
$\mathcal{N}$	(3.518)	Gyrogauge field
$P_\parallel$	(3.304)	Parallel projection operator
$P_\perp$	(3.305)	Perpendicular projection operator
$\mathcal{P}$	(5.798)	Smoothing projection operator
Q_ℓ	(4.643)	Special Function
$\mathcal{Q}$	(3.384)	
r		Particle spacetime position
R_ℓ	(4.644)	Special function
$\mathcal{R}$	(3.385)	

$\Re$		The set of real numbers
S	(5.856)	Guiding-center spin angular momentum tensor
$\tilde{S}$	(5.923)	Wave contribution to guiding-center spin angular momentum tensor
S_{gc}	(5.768)	Guiding-center action
S_m	(5.771)	Maxwell action
$\tilde{S}_m$	(5.867)	Maxwell action due to eikonal wave
T	(5.840)	Guiding-center stress-energy tensor
$\tilde{T}$	(5.915)	Wave contribution to guiding-center stress-energy tensor
$\hat{\mathbf{t}}$	(3.343)	Member of orthonormal basis tetrad
U	(3.562)	Boostgauge-invariant coordinatization of guiding-center parallel velocity
u	(3.289)	Particle four-velocity
$\mathbf{v}$		Three-velocity
W	(3.485)	Radial polar coordinate for perpendicular part of guiding-center four velocity
w	(3.341)	Radial polar coordinate for perpendicular part of particle four velocity
x		Spacetime coordinates
Z		Generic coordinates
:		Double index contraction: $A : B \equiv A_{\mu\nu} B^{\mu\nu}$.

Appendix B

Vector Spaces, Dual Spaces, Algebras, and Modules

This appendix is included to establish the set-theoretical foundations of tensor calculus and exterior algebra, as these ideas are used extensively in this thesis. It is intended to provide a review for people already familiar with these topics, and to establish notation. The reader is expected to be familiar with linear algebra and with the topology of the real numbers. If anything herein is unfamiliar, the reader is urged to consult one of the above-mentioned introductory references.

We begin with some set-theoretical notation: Given two sets, A and B, we define the *Cartesian product*, $A \times B$, to be the set of all ordered pairs, (a, b), such that $a \in A$ and $b \in B$. The symbol $\forall$ is read "for all," and the symbol $\exists$ is read "there exists." A set is said to be *partitioned* if there exist subsets such that each and every element of the set is a member of one and only one subset. A map that associates an element of a set, B, to each element of a set, A, is denoted by $A \mapsto B$.

A *relation*, R, among the elements of a set, A, is defined to be a subset of

$A \times A$; we write $R \subset A \times A$. Two elements of A, say a_1 and a_2, are then said to be *related* if $(a_1, a_2) \in R$. In this case, we may write $a_1 \sim a_2$. A relation is *reflexive* if $a \sim a$ for all $a \in A$. A relation is *symmetric* if $a \sim b$ implies $b \sim a$ for all $a, b \in A$. A relation is *transitive* if $a \sim b$ and $b \sim c$ implies $a \sim c$ for all $a, b, c \in A$. A relation that is reflexive, symmetric and transitive is called an *equivalence relation.* An equivalence relation naturally partitions a set into subsets called *equivalence classes.* Any two members of the same equivalence class are related to each other by the equivalence relation, and members of different equivalence classes are not related by the equivalence relation. For example, the equivalence relation of "similarity" partitions the set of all triangles into an infinity of equivalence classes, and the equivalence relation of "equality modulo three" partitions the set of integers into three classes. The relation "is the same height or taller than" is *not* an equivalence relation on the set of all trees, because, although it is reflexive and transitive, it is not symmetric, etc.

The set of all real numbers will be denoted by $\Re$. The set of all n-tuples of real numbers will be denoted by $\Re^n$, and the reader is assumed to have some familiarity with its usual topology. In particular, by using, say, the Euclidean norm, it is possible to define open sets as neighborhoods, and thus to have a concept of nearness, continuity, convergence, etc.

Let $\mathcal{V}$ be a set with $U, V, W, \ldots \in \mathcal{V}$, and let $a, b, c, \ldots \in \Re$. Let $+$ denote an operation that takes two elements of $\mathcal{V}$ and returns a third one; that is, $+$ is a map $\mathcal{V} \times \mathcal{V} \mapsto \mathcal{V}$. Let $\cdot$ denote an operation that takes an element of $\Re$ and an element of $\mathcal{V}$ and returns an element of $\mathcal{V}$; that is, $\cdot$ is a map $\Re \times \mathcal{V} \mapsto \mathcal{V}$. Then $\mathcal{V}$ is a *vector space* over the field of real numbers if and only if the following conditions hold:

Condition B.0.8.1 $\forall U, V, W \in \mathcal{V} : U + (V + W) = (U + V) + W.$

Condition B.0.8.2 $\forall U, V \in \mathcal{V} : U + V = V + U.$

Condition B.0.8.3 $\exists 0 \in \mathcal{V} : \forall V \in \mathcal{V} : V + 0 = V.$

