LBL--20147

DE86 001485

Geometric Perturbation Theory and Plasma Physics

by

Stephen Malvern Omohundro

Ph.D. Thesis

Accelerator and Fusion Research Division
Lawrence Berkeley Laboratory
and
University of California
Berkeley, California 94720

April 4, 1985

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

Geometric Perturbation Theory and Plasma Physics

Stephen Malvern Omohundro

Geometric Perturbation Theory and Plasma Physics

Stephen Malvern Omohundro

ABSTRACT

Modern differential geometric techniques are used to unify the physical asymptotics underlying mechanics, wave theory and statistical mechanics. The approach gives new insights into the structure of physical theories and is suited to the needs of modern large-scale computer simulation and symbol manipulation systems. A coordinate-free formulation of non-singular perturbation theory is given, from which a new Hamiltonian perturbation structure is derived and related to the unperturbed structure in five different ways. The theory of perturbations in the presence of symmetry is developed, and the method of averaging is related to reduction by a circle group action. The pseudo-forces and magnetic Poisson bracket terms due to reduction are given a natural asymptotic interpretation. Similar terms due to changing reference frames are related to the method of variation of parameters, which is also given a Hamiltonian formulation. These methods are used to answer a long-standing question posed by Kruskal about nearly periodic systems. The answer leads to a new secular perturbation theory that contains no ad hoc elements, which is then applied to gyromotion. Eikonal wave theory is given a Hamiltonian formulation that generalizes Whitham's Lagrangian approach. The evolution of wave action density on ray phase space is given a Hamiltonian structure using a Lie-Poisson bracket. The relationship between dissipative and Hamiltonian systems is discussed. A theory motivated by free electron lasers gives new restrictions on the change of area of projected parallelepipeds under canonical transformations. A new type of attractor

is defined which attracts both forward and backward in time and is shown to occur in infinite-dimensional Hamiltonian sytems with dissipative behavior. The theory of Smale horseshoes is applied to gyromotion in the neighborhood of a magnetic field reversal and the phenomenon of reinsertion in area-preserving horseshoes is introduced. The central limit theorem is proved by renormalization group techniques. A natural symplectic structure for thermodynamics is shown to arise asymptotically from the maximum entropy formalism in the same way the structure for classical mechanics arises from quantum mechanics via path integrals. The new structure for thermodynamics is used to generalize Maxwell's equal area rule.

Acknowledgements

It is a pleasure to here thank those people who made this thesis possible and this period of my life positive, productive, and exciting.

My mother, Hope G. Wallace, who through her example has taught me to be persistent, organized, and to love learning, and who has given me her love and support in every possible way. My father, Stuart W. Omohundro, who taught me to pursue philosophical questions and to love oration and vigorous discussion, and who with his wife Astrid Omohundro, has helped support me financially over the years. My stepfather Dr. Leon Wallace, who by example showed me how to pursue scientific thought and research, and who has always given generously to me, both financially and emotionally.

To my advisers I owe a great debt of thanks. They have influenced every aspect of this thesis, both in spirit and in content and many of their ideas appear within it.

Allan N. Kaufman has created a near perfect research environment at the Lawrence Berkeley Laboratory, with just the right mix of individual freedom and group endeavor. He has been an ideal mentor, constantly including me in his research, beginning with his stimulating year and a half plasma course, on through the great ferment of ideas he is continually generating ("Let me show you this great idea I thought of at 3:00 this morning!") in the development of his unique unified approach to theoretical plasma physics. He has given me immense freedom and has been extremely supportive of the paths I have taken, even when they didn't

seem to be along the main thrust of our group's research (eg. cellular automata, knotted dynamical orbits). He has been very innovative in bringing into physics ideas from other disciplines through his wonderful free-for-all dynamics seminar, in which we tackled head on some rather abstract mathematics reading, and through his close contact with mathematicians, bridging some traditionally wide gaps. With his deep insights into physics, his interminable curiosity, and continual probing into the structure of basic theory, Allan has taught me to be a physicist.

Alan Weinstein, through his course on geometrical mechanics in my first year of graduate school and through his insightful comments and lectures since, has taught me differential geometry and its absolutely central place in physical thinking. Through his example, I have learned that a good picture can be worth a thousand equations and that parsimony and elegance often arise from abstraction. At weekly meetings with his students, Alan continually makes those short insightful remarks that change the whole direction of research. Through his patient explanations and striking insights, Alan has taught me mathematics.

Maureen Saunders for her love and support during often stressful times, and for enlivening my life even during her own struggles at medical school.

Peter Blicher, my roommate during all five years of graduate school for hundreds of great discussions about everything (including many of the ideas in this thesis) and for firsthand advice on using TEX for writing a thesis. He and Lynn Hall are responsible for my going out to dinner and seeing movies many more times than I would have otherwise.

Jerry Marsden for many helpful comments and discussions on the material of

this thesis, especially for suggesting averaging as an interesting procedure to look at geometrically, and for including me in many seminars, gatherings and conferences which have really shaped my understanding of mathematical physics. His excellent profuse writings and boundless intellectual energy will always serve as an example for me.

Robert Littlejohn, for insightful discussions on many subjects, particularly for his explorations into eikonal wave theory and for the observation of the many similarities between wave theory and statistical mechanics. His writings on guiding center theory have served to stimulate me (and many other people) to think about perturbation theories in new ways.