Condition B.0.8.4 $\forall U \in \mathcal{V} : \exists V \in \mathcal{V} : U + V = 0.$

Condition B.0.8.5 $\forall a, b \in \Re, U \in \mathcal{V} : (ab) \cdot U = a \cdot (b \cdot U).$

Condition B.0.8.6 $\forall a, b \in \Re, U \in \mathcal{V} : (a + b) \cdot U = a \cdot U + b \cdot U.$

Condition B.0.8.7 $\forall a \in \Re, U, V \in \mathcal{V} : a \cdot (U + V) = a \cdot U + a \cdot V.$

Condition B.0.8.8 $\forall U \in \mathcal{V} : 1 \cdot U = U.$

A set of vectors, $U_1, \ldots, U_n$, is said to be *linearly independent* if and only if the only real numbers, $c_1, \ldots, c_n$, satisfying

$$c_1 \cdot U_1 + \cdots c_n \cdot U_n = 0 \tag{B.925}$$

are $c_1 = \cdots = c_n = 0$. Otherwise, the vectors are said to be *linearly dependent.* The number of elements in the largest possible set of linearly independent vectors is called the *dimension* of the vector space. If a vector space has dimension n, then any set of n linearly independent vectors constitutes a *basis* for that vector space. If $V_1, \ldots, V_n$ is a basis for $\mathcal{V}$, then any vector, U, in $\mathcal{V}$ can be expressed

$$U = a_1 \cdot V_1 + \cdots a_n \cdot V_n, \tag{B.926}$$

where the real constants, a, are uniquely determined by U, and can be computed by standard techniques of linear algebra. In this case, we say that the basis *spans* the vector space. A *vector subspace* of a vector space, $\mathcal{V}$, is a subset of $\mathcal{V}$ that is itself a vector space closed under $\cdot$ and $+$. The dimension of the vector subspace is the minimal number of basis vectors needed to span it.

Vector spaces can be finite or infinite dimensional. An example of an infinite dimensional vector space is the space of all infinitely differentiable (C^∞) real-valued functions on $\Re$. The addition and multiplication operations are then

$$(f + g)(x) = f(x) + g(x) \tag{B.927}$$

and

$$(a \cdot f)(x) = a \cdot f(x). \tag{B.928}$$

This very important space will be called $\Lambda(\Re)$. A basis for this vector space would have to contain an infinite number of elements; the theory of Fourier series provides an example of how to go about constructing and using such bases. The set of all polynomial functions of a real argument is a vector subspace of $\Lambda(\Re)$.

A *functional*, U^*, operating on a vector space, $\mathcal{V}$, is a map $\mathcal{V} \mapsto \Re$. Equivalently, we can think of functionals as objects which pair with vectors to yield real numbers. The notation for this pairing is $\langle U^*, V \rangle \in \Re$. Note that we frequently denote functionals with superscripted stars. It is possible to define operations of addition and real number multiplication on the space of functionals as follows:

$$\langle U^* + V^*, W \rangle = \langle U^*, W \rangle + \langle V^*, W \rangle \tag{B.929}$$

and

$$\langle a \cdot U^*, W \rangle = a \langle U^*, W \rangle. \tag{B.930}$$

It is readily verified that these operations make the space of all functionals operating on $\mathcal{V}$ into a vector space which we shall denote by $\mathcal{V}^*$, and which we shall call the *dual space* to the vector space, $\mathcal{V}$. Furthermore, it is also readily verified that the dimensions of $\mathcal{V}$ and $\mathcal{V}^*$ are equal. An example of this from linear algebra may be instructive: The dual space to the vector space of column vectors may be identified with the vector space of row vectors, since a row vector and a column vector pair to yield a real number under matrix multiplication.

If a vector space, $\mathcal{V}$ is endowed with a further bilinear operation that maps $\mathcal{V} \times \mathcal{V} \mapsto \mathcal{V}$, then it is called an *algebra*. Since this operation pairs vectors with other vectors, it can be written in the form $(U, V) \in \mathcal{V}$. By "bilinear," we mean

$$(a \cdot U + b \cdot V, W) = a \cdot (U, W) + b \cdot (V, W) \tag{B.931}$$

and

$$(U, a \cdot V + b \cdot W) = a \cdot (U, V) + b \cdot (U, W). \tag{B.932}$$

An algebra is *commutative* if $\forall U, V \in \mathcal{V} : (U, V) = (V, U)$. An algebra is associative if $\forall U, V, W \in \mathcal{V} : (U, (V, W)) = ((U, V), W)$. The set of real numbers, $\Re$, becomes a commutative, associative algebra when equipped with the operation of multiplication of real numbers. The space $\Lambda(\Re)$ described above is also a commutative, associative algebra if we equip it with the multiplication

$$(fg)(x) = f(x)g(x). \tag{B.933}$$

In linear algebra, the set of all n by n square matrices is a vector space of dimension n^2 with the usual definitions of matrix addition and multiplication by real numbers; it becomes an associative (but not commutative) algebra when equipped with matrix multiplication.

An algebra, $\mathcal{V}$, is called a *Lie algebra* if and only if it is anticommutative

$$\forall U, V \in \mathcal{V} : (U, V) = -(V, U), \tag{B.934}$$

and satisfies the *Jacobi identity*

$$\forall U, V, W \in \mathcal{V} : (U, (V, W)) + (V, (W, U)) + (W, (U, V)) = 0. \tag{B.935}$$

The space of vectors in $\Re^3$ becomes a Lie algebra when equipped with the usual cross product.

A vector subspace of an algebra is called a *subalgebra* if it is closed under the algebra's multiplication rule. For example, the space of all polynomial functions of a real argument is a subalgebra of $\Lambda(\Re)$. A subalgebra of a Lie algebra is called a *Lie subalgebra.*

We can generalize the concept of a vector field somewhat by relaxing the requirement that a and b in Conditions B.0.8.5 through B.0.8.8 above are real numbers. Suppose instead that they are members of any associative algebra, A.

Then Conditions B.0.8.5 through B.0.8.8 still make sense, though the number 1 that appears in Condition B.0.8.8 must be reinterpreted to refer to the identity element of the algebra, A. In this case, $\mathcal{V}$ is said to be a *module* over the algebra, A. For example, in linear algebra, the space of column vectors is a module over the above-described algebra of square matrices.

Given an algebra, $\mathcal{V}$, with subspace, $\mathcal{U}$, we say that $\mathcal{U}$ is an *ideal* of $\mathcal{V}$ if and only if $(U, V) \in \mathcal{U}$ and $(V, U) \in \mathcal{U}$ for all $U \in \mathcal{U}$, and $V \in \mathcal{V}$. For example, let $\mathcal{V}$ be the vector space of all polynomial functions of a real argument, x; Recall that this is a subalgebra of $\Lambda(\Re)$. Then, the subspace, $\mathcal{U} \subset \mathcal{V}$, of all polynomials with zeros at some particular location(s) is an ideal of $\mathcal{V}$.

Throughout this thesis, when a scalar multiplies a vector, the dot is suppressed; that is, $a \cdot V$ is written simply aV. The dot notation is used for other things. Also, boldface type is used to denote a vector, though its components in a given coordinate system are denoted by the same letter in ordinary typeface (with a superscripted index to label components).

Appendix C

Gyrofrequency Shift for Two-Dimensional Nonrelativistic Guiding-Center Motion

As a straightforward but nontrivial example of the vector Lie transform technique, we consider two-dimensional nonrelativistic guiding-center motion in a magnetic field of the form

$$\mathbf{B} = B(x,y)\hat{\mathbf{z}}, \tag{C.936}$$

and a perpendicular electric field of the form

$$\mathbf{E} = E_x(x,y)\hat{\mathbf{x}} + E_y(x,y)\hat{\mathbf{y}}. \tag{C.937}$$

To lowest order, the gyrofrequency is given by $\Omega = eB/mc$. We shall address the problem of computing the correction to this quantity due to the spatial depen-

dence of **B** and **E**.

The single-particle equations of motion are

$$\begin{aligned}\dot{x} &= u \\ \dot{y} &= v \\ \dot{u} &= \frac{e}{m}E_x + \Omega v \\ \dot{v} &= \frac{e}{m}E_y - \Omega u.\end{aligned} \tag{C.938}$$

Introduce the perpendicular velocity and the gyroangle,

$$\begin{aligned}w &= \sqrt{u^2 + v^2} \\ \theta &= \arg(-v - iu),\end{aligned} \tag{C.939}$$

so that

$$\begin{aligned}u &= -w\sin\theta \\ v &= -w\cos\theta.\end{aligned} \tag{C.940}$$

In terms of w and θ the equations of motion are found to be

$$\begin{aligned}\dot{x} &= -w\sin\theta \\ \dot{y} &= -w\cos\theta \\ \dot{w} &= -\frac{e}{m}\left(E_x\sin\theta + E_y\cos\theta\right) \\ \dot{\theta} &= \frac{1}{\epsilon}\Omega - \frac{e}{mw}\left(E_x\cos\theta - E_y\sin\theta\right)\end{aligned} \tag{C.941}$$

Here we have introduced the formal ordering parameter ϵ, and have ordered the equations of motion by the prescription $e \mapsto e/\epsilon$ and $\mathbf{E} \mapsto \epsilon\mathbf{E}$.

Though it is most useful and quite elegant to treat this problem with Hamiltonian perturbation theory, we shall instead use Lie transforms directly on the dynamical vector field. We do this for the purposes of illustration. In Chapter 3 of this thesis, we treat the much more general problem of relativistic guiding-center motion in arbitrary electromagnetic field geometry in space-time (including perpendicular electric fields that may be order unity in the guiding-center expansion

parameter, ϵ), and there we make full use of the Hamiltonian nature of the equations of motion and we spend a great deal of time studying the associated Poisson structure. It is useful to compare the two approaches.