Richard Montgomery for being a sounding board for virtually everything in this thesis and for his contributions which appear throughout (especially in the sections on geometric perturbation theory and guiding center theory). At our weekly meetings he suggested many useful directions and explained many mysteries to me. Debbie Lewis for her friendship and encouragement, Bruce Boghosian for many discussions and ideas (especially on dissipative Hamiltonian systems, reversibility, infinite dimensional Lie transforms, and relativistic guiding center theory), Jonathan Wurtele for many discussions and for explaining FEL's to me, Andy Sessler for suggesting the questions about the projection of area under Hamiltonian flow, Philippe Similon for many discussions, especially on ponderomotive stabilization, adiabatic invariants and guiding centers, Ted Courant for discussions of Hamiltonian mechanics and guiding centers, John David Crawford for many discussions and insights into dynamical systems and plasma physics and for collaborating on [Crawford and

Omohundro, 1984], Steve McDonald for his insights into eikonal wave theory, Celso Grebogi for discussions on dynamical systems and ponderomotive theory, Darryl Holm for being responsible for a wonderful summer that I spent at Los Alamos and for his constant interest in and encouragement of my work, Doyne Farmer who has given me many insights into life, dynamics and statistics and who has invited me back to Los Alamos for several great stays, Norm Packard and Jim Crutchfield for many discussions of dynamical systems theory, Alice Roos for insights into integrability and for sharing her perspectives on perturbation theory, Bill Baird whose drive for knowledge and quest for insights is truly inspiring, and Yukkei Hui and Iña Chang who are continuing the application of modern mathematical tools to plasma physics.

This thesis was written on the ccc machine at Livermore using Donald Knuth's excellent mathematical typesetting language TEX and on an Apple Macintosh computer. These tools have allowed me to proceed much more quickly and with much greater ease than I might have otherwise.

The work reported in this thesis was supported by the Office of Basic Energy Sciences of the U. S. Department of Energy under Contract No. DE-AC03-76SF00098. This support is due to the far-reaching insight of Oscar Manley who has the vision to support research into basic issues.

The writings of Vladimir Arnold and, in modern guise, the fundamental ideas of Leonhard Euler and Henri Poincaré have constantly served me as inspiration.

Short Table of Contents

Table of Contents

Chapter 1: Introduction

"One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity."—J.W. Gibbs

1.1. Summary and Motivation

This thesis presents the underlying theoretical basis for an ambitious program to develop a unified, coordinate-free theory of asymptotic perturbation methods in the three major areas of physics: mechanical systems, wave systems, and statistical systems. This program has far-reaching consequences, both practical and theoretical, which we will outline here. It is quite clear that we are entering a new era in physics and engineering in which powerful computers will play a major role. The two major applications of the computer will be simulating physical systems and performing symbolic computations. Both of these areas are fundamentally impacted by our research.

It has long been known that many of the most important problems in physics and engineering design are analytically intractable. Recent theoretical developments have shown that this intractability can be inherent in the problem, and not due simply to insufficient mathematical technique. Rigorous results from dynamical systems theory have shown that almost all systems of interest have chaotic behavior.

which absolutely precludes the development of analytic solutions (see for example [Guckenheimer and Holmes, 1983]). Recent work in cellular automata theory has shown that there are problems whose behavior cannot be predicted by any algorithm which is shorter than direct simulation [Wolfram, 1984].

Computer simulation is therefore destined to play a fundamental role in the study of physical problems. Unfortunately, the very same exponential divergences of neighboring orbits which cause difficulties in analytic treatments also plague computer simulations. For a given accuracy of prediction, the computing power required typically grows exponentially with the time-scale to be studied measured in units of the smallest important physical time-scale. Advances in parallel computation will increase the power of computers by factors of about a million in the next twenty years. While large, this number pales against the spectre of needed exponential growth and the new scales posed by fully three-dimensional simulations. It is therefore necessary to develop reliable analytic theories for preconditioning problems prior to simulation. An important example of this concept is utilized in studying the gyromotion of a charged particle in slowly-varying magnetic and electric fields. The particle motion consists of fast gyrations near the gyrofrequency, on top of slow drifts of the center of gyration. If one simulates the particle motion directly, the simulation errors accumulate on the time-scale of the gyroperiod, quickly leading to incorrect results on the drift time-scale. By first introducing asymptotics and analytically removing the fast gyration, we obtain the so-called guiding center description. These equations have significant evolution on only the drift time-scale, and the simulation retains validity for much longer times with the same computing

resources. A similar motivation is behind the description of a particle in a wave using oscillation centers and the description of the evolution of an eikonal wave in terms of modulational equations.

The need for accurate asymptotics performed to very high order will eventually be met by using symbolic manipulation programs such as MACSYMA and SMP. For this to be possible, it is absolutely essential that we have methods that are precisely defined with a precisely defined domain of applicability, are systematic, and require no *ad hoc* choices in their implementation. It is also very important to design simulations to take advantage of any precise knowledge of the dynamics, such as energy or momentum conservation. Insight into the theoretical structure of a physical theory leads to much more efficient and testable computer code and may mean the difference between a successful theory and a useless one.

For these practical reasons, as well as for fundamental theoretical ones, we have taken a new approach to physical asymptotics. The approach we take is based on some revolutionary new ideas that are changing the underlying mathematical structure of physics. The introduction of the vector calculus by Heaviside in 1882 led not only to the streamlining and simplification of calculations, but to new conceptual insights into the structure of physical law. The application of coordinate-free differential geometry, developed primarily by Elie Cartan in the 1930's, is having an even more profound effect on both the expression and the conceptual structure of physics. Coordinates are not physical, and the new formulations encourage the use of only physically relevant concepts, never *ad hoc* artifacts of some coordinate description. Hamiltonian mechanics has been particularly revolutionized by this

reexpression. It was discovered in the 1960's that, just as Riemannian geometry describes the structure of space-time, *symplectic* geometry describes the structure of Hamiltonian phase space. An introduction to this theory in the context of our work is presented in chapter 2. The full impact of the reformulation of Hamiltonian mechanics has been felt only recently. In the past few years virtually every area of physics has been Hamiltonianized. The new perspective has shed light on the underlying symmetry structure of these theories (including the elucidation of automatically conserved quantities, called Casimir functions), has yielded improved nonlinear stability results based on Arnold's stability method, and has given insight into the reasons for the integrability of certain systems.

Hamiltonian structures were originally introduced by Lagrange to simplify and to check perturbation calculations. They will play a similar role in modern computer simulations. Most of the recent developments in Hamiltonian mechanics have dealt with fundamental models of physical systems and have not studied the structure of perturbation equations. We have initiated a study of such equations and discovered several important results.