We denote the phase-space coordinates by $\mathbf{z} = (x, y, w, \theta)$, and the equations of motion by

$$\dot{\mathbf{z}} = \frac{1}{\epsilon}\mathbf{V}_0 + \mathbf{V}_1, \tag{C.942}$$

where the dynamical vector field is described by

$$\begin{aligned} V_0^x &= 0 \\ V_0^y &= 0 \\ V_0^w &= 0 \\ V_0^\theta &= \Omega \end{aligned} \tag{C.943}$$

and

$$\begin{aligned} V_1^x &= -w\sin\theta \\ V_1^y &= -w\cos\theta \\ V_1^w &= -\frac{e}{m}\left(E_x\sin\theta + E_y\cos\theta\right) \\ V_1^\theta &= -\frac{e}{mw}\left(E_x\cos\theta - E_y\sin\theta\right). \end{aligned} \tag{C.944}$$

The unperturbed problem, $\dot{\mathbf{z}} = \mathbf{V}_0/\epsilon$, thus has the solution

$$\begin{aligned} x &= x_0 \\ y &= y_0 \\ w &= w_0 \\ \theta &= \theta_0 + \Omega t/\epsilon, \end{aligned} \tag{C.945}$$

so that averages over the unperturbed motion are equivalent to averages over θ.

At first order, Eq. (2.222) tells us that

$$\mathcal{V}_1 = \mathbf{V}_1 - \mathcal{L}_1\mathbf{V}_0, \tag{C.946}$$

where $\mathcal{V}$ denotes the Lie transformed dynamical vector field at first order. The separate components of the above equation are then

$$\begin{aligned}
\Omega\frac{\partial}{\partial\theta}g_1^x &= \mathcal{V}_1^x + w\sin\theta \\
\Omega\frac{\partial}{\partial\theta}g_1^y &= \mathcal{V}_1^y + w\cos\theta \\
\Omega\frac{\partial}{\partial\theta}g_1^w &= \mathcal{V}_1^w + \frac{e}{m}\left(E_x\sin\theta + E_y\cos\theta\right) \\
\Omega\frac{\partial}{\partial\theta}g_1^\theta &= \mathcal{V}_1^\theta + \frac{e}{mw}\left(E_x\cos\theta - E_y\sin\theta\right) + g_1^x\Omega_{,x} + g_1^y\Omega_{,y}.
\end{aligned} \tag{C.947}$$

We demand that the generator vector $\mathbf{g}_1$ be purely oscillatory (single-valued in θ). Thus, averaging the above equations immediately yields

$$\mathcal{V}_1 = 0. \tag{C.948}$$

Then, we can solve Eqs. (C.947) for the components of $\mathbf{g}_1$. We get

$$\begin{aligned}
g_1^x &= -\frac{w}{\Omega}\cos\theta \\
g_1^y &= \frac{w}{\Omega}\sin\theta \\
g_1^w &= \frac{e}{m\Omega}\left(-E_x\cos\theta + E_y\sin\theta\right) \\
g_1^\theta &= \frac{e}{mw\Omega}\left(E_x\sin\theta + E_y\cos\theta\right) - \frac{w\Omega_{,x}}{\Omega^2}\sin\theta - \frac{w\Omega_{,y}}{\Omega^2}\cos\theta.
\end{aligned} \tag{C.949}$$

Thus we have completely removed the perturbation in the dynamical vector field at first order. The guiding-center equations of motion will appear at the next order, as will the desired correction to the gyrofrequency.

At second order, Eq. (2.223) tells us that

$$\mathcal{V}_2 = -\mathcal{L}_2\mathbf{V}_0 - \mathcal{L}_1\mathbf{V}_1 + \frac{1}{2}\mathcal{L}_1^2\mathbf{V}_0 = -\mathcal{L}_2\mathbf{V}_0 - \frac{1}{2}\mathcal{L}_1\mathbf{V}_1. \tag{C.950}$$

The generator $\mathbf{g}_2$ must be chosen so that $\mathcal{V}_2$ is purely averaged. Thus, without having to actually compute $\mathbf{g}_2$, we can deduce

$$\mathcal{V}_2 = \left\langle -\frac{1}{2}\mathcal{L}_1\mathbf{V}_1 \right\rangle. \tag{C.951}$$

To get the shift in gyrofrequency, we need only $\mathcal{V}_2^\theta$. Because both $\mathbf{V}_1$ and $\mathbf{g}_1$ contain oscillatory terms, the Lie derivative of one with respect to the other will contain products of oscillatory terms, and some of these will not average to zero. After some tedious algebra, we find

$$\mathcal{V}_2^\theta = -\frac{e\Omega}{2m}\nabla\cdot\left(\frac{\mathbf{E}}{\Omega^2}\right) + \frac{w^2}{4}\nabla\cdot\left(\frac{\nabla\Omega}{\Omega^2}\right). \tag{C.952}$$

This is the gyrofrequency shift. The first term is the shift due to the spatial dependence of the perpendicular electric field, and the second term is the shift due to the spatial dependence of the magnetic field. The first of these terms was discovered by Kaufman [47] in 1960, who also showed that it gives rise to the phenomenon of *gyroviscosity.*

It is interesting to note that, when the results of Chapter 3 are cast into "1 + 3" notation and the nonrelativistic limit is taken, the first of the above pair of terms is present but the second is not. This is because the ordering scheme used is quite different. In this appendix, we treated the perpendicular electric field as an order ϵ quantity, whereas in Chapter 3 we took it to be order unity. Thus both terms appear at the same order above (the first term has a spatial gradient and an electric field, and the second term has two spatial gradients), whereas in Chapter 3 the second term would appear at one higher order than the first term (and we did not calculate to high enough order there to see it). It is also interesting to note that the term involving $\mathcal{R}$ in Eq. (3.511) of Chapter 3 is a three (or higher) dimensional effect, and has no analog in two-dimensional guiding-center motion.

Appendix D

Properties of the Special Functions

The following is a list of properties of the Q_ℓ and R_ℓ functions that follow directly from their definitions given in Section 4.4.

D.0.9 The Q Functions

Property D.0.9.1 (Defining Integral)

$$Q_\ell(x) \equiv \frac{1}{2\pi} \int_0^{2\pi} d\xi \left(\frac{e^{ix \sin \xi} - 1}{ix \sin \xi} \right) e^{-i\ell\xi}$$

Property D.0.9.2 (Relationship with Bessel Functions)

$$\frac{d}{dx} \left[x Q_\ell(x) \right] = J_\ell(x)$$

Property D.0.9.3 (Power Series)

$$Q_\ell(x) = \sum_{j=0}^{\infty} \frac{(-1)^j (x/2)^{2j+\ell}}{(2j+\ell+1)j!(\ell+j)!} = \frac{(x/2)^\ell}{(\ell+1)!} + \cdots$$

Property D.0.9.4 (Asymptotic Behavior for Large Argument)

$$Q_\ell(x) \sim \frac{1}{x} + \sqrt{\frac{2}{\pi x^3}} \sin\left(x - \frac{\pi}{2}\ell - \frac{\pi}{4}\right) + \cdots$$

Property D.0.9.5 (Recursion Relations)

$$Q_{\ell-1}(x) + Q_{\ell+1}(x) = \frac{2\ell}{x} \int_0^x dy \frac{J_\ell(y)}{y}$$

$$Q_{\ell-1}(x) - Q_{\ell+1}(x) = \frac{2}{x} J_\ell(x)$$

Property D.0.9.6 (Formula for Derivative)

$$Q'_\ell(x) = \frac{1}{x}\left[J_\ell(x) - Q_\ell(x)\right]$$

Graphs of the Q functions are presented in Fig. D.1.

D.0.10 The R Functions

Property D.0.10.1 (Defining Integral)

$$R_\ell(x) \equiv \frac{1}{\pi} \int_0^{2\pi} d\xi \left(\frac{(1 - ix\sin\xi)e^{ix\sin\xi} - 1}{x^2 \sin^2\xi} \right) e^{-i\ell\xi}$$

Property D.0.10.2 (Relationship with Bessel Functions)