Our first goal was to study ordinary non-singular perturbation theory. Chapter 2 gives a coordinate-free description of this technique in terms of the mathematical theory of jets. This is important for systematizing perturbation methods and bringing them under the purview of modern mathematical methods. Many physical systems have state spaces that are naturally manifolds, and many are naturally infinite-dimensional. Before our work, it was not clear that one obtained the same perturbation expressions when one worked in different coordinate systems and there

was no systematic means for dealing with the geometrically more complex cases.

Chapter 4 studies this perturbation theory in a Hamiltonian context. Again we find that the theory is expressible in coordinate-free language. We have discovered that a remarkable and important Hamiltonian structure governs the perturbation equations themselves. The fact that the perturbation equations are Hamiltonian means that, for the first time, all of the powerful techniques of Hamiltonian mechanics (including Noether's theorem, energy conservation, Liouville's theorem, and stability techniques) may be applied directly to perturbed systems. We have discovered that the new perturbation structure is related to the key elements of modern Hamiltonian mechanics in *five* different ways. (We have worked out the direct coordinate description, the relationship with natural structures on the so-called iterated tangent bundle, the relationship with a natural structure on a path space, a new sense in which the perturbation state space can be viewed as a "derivative" which unifies our method with previous work, and finally an extensive theory of perturbations in the presence of symmetry.) The theory describing the relationship of perturbations and symmetries extends to perturbation problems what is perhaps the most important idea of modern Hamiltonian mechanics: *reduction*. This is a procedure, formulated by Marsden and Weinstein and discussed in chapter 2, which is a far-reaching generalization of Noether's theory of simplification in the presence of symmetry. Every modern Hamiltonian structure, including those arising in plasma physics, magneto-hydrodynamics, fluid dynamics, general relativity, electromagnetism, quantum mechanics, superfluidity, and superconductivity, may be obtained from underlying canonical structures by means of reduction. In most

of these cases, the resulting Poisson bracket is intimately related to the so-called Lie-Poisson bracket, which arises from reduction. We have broadened this theory to encompass perturbed systems and have shown that the perturbation structures are intimately related to certain "jet-group" symmetries. Our Hamiltonian perturbation structure arises from the Lie-Poisson bracket of this larger group.

Perhaps the most important application of non-singular perturbation theory is as a component of a singular or secular perturbation calculation. Using geometric methods, we have been able to make fundamental advances in secular perturbation theory as well. We first demonstrate in section 2.9 that the method of averaging can be viewed as an application of reduction by an approximate circle action. This allows the method of averaging to be incorporated into any situation amenable to reduction. We explicitly calculate the Poisson structure for $E \times B$ drift using these new methods in section 2.10. Nowhere in the calculation does one need to introduce the unphysical or *ad hoc* elements usually required. The resultant bracket is derived with much less calculation than by any previous method.

To reach higher order than the method of averaging, we reexamine the seminal theory of Kruskal in the light of the new methods in chapter 5. He introduced the first systematic, order by order calculation of adiabatic invariants to all orders. His technique requires one to make changes of coordinates order by order. Each change, however, requires one to make certain arbitrary choices leading to an ad hoc element in the perturbation algorithm. Kruskal was able to show that nonetheless the vector field generated by the adiabatic invariant (which we would now call an approximate circle symmetry) was uniquely defined to all orders. This vector field

is the real goal of the perturbation technique and Kruskal was moved to comment in his paper: "It does not appear obvious whether an explicit recursion formula to determine R [the symmetry vector field] in terms of f [the dynamical vector field] can be found. If so, the whole theory of this paper might be simplified and rendered less deep." Indeed, our approach explicitly gives just such a formula and relates it to the developments in Hamiltonian mechanics listed above. Our algorithm is completely well defined with no *ad hoc* elements and so is ideally suited for symbolic implementation on a computer. As an example, we compute the symmetry vector field for two dimensional gyromotion to second order in chapter 6.

We have found that Kruskal's results (and others of the same type) are often misunderstood and misused. While Kruskal showed that the accuracy of the adiabatic invariant conservation may be made exponentially small in the small parameter, the *time-scale* over which this is valid is only the reciprocal of the small parameter. Kruskal's result is often quoted as: "adiabatic invariants are preserved to all orders in the small parameter" and wrongly interpreted to mean exponentially small error for exponentially long times. This misinterpretation can be very dangerous in the situations where the theorem is applied, and so we have constructed a number of counterexamples to illustrate the limits of the theory in section 5.4.

We have also successfully considered the method of variation of parameters in this light and have explicitly demonstrated the connection with the notions of "pseudoforces" and "magnetic curvature terms" in the Poisson brackets of reduced systems in chapter 3. We have shown a new way of interpreting the Coriolis force (as a term in the Poisson bracket due to reduction) which is extendible to any other

system expressed in a changing reference frame.

The next class of systems we study are those describing the asymptotic evolution of eikonal waves. We have developed a number of new theoretical constructs in this field including a precise asymptotic definition of local Fourier transform. These ideas are presented along with a summary of the geometric approach to eikonal wave theory in chapter 7. We have also succeeded in our main goal, which was to develop a systematic method for finding a Hamiltonian description of modulational equations given the underlying Hamiltonian wave system. Our technique is algorithmic and does not depend on linearity, nor on any other special features of the system. We present the theory for the Klein-Gordon equation in detail in chapter 8. Our approach encompasses the systems amenable to Whitham's averaged Lagrangian technique, but applies in addition to Hamiltonian systems that do not arise from a Lagrangian (for example, any of the systems with Casimirs or any system based on a Lie-Poisson bracket). It also unifies the study of eikonal waves with the other Hamiltonian systems we have discussed.