$$\frac{d}{dx}\left[x^2 R_\ell(x)\right] = 2x J_\ell(x)$$

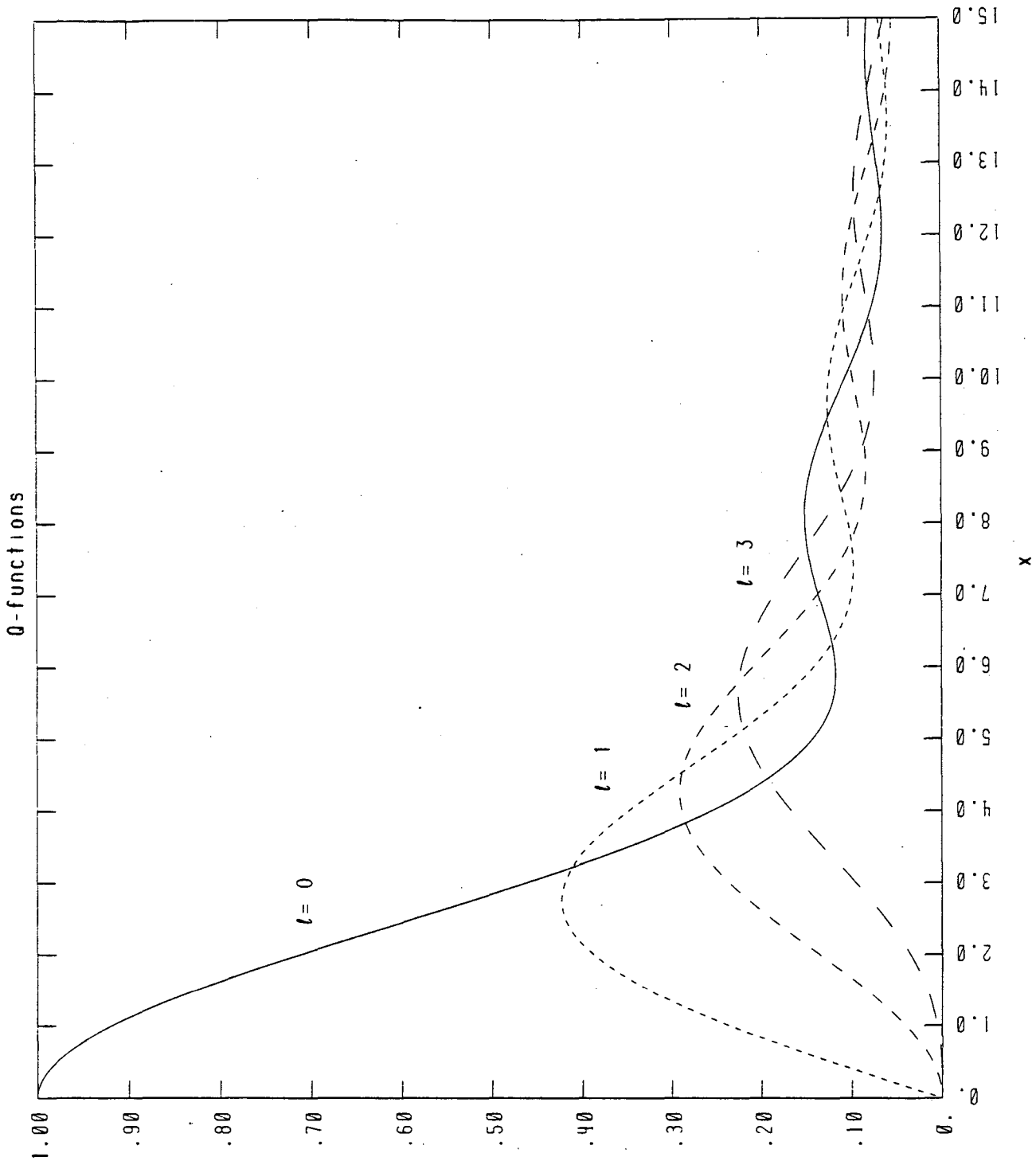


Figure D.1: The Q Functions

Property D.0.10.3 (Power Series)

$$R_\ell(x) = 2\sum_{j=0}^{\infty} \frac{(-1)^j (x/2)^{2j+\ell}}{(2j+\ell+2)j!(\ell+j)!} = \frac{2(x/2)^\ell}{(\ell+2)\ell!} + \cdots$$

Property D.0.10.4 (Asymptotic Behavior for Large Argument)

$$R_\ell(x) \sim \sqrt{\frac{2}{\pi x^3}} \sin\left(x - \frac{\pi}{2}\ell - \frac{\pi}{4}\right) + \cdots$$

Property D.0.10.5 (Recursion Relations)

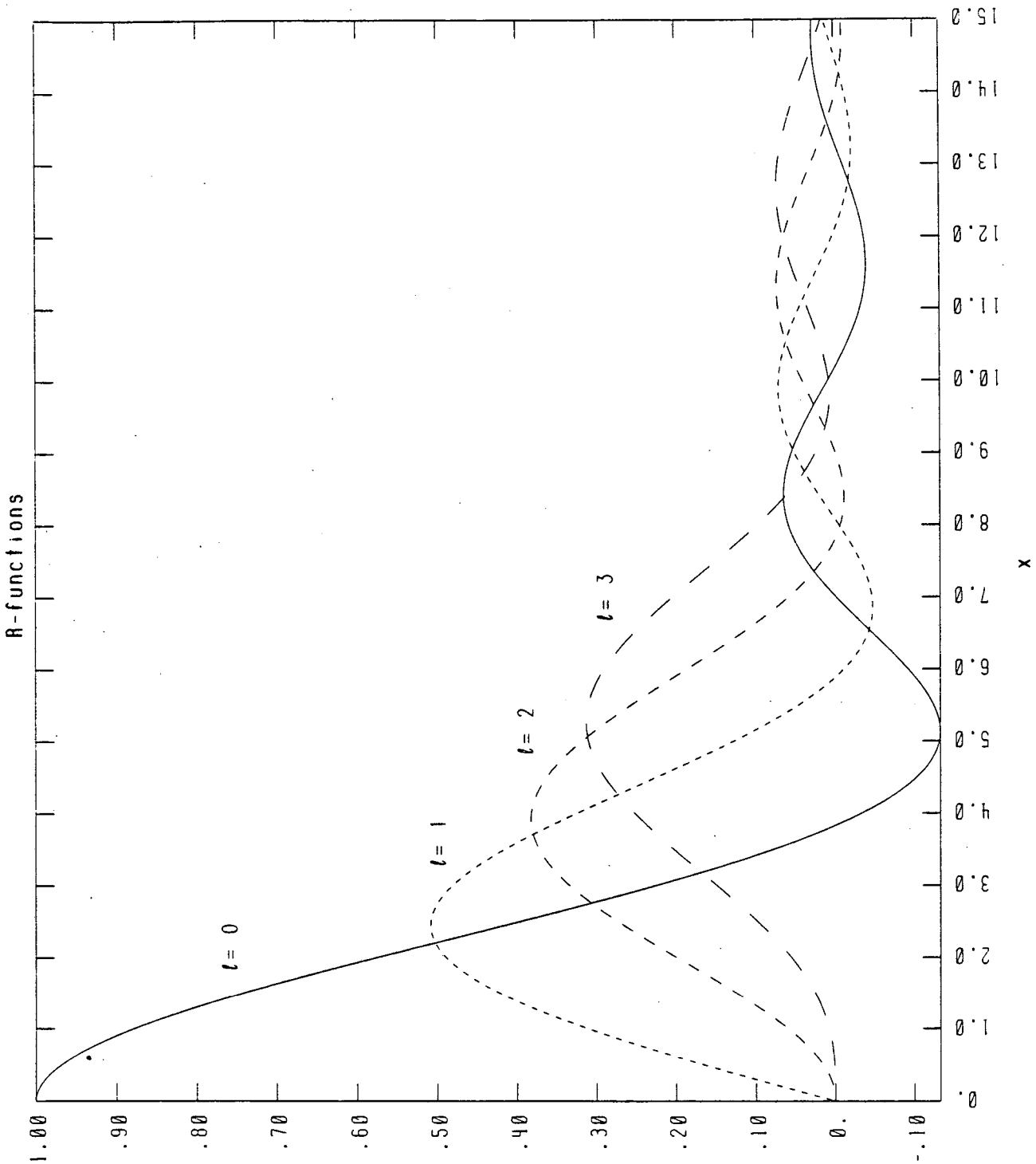
$$R_{\ell-1}(x) + R_{\ell+1}(x) = \frac{4\ell}{x} Q_\ell(x)$$

$$R_{\ell-1}(x) - R_{\ell+1}(x) = \frac{4}{x}\left[J_\ell(x) - Q_\ell(x)\right]$$

Property D.0.10.6 (Formula for Derivative)

$$R'_\ell(x) = \frac{2}{x}\left[J_\ell(x) - R_\ell(x)\right]$$

Graphs of the R functions are presented in Fig. D.2.



XBL 876-3020

Figure D.2: The R Functions

Appendix E

Useful Bessel Function Sums

All of the Bessel function summation formulas used in Chapter 4 can be derived from the following theorems:

$$\sum_{\ell} J_{\ell+k}(z)J_{\ell-k}(z) = \delta_{k0} \tag{E.953}$$

and

$$\sum_{\ell} J_{\ell+k+1}(z)J_{\ell-k}(z) = 0, \tag{E.954}$$

the usual Bessel function recursion relations

$$J_{\ell-1}(z) + J_{\ell+1}(z) = \frac{2\ell}{z}J_{\ell}(z) \tag{E.955}$$

and

$$J_{\ell-1}(z) - J_{\ell+1}(z) = 2J'_{\ell}(z), \tag{E.956}$$

and the parity rule

$$J_{-\ell}(z) = (-1)^{\ell}J_{\ell}(z). \tag{E.957}$$

To verify Eq. (E.953), let us define

$$f_k(z) \equiv \sum_\ell J_{\ell+k}(z) J_{\ell-k}(z), \tag{E.958}$$

and differentiate with respect to z to get

$$\begin{aligned} f_k'(z) &= \sum_\ell \left(J'_{\ell+k} J_{\ell-k} + J_{\ell+k} J'_{\ell-k}\right) \\ &= \frac{1}{2} \sum_\ell \left[\left(J_{\ell+k-1} - J_{\ell+k+1}\right) J_{\ell-k} + J_{\ell+k}\left(J_{\ell-k-1} - J_{\ell-k+1}\right)\right] \\ &= \frac{1}{2} \sum_\ell \left(J_{\ell+k-1} J_{\ell-k} - J_{\ell+k+1} J_{\ell-k} + J_{\ell+k+1} J_{\ell-k} - J_{\ell+k-1} J_{\ell-k}\right) \\ &= 0, \end{aligned} \tag{E.959}$$

where we have used Eq. (E.956) in the second line and have redefined the summation variable in the third line (we have also omitted explicit indication of the functional dependence of J_ℓ on z since no ambiguity can result from doing so). This means that $f_k(z)$ cannot depend on z, so it is a constant for each value of k. To find the value of this constant, set z equal to zero in Eq. (E.953). Recalling that $J_\ell(0) = \delta_{\ell 0}$, we see that $f_k(z) = \delta_{k0}$, and the theorem is proved.

To verify Eq. (E.954), use the parity rule, Eq. (E.957). We have

$$\begin{aligned} \sum_\ell J_{\ell+k+1} J_{\ell-k} &= \frac{1}{2} \sum_\ell \left(J_{\ell+k+1} J_{\ell-k} + J_{-\ell-k-1} J_{-\ell+k}\right) \\ &= \frac{1}{2} \sum_\ell \left(J_{\ell+k+1} J_{\ell-k} + (-1)^{2\ell+1} J_{\ell+k+1} J_{\ell-k}\right) \\ &= \frac{1}{2} \sum_\ell \left(J_{\ell+k+1} J_{\ell-k} - J_{\ell+k+1} J_{\ell-k}\right) \\ &= 0, \end{aligned} \tag{E.960}$$

where we have redefined the summation variable in the first line ($\ell \mapsto -\ell$ in the second term), and used the parity rule in the second line.