We also study the common case of linear waves described by a canonical Hamiltonian structure with wave amplitude and phase as conjugate variables. We have shown that these systems are naturally described by a wave action density on the entire x, k phase space. The evolution of this action density is itself Hamiltonian, but with respect to a natural Lie-Poisson bracket, as is demonstrated in chapter 9. The symmetry group underlying this theory is the same as that for the Vlasov equation.

Using new mathematical techniques, we have discovered several novel results

regarding some long-standing questions about the relationship between dissipative and non-dissipative systems. In chapter 10, we show by explicit construction that any dynamical system (with any amount of dissipation) can be imbedded in a Hamiltonian system of twice the dimension or a Poisson system of only one higher dimension. We also show by explicit construction that there are a Hamiltonian system and a Poisson system of only one dimension greater that project (by ignoring the value of a coordinate) to become any arbitrary dynamical system. These results show at once that Hamiltonian systems are very general and that it can be very dangerous to indiscriminately add new variables to a system. In particular, we show that by introducing time-dependent changes of coordinates, one can make any system look integrable, or coerce Liapunov exponents to take any desired values. We give several explicit examples of seemingly harmless yet truly dangerous operations. We show in an explicit example involving a resonance that, by indiscriminate use of the metbod of Lie transforms, one can inflict mortal injury to the underlying physics.

In chapter 11 we use modern symplectic geometry to study some questions that arise in the design of free electron lasers and other accelerators. One often wants to force the particles in a beam into a more confined region in phase space. Most devices act on the particles in a (time-dependent) Hamiltonian way and so we consider the effect of canonical transformations on regions of phase space. Liouville's theorem says that it is impossible to change the volume of the region in phase space. Often, though, we are interested in the projection of our region onto some subset of the degrees of freedom (e.g., the longitudinal position and momentum

variables). One might attempt to shrink the image of such a projection. Courant addressed this question for linear systems, but only under very special conditions: particles in ellipsoids or parallelepipeds with axes aligned with the coordinate axes both before and after the transformation. Weinstein has given a general theorem that applies only for sufficiently small transformations. We generalize Courant's theory in a fundamental way and obtain results for arbitrary linear canonical transformations. This is accomplished by positing and proving a new theorem about the structure of parallelepipeds in high dimensional spaces. We give several examples and counterexamples and discuss the possibility of a generalization to statements about projected measures as opposed to volumes.

In chapter 12 we introduce a new class of attractors into dynamical systems theory that sheds light on classical irreversibility paradoxes. By construction, we exhibit systems possessing a fixed point which is attracting both forward and backward in time. In chapter 13 we show that exactly this behavior occurs in many infinite dimensional Hamiltonian systems and is responsible for the appearance of dissipation in many cases. Examples include a variety of wave systems, Landau damping, resonant coupling, and the decay of correlations in chaotic maps.

The fundamental mathematical structure behind chaos in dynamical systems is known as Smale's horseshoe. This is a very commonly found piece of nonlinear mappings which guarantees the existence of orbits that hop between two regions according to any sequence of random 1's and 0's. Any map with a horseshoe has dynamics which is as unpredictable as a sequence of coin tosses. In chapter 14 we show that periodically perturbed gyromotion in the neighborhood of a magnetic

field reversal leads to horseshoes in the particle dynamics. This has the consequence that there are particle orbits that loop from side to side of the field reversal according to any random sequence. The detailed structure of horseshoes is extremely complex and is only beginning to be understood. Horseshoes are responsible for chaos in both dissipative and Hamiltonian systems but have been studied mostly in the dissipative case. We have discovered a fascinating apparent paradox regarding Hamiltonian horseshoes whose resolution indicates a much more intricate structure for the Hamiltonian than for the dissipative case. We call the new phenomenon "reinsertion" because it forces the tongues of the unstable manifold to reinsert into other tongues an infinite number of times. It is possible that this new structure is responsible for some extremely complex phenomena that have been recently observed by Holmes and Whitley in the transition from dissipative to Hamiltonian horseshoes.

In chapter 15 we give a dynamical systems description of the idea of renormalization and prove the central limit theorem using renormalization group techniques. Using renormalization group methods, Feigenbaum discovered a universal scaling constant for period-doubling cascades. He used a Cray supercomputer to calculate this constant to high accuracy. In section 15.4 we present a poor man's approach that gives the constant to within 25 percent on the back of an envelope.

The last type of physical asymptotics that we have incorporated into our theoretical structure is the transition from statistical mechanics to thermodynamics. This is perhaps the most profound aspect of our work and introduces structures that are of fundamental physical significance. As discussed in chapter 16, we have

discovered that the asymptotic averaging process that occurs in the transition from statistical mechanics to thermodynamics is exactly analogous to the asymptotic averaging process that takes us from eikonal wave theory to rays (e.g., from quantum mechanics to classical mechanics). We show that a natural symplectic structure for thermodynamics arises from this asymptotics in exactly the same way that the symplectic structure of Hamiltonian mechanics (that has been so fruitful in recent applications) arises from wave asymptotics. Our theory now allows the same powerful results which have caused Hamiltonian dynamics to flourish in recent years to be applied to statistical systems.

The underlying statistical foundation for the new theory is the principle of maximum entropy. We develop several new interpretations for this principle in section 16.2, in which it plays exactly the same role as the principle of least action plays in mechanics. Just as the principle of least action arises asymptotically from the method of stationary phase applied to the Feynman integral over all paths, the principle of maximum entropy arises from the method of steepest descents applied to an integral over all probability distributions. This formulation is new and is expected to lead to the same benefits in statistical mechanics that are derived from the path integral formulation in quantum mechanics.

The integral over paths can be done by integrating over paths with a given constraint and then integrating over the constraint. This leads to the action defined on phase space and the description of dynamics in terms of it. The integral over distributions may be done by integrating over distributions obeying a given constraint and then integrating over the constraint. This leads to the entropy de-

fined on the thermodynamic state space and the description of the equation of state in terms of it. We obtain the wave phase (which is the action) at a given point by introducing Lagrange multipliers that are canonically conjugate to the spatial variables and choosing them so that the extremal action occurs at the point of interest. The value of the multipliers is the wave-vector or *momentum* conjugate to the position and has a value equal to the derivative of the action. This is the origin of the symplectic structure of mechanics. We obtain the entropy for given values of the mechanical quantities by introducing Lagrange multipliers that are thermodynamically conjugate to the mechanical variables and choosing them to make the maximum entropy occur at the state of interest. The value of the multipliers is the derivative of the entropy with respect to the mechanical variables. This is the origin of the symplectic structure of thermodynamics. The Lagrange multipliers cause our path integral to be a Fourier transform and our probability integral to be a Laplace transform. The asymptotic evaluation of these transforms by stationary phase and steepest descents, respectively, leads to the Legendre transforms that are so central in mechanics and thermodynamics.

As a first application of the new structure of thermodynamics, we reexamine the classical Gibbsian theory of phase transitions in section 16.12. We endow the Maxwell equal area rule with a natural geometric interpretation within the new theory. It is then generalized to describe the phase transition surface in an arbitrary direction instead of only along the isothermal surfaces considered by Maxwell. If the flood of new practical results that arose from the symplectic interpretation of Hamiltonian mechanics is any indication, we can expect that this incorporation of

thermodynamics into the modern framework will soon yield many other significant results.

1.2. Philosophical Approach

What is the job of theoretical physics? The usual answer is that it is to suggest new physical laws which are then verified by experiment. This is indeed what occured in the great revolutions of Newtonian mechanics, special and general relativity and quantum mechanics. This is not what most theoretical physicists do however. They start with the fundamental laws (be they quantum electrodynamics or some other model) and determine what behaviors these laws imply in special situations. Ideally, this enterprise is one of mathematical deduction, but practically one must often make intuitive leaps either because current mathematical techniques aren't powerful enough or simply because the underlying mathematical model of the physical situation is inadequate. The ultimate goals of such studies are the prediction and understanding of physical behaviors. Sometimes these come in the form of numbers to compare with experiment, but more often one is interested in qualitative features like the stability of an equilibrium state or the type of evolution expected of a given system. At the highest level, one finds general principles which apply to many situations and give reasons for the qualitative behaviors observed.

From this viewpoint, the enterprise of theoretical physics may be thought of as the creation of a succession of models. The highest models are extremely general in their domains of application but are very intractable and, because of their generality, give little insight into actual behavior. The lower models specialize the general ones to smaller classes of situations and make more and more precise predictions until finally the lowest models describe a single experimental setup and predict the numerical values of individual measurements.

The goal of this thesis is to examine very general features of this grand hierarchy of models that is physics. The sorts of questions we are interested in are: When does a model simplify? What makes a model simplify? How robust are the features of the simplification?

The fundamental theoretical structure which unifies the models of physics at all levels is the Hamiltonian structure, as may be seen in the beautiful compendium [Landau and Lifshitz, 1960–1981]. The fundamental equations of physics all appear to be Hamiltonian and many of the simplification procedures respect this Hamiltonian structure. The most basic circumstance which allows simplification is the presence of symmetry. Emmy Noether discovered that for systems with a Hamiltonian structure, the presence of a dimension of symmetry allows one to eliminate two dimensions of state space from consideration. Recently, the reduction of fundamental models to more specific ones has been accomplished within a Hamiltonian framework for many examples using the symmetries present in the underlying situation.

Many times, however, we are not exactly in the symmetric situation, but we are close to it in some sense. The physics is described asymptotically using the diverse methods of perturbation theory which have come to be the mainstay of much of theoretical physics. One can make a case that all of the actual calculations carried out in quantum electrodynamics, plasma physics, solid state physics and many other fields are expressions of divers perturbation approaches to the underlying equations.

The other great simplifying tool is statistics. As for perturbation theory, the fundamental theoretical justification for using statistical approximations is almost

always lacking, and yet these techniques have led to some of the most useful and accurate theories in any domain of study.

Our purpose in this thesis is to examine in a new light some of these simplification procedures at the heart of physics. Traditionally, physical calculations have been carried out in special, arbitrary coordinate systems which may simplify the calculations but obscure the distinction between what is intrinsic to the physical situation and what is arbitrary. The mathematical physics community has recently been moving to reexpress the fundamental ideas of physics in the coordinate-free language of differential geometry, developed by Elie Cartan. This has led to some resounding successes and has identified many new structures that have direct relevance to physics. Much of the huge body of traditional work in applied mathematics has not yet benefitted from this new viewpoint, however. Books and journal articles on perturbation methods typically describe these methods in terms of a specific example and little or no attempt is made to delineate what physical features of the model have made the method work. The result is a morass of disconnected special cases. Workers in the field have developed intuitions as to what will work where, but this has not been codified into a theory.

The underlying philosophy here is that there is no magic. If a situation simplifies, if a problem is tractable, or if there is some effect which is universal enough to be given a name, then there must be a definite physical reason for it. We wish to understand these reasons. This thesis, of course, is only a beginning in this direction. We find the underlying geometrical content of some of the central perturbation methods used in physical situations. We relate this to the Hamiltonian

structures involved, redo classical derivations and produce new results. To make this work somewhat self-contained, we give intuitive versions of needed background mathematical results. We show how they fit into the picture presented here without too much duplication of material that can be found in standard reference works.

1.3. Guide for the Reader

Unfortunately, unifications of frontier material from difficult subjects do not often make for easy reading. We have tried to alleviate this problem here in several ways. Throughout the thesis, we use concepts from differential geometry, geometric mechanics, and dynamical systems theory. Whenever we use a new concept from these fields, we give an intuitive discussion of the basic ideas involved and refer you to more detailed discussions. Fortunately, this background material is becoming widely known and used in the physics community, primarily because of the existence of several excellent texts.

Geometric mechanics is beautifully presented in [Abraham and Marsden, 1978], [Arnold, 1978], and [Thirring, 1978]. [Arnold, 1978] gives the most intuitive discussions, [Abraham and Marsden, 1978] is the most complete and mathematically precise, and [Thirring, 1978] covers several additional topics like the KAM theorem.

Both [Abraham and Marsden, 1978] and [Thirring, 1978] begin with introductions to differential geometry. An expanded version of this introduction and infinite-dimensional versions of the ideas are given in [Abraham, Marsden, and Ratiu, 1983]. We give specific citations to this reference as the fundamental geometric concepts appear. An intuitively appealing treatment of differential geometry may also be found in [Spivak, 1979].

The ideas of dynamical systems theory are discussed in [Abraham and Marsden, 1978], [Arnold, 1983], and [Guckenheimer and Holmes, 1983]. [Guckenheimer and Holmes, 1983] gives a very nice treatment of many examples in addition to presenting the pure theory.

Two useful compendiums of the various perturbation methods, with many physical examples are Kevorkian and Cole, 1981 and Nayfeh, 1973. The asymptotics of wave theory is given a nice treatment in Whitham, 1974. Guillemin and Sternberg, 1977 is a rich source of mathematical insight into waves, but is fairly difficult to read without a mathematical background.

A unified treatment of the classical statistical physics we need is presented in the statistical physics volume of Landau and Lifshitz, 1960 1981]. Jaynes, 1983] presents the maximum entropy viewpoint that is central to our approach.

Because there is no index, we have made the table of contents very detailed. We have also provided an alphabetical list of key concepts and the sections in which they appear in chapter 17.

The remainder of this introduction is intended to be readable without extensive mathematical background. It presents the basic conceptual structure of the thesis and motivates some of the mathematical concepts.

The body of the thesis is broken up into three parts: I) mechanics, II) waves, and III) dissipation and statistics. While one of our goals is to connect these disciplines, for the most part the three parts may be read independently of one another.

Chapter 2 is intended to introduce the geometric approach to mechanics while introducing our approach to non-singular perturbation theory. It gives intuitive descriptions of the basic concepts of differential geometry with pointers to more detailed reference works. We have tried to point out the key ideas of the geometric approach and to make our discussions easier to follow at a heuristic level than the reference works.

Chapter 3 rests on detailed geometric mechanics and will be most accessible to readers with this background. No other sections depend on this material.

Chapter 4 rests on the material of chapter 2 and will be most accessible to readers with some mathematical background. The philosophy and basic results were presented in chapter 2 and chapter 4 may be viewed as a reference for the details and methods of the approach. The results are used in sections 5.3 and 9.3.

Chapter 5 extends some standard secular perturbation methods. We give an introduction and critique of this theory but the reader may wish to look at standard references and the paper [Kruskal, 1962] while reading this chapter. Section 5.3 rests on geometric notions from chapter 4 and may be omitted by the uninterested.

Chapter 6 is an application of the methods of chapter 5 to gyromotion. We have given fairly complete details of our calculations so that they may be used on other problems. Later chapters do not depend on this one.

Chapter 7 is an introduction to the geometry of eikonal wave theory. Extensive use of the symplectic geometry introduced in chapter 2 is made in this chapter.

Chapter 8 develops a Hamiltonian perturbation technique for eikonal waves which is founded in the geometric ideas of chapter 7 but takes the form of an explicit algorithmic calculation. The examples we present may serve as a model for similar problems.

Chapter 9 again presents an explicit calculation whose theory rests on the ideas of Lie-Poisson brackets presented in chapter 2. The first section may be read independently of this material (though the motivation might seem obscure) but the second section rests heavily on it.

Chapter 10 uses only the elementary ideas of chapter 2 and should be fairly easy to read.

Chapter 11 rests on some symplectic geometry, but the results are easily understood and may be useful in general situations. The proof of the main result uses only linear algebra and induction.

Chapter 12 may be read on its own and serves as the background for chapter 13.

Chapter 14 introduces the dynamical systems concepts necessary and may be read independently of the rest of the thesis. It may be useful for the reader unfamiliar with the ideas of chaotic dynamics to consult some of the more detailed works.

Chapter 15 rests on some ideas of probability theory, dynamical systems theory, and renormalization group theory. It may be read independently of the rest of the thesis but consultation of the reference works listed in that section may provide useful background material for the reader.

Chapter 16 makes many references to chapter 7. It also rests on the ideas of maximum entropy which are quickly sketched here but may be studied in detail from 'Jaynes, 1983'. It is also necessary to introduce many ideas from contact geometry, which clarify many aspects of thermodynamics. It might be useful to look at the more detailed references on this material.

1.4. Intuitive Discussion of the Conceptual Framework

Plasma physics is a fascinating discipline in part because it is at the crossroads of what I consider to be the three fundamental types of models in physics: mechanics, wave theory, and statistical mechanics. Each of these areas gets its richest and most powerful models and elementary concepts from asymptotic approximations to the real physical system. It is this asymptotics which allows us to introduce the concepts central to simplified descriptions of physical phenomena. This thesis explores some amazing relations among these three seemingly disparate disciplines and develops a unified way of understanding the structures that make them work.

We are trying to understand how systems simplify. The key feature of such systems is that they project onto a sub-piece which moves according to its own dynamics which is approximately independent of the exact state of the forgotten pieces. The interesting physics arises from the fact that the two halves are in no sense decoupled, and the effect of the forgotten piece is felt in the kept piece as a new physical effect. Each of our three main types of system can undergo such a simplification, and together these systems span the breadth of physics. There are simple dynamical systems with *fast oscillatory* degrees of freedom, wave systems in the *eikonal limit* of short wavelength, and chaotic systems of a large number of degrees of freedom in the limit in which *statistics* is valid. In each, we make a separation of scales by first introducing a formal parameter (often only implicitly) which increases the separation as it vanishes, and then by doing asymptotics in that parameter.

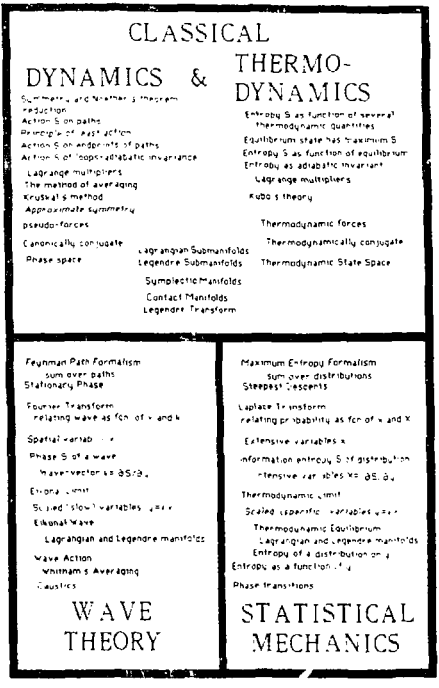

Figure 1.1: Some links between the three disciplines of mechanics, waves, and statistical mechanics.

1.4.1. The Analogy Between Entropy and Action

We will forge links between these three disciplines. Let us first state the various relationships and then give intuitive examples of them. Oddly enough, in each case the classical notation associates the letter "S" with the fundamental unifying quantity. The relation between quantum and classical mechanics is based on the classical action S along a path in phase space, which also represents the quantum phase corresponding to that path. The information entropy S of a probability distribution similarly gives the connection between statistical mechanics and thermodynamics in the maximum entropy approach. In the asymptotic eikonal limit, we may define the action S as a function on the space of only real dynamical paths. In the case of a system all of whose orbits are periodic, we may define the action S as a function on phase space. This function is an adiabatic invariant for slow variations of parameters, and this constancy forces the exchange of energy between the system moving the parameters and the fast periodic oscillations. This exchange causes the slow system to behave as if it had new pseudoforces acting on it, which are expressed in terms of S and are Hamiltonian. When we think of wave systems as themselves being dynamical systems, we can define a local wave action density. Again, this is adiabatically a locally conserved quantity, which causes energy to move around so as to stay constant. In the thermodynamic limit, we can view the entropy as a function S on the space of only the real equilibrium distributions. Asymptotically, in the case of ergodicity, we can assign it to be the function S on the system phase space given by the logarithm of the volume of particle orbits. When we think of a statistical system as itself being a dynamical system, this entropy becomes an

adiabatic invariant and leads to pseudoforces. These are the thermodynamic forces, and they lead to a thermodynamic symplectic state space on which we can define S, and in terms of which we get the thermodynamic equations of state.

Because this sequence of connections is the central unification around which this thesis revolves, we will now give some easily understood examples of the concepts involved.

1.4.2. Adiabatic Invariants and Pseudo-forces

Lorentz was first to ask the question that led to the notion of adiabatic invariance. He wondered how the energy of a simple harmonic oscillator varied as its spring constant slowly changed. It was discovered that, although the energy and frequency both change, their ratio remains asymptotically constant for slow variations. This ratio, H/ω, is an *adiabatic invariant* for the oscillator. A precise definition of the concept of adiabatic invariance and some of its limitations are given in sections 2.9 and 5.4. H/ω is equal to the *action* of the oscillator over one cycle. The action of a closed orbit is the area encircled by that orbit in the (q,p) phase plane. It is a quite general result that this area is an adiabatic invariant for slow variations of the parameters of both linear and nonlinear oscillators. We show how to understand this result geometrically in section 2.9 and how to use this understanding to get more refined results in section 5.4. That we consider area as measured in the coordinates q and p is absolutely essential here. If we had instead coordinatized phase space using velocity rather than momentum, the area would have had no special significance.

In higher dimensions, the action of a closed loop can be defined as the integral of $p^i dq_i$ around the loop where we adopt the usual Einstein convention that repeated indices are summed over. This combination $p^i dq_i$ therefore has a deep physical significance. Geometrically, the choice of individual coordinates q_i and their conjugate p^i is irrelevant and only the combination $p^i dq_i$ is significant. A geometric entity that one can integrate along one-dimensional paths is called a *one-form*. $p^i dq_i$ is intrinsically built into the structure of physical phase space and is therefore called the *canonical one-form* and is usually denoted by θ. Another way to obtain the action of a closed loop in phase space is to find the area of a disc whose edge is the loop. "Area" must be defined in a special way to get an answer that is independent of the means used to obtain it. Applying the (generalized) Stokes' theorem to the line integral of $p^i dq_i$, we see that the action can be defined as the surface integral of $dq_i dp^i$ over the disc. It is important here to keep track of orientation. A geometric structure that one can integrate over two-dimensional surfaces is called a *two-form*. The standard notation uses a wedge to keep track of orientation. The two-form $dq_i \wedge dp^i$ is usually denoted by ω and is called the *symplectic* form on phase space. It is uniquely defined from θ and so also is intrinsic to the geometry of phase space. Since the evolution of a Hamiltonian system preserves the action of closed loops (this is Poincaré's first integral invariant), the notion of area with respect to the symplectic structure ω is also preserved. From the modern perspective of geometry introduced in Felix Klein's Erlanger program, a "geometry" is defined by a mathematical structure and the group of symmetry transformations that preserve that structure. For example, Riemannian geometry studies the concepts that are invari-

ant under the isometries of a metric tensor. The geometry of the phase space of Hamiltonian mechanical systems is *symplectic geometry*, since it is the symplectic structure which is invariant under the canonical transformations describing time evolution.

The real conceptual and computational advantage in the phenomenon of adiabatic invariance is that we can forget about the phase of the fast degree of freedom. Imagine a complex, slowly moving piece apparatus (for example, one of the designs of Rube Goldberg). There is a small weight attached to a string that hangs out of a hole in the side of the machine. As the apparatus moves, the string is slowly pulled in and out of the hole. The part of the string that is hanging out forms a pendulum with the weight. The weight rapidly swings back and forth (say thousands of times before the string length varies appreciably). In trying to understand the operation of this device, one first thinks that the exact state of the machine, including the phase of the pendulum, will be essential in determining the time evolution. The idea of adiabatic invariance tells us that we only need to know the *action* of the pendulum (say by observing the initial amplitude of swing) and not the phase in order to determine the average effect on the apparatus. As the machine pulls the rope in, the energy of the pendulum will change so as to keep its action constant. It will therefore give and take energy from the rest of the machine. The pendulum energy looks like a function purely of string length. From inside the machine, we may regard the string as attached to a nonlinear spring whose potential energy represents the entire pendulum energy. The oscillations have been replaced by a "pseudopotential".

1.4.3. Symmetries and Exact Invariants

One sees similar "pseudopotentials" when one forgets about coordinates in systems with symmetry. In studying a particle moving in a central potential, we may ignore the angular position of the particle, since all angles lead to the same particle behavior (i.e. there is a rotational symmetry). Noether tells us that this symmetry leads to a conserved quantity, namely the angular momentum of the particle. The radial motion must behave in such a way as to keep the angular momentum constant. The *energy* in the angular direction is not constant and so the radial dynamics *must supply and receive the extra* angular energy. This comes out looking like a new radial pseudopotential that gives rise to the centrifugal force. The adiabatic invariant case is exactly analogous: we forget about the fast part's phase; the fast part's energy must change so as to keep the adiabatic invariant constant; this energy must come from the slow part of the system; and the net result is a new "pseudopotential" and corresponding pseudoforce acting on the slow system. We make this connection precise in section 2.9.2 and show that these "forgetting operations" are part of a procedure known as reduction.

In going to rotating frames of reference, there is a change to the Poisson structure, corresponding to the Coriolis force, in addition to the change in the Hamiltonian. In chapter 3 we give the underlying structure behind this and show how the method of variation of parameters extends this kind of change to an asymptotic setting. The simplification process which is applied to asymptotic systems with adiabatic invariants is therefore closely connected to the one applied to systems with exact symmetries amenable to the process of reduction. Constants of

the motion which result from symmetry are codified geometrically in the notion of the *momentum map* of a group action. The powerful setting in which this concept *makes* general sense is discussed in sections 2.5 through 2.7. Adiabatic invariants are the generators of asymptotic symmetries. This viewpoint is developed into a powerful perturbation method in chapter 5.

1.4.4. Thermodynamic Forces

The next connection is with thermodynamic systems. If one slowly varies the parameters of a thermodynamic system in equilibrium, there is again an adiabatically invariant quantity known as the *entropy*. The entropy change along a path in thermodynamic phase space can be defined as the integral of the one-form $(1/T)dU + (p/T)dV$ (where T is temperature, p is pressure, U is energy, and V is volume). The choice of coordinates $(1/T)$ and (p/T), thermodynamically conjugate to U and V, is absolutely essential to obtaining an adiabatically invariant integral. As in mechanics, this canonical one-form $(1/T)dU + (p/T)dV$ plays a fundamental role in the structure of thermodynamics. The net change in entropy in a cyclic process can be obtained by integrating the corresponding symplectic two-form

$$\omega = dU \wedge d(\frac{1}{T}) + dV \wedge d(\frac{p}{T}) \tag{1.1}$$

over any two-dimensional disc bounded by the loop representing the cycle in phase space. This thermodynamic symplectic structure is discussed in section 16.7.1.2.

Picture a slowly moving apparatus attached to a piston that compresses a gas confined in a cylinder. When we first study the system, we might think that we

have to keep track of the detailed dynamics of every molecule of the gas in order to understand the operation of our device. The adiabatic invariance of the entropy tells us, however, that we really need only the entropy of the gas (obtained by initially measuring the internal energy of the gas, say). As the piston moves, the energy of the gas will change so as to keep the entropy constant. This energy comes from our apparatus, and the effect of the gas is just like a nonlinear spring. We can forget the gas by introducing a new "pseudopotential" into the dynamics of the machine. This pseudopotential gives rise to *thermodynamic forces*. This then connects thermodynamic forces with adiabatic pseudoforces. A more detailed discussion of this connection is given in sections 2.9 and 5.4.4.

1.4.5. Wave Action

The next connection is with wave systems. First consider the example of linear sound waves in a closed room. The wave evolution may be expressed as a superposition of room eigenmodes, each of which are purely oscillatory at a corresponding eigenfrequency. What happens if we excite an eigenmode and then slowly vary the shape of the room? In an underlying infinite-dimensional Hamiltonian phase space for the waves, the eigenmode's evolution describes a closed loop (since it is periodic). Just as in the purely mechanical case, we may define the action of the oscillatory wave to be the integral of the symplectic form over a disc bounded by this loop. For linear waves, the action is again the energy of the wave divided by the frequency. As we slowly vary the room, the action of the eigenmode remains constant. Because the eigenfrequency typically varies as we change the room, this

forces the wave energy to change. This extra wave energy comes from the work done against the wave in altering the room.

One might think that our eigenmode could excite other eigenmodes as we vary the room. As long as the eigenfrequencies are distinct, this coupling is exponentially small in the slowness of room variation. For typical one-parameter room variations it turns out that the eigenfrequencies remain distinct. This phenomenon is sometimes called "level repulsion" and is due to the fact that the space of two by two Hermitian matrices (representing the possible couplings of two modes for all room parameters) is 4-dimensional, while the subspace of matrices with equal eigenvalues is only 2-dimensional (parameterized by the eigenvalue and the imaginary off-diagonal antisymmetric element). It therfore takes two parameters to force a degeneracy.

Imagine a slowly varying apparatus which, as it moves, changes the shape of the room containing the eigenmode. At first we might think that we need to know the phase of the mode to determine the evolution. Adiabatic invariance allows us to get by with just the wave action. The apparatus moves as if it had a new pseudopotential. When the wave is a light wave, the corresponding force is the *radiation pressure*.

The same idea may be applied to quantum wave-functions in a slowly varying potential. The classical limit of the wave result applied to a particle in an oscillator