These theorems can be used to derive sum rules with summands that are quadratic in the Bessel functions. To do this, note first that setting $k = 0$ in

Eqs. (E.953) and (E.954) immediately yields

$$\sum_\ell J_\ell^2 = 1 \tag{E.961}$$

and

$$\sum_\ell J_{\ell+1} J_\ell = \sum_\ell J_\ell J_{\ell-1} = 0. \tag{E.962}$$

To derive a sum rule that includes ℓ raised to some power, first use Eq. (E.955) to get rid of the power of ℓ. To derive a sum rule that includes a derivative of a Bessel function, first use Eq. (E.956) to express the Bessel function derivative in terms of undifferentiated Bessel functions; alternatively, if a sum rule that includes a Bessel function derivative can be expressed as the derivative of another sum rule with undifferentiated Bessel functions, then this is usually a better way to proceed.

As an example of some generality, consider the sum over ℓ of $\ell^4 J_\ell J'_\ell$. This can be expressed as follows:

$$\sum_\ell \ell^4 J_\ell J'_\ell = \frac{1}{2}\frac{d}{dz}\sum_\ell \ell^4 J_\ell^2. \tag{E.963}$$

Now note

$$\begin{aligned}
\ell^4 J_\ell^2 &= \ell^2 (\ell J_\ell)^2 \\
&= \frac{\ell^2 z^2}{4}\left(J_{\ell-1}^2 + 2J_{\ell-1}J_{\ell+1} + J_{\ell+1}^2\right) \\
&= \frac{z^2}{4}\Big\{\left[(\ell-1)^2 + 2(\ell-1) + 1\right] J_{\ell-1}^2 \\
&\quad + 2\left[(\ell-1)(\ell+1) + 1\right] J_{\ell-1}J_{\ell+1} \\
&\quad + \left[(\ell+1)^2 - 2(\ell+1) + 1\right] J_{\ell+1}^2\Big\} \\
&= \frac{z^2}{4}\Big\{\left[\frac{z^2}{4}\left(J_{\ell-2} + J_\ell\right)^2 + 2\frac{z}{2}\left(J_{\ell-2} + J_\ell\right) J_{\ell-1} + J_{\ell-1}^2\right] \\
&\quad + 2\left[\frac{z^2}{4}\left(J_{\ell-2} + J_\ell\right)\left(J_\ell + J_{\ell+2}\right) + J_{\ell-1}J_{\ell+1}\right] \\
&\quad + \left[\frac{z^2}{4}\left(J_\ell + J_{\ell+2}\right)^2 - 2\frac{z}{2}\left(J_\ell + J_{\ell+2}\right) J_{\ell+1} + J_{\ell+1}^2\right]\Big\},
\end{aligned} \tag{E.964}$$

so that application of our theorems to this last equation yields

$$\begin{aligned}\sum_\ell \ell^4 J_\ell^2 = \frac{z^2}{4}\Big\{&\left[\frac{z^2}{4}(1+2\cdot 0+1)+z(0+0)+1\right]\\ &+2\left[\frac{z^2}{4}(0+0+1+0)+0\right]\\ &+\left[\frac{z^2}{4}(1+2\cdot 0+1)-z(0+0)+1\right]\Big\}\\ &=\frac{z^2}{2}+\frac{3z^4}{8}.\end{aligned} \tag{E.965}$$

Thus, we finally get

$$\sum_\ell \ell^4 J_\ell J_\ell' = \frac{z}{2}+\frac{3z^3}{4}. \tag{E.966}$$

The following is a list of useful results that can be established in the above manner:

$$\sum_\ell J_\ell \mathcal{J}_\ell^+ = 0 \tag{E.967}$$

$$\sum_\ell \ell J_\ell \mathcal{J}_\ell^+ = -\frac{\rho}{\sqrt{2}\lambda_B} F_0 \cdot \mathbf{k} \tag{E.968}$$

$$\sum_\ell \mathcal{J}_\ell^{-*} \mathcal{J}_\ell^+ = \frac{i}{\lambda_B} F_0 \tag{E.969}$$

$$\sum_\ell \ell \mathcal{J}_\ell^{-*} \mathcal{J}_\ell^+ = P_\perp \tag{E.970}$$

$$\sum_\ell (J_{\ell-1} - J_{\ell+1})\, \mathcal{J}_\ell^{-*} = -\frac{\sqrt{2}}{k_\perp \lambda_B} F_0 \cdot \mathbf{k}. \tag{E.971}$$

These sum rules are needed in the proof that the results for K_2 in Eqs. (4.717), (4.719) and (4.735) are indeed the same.

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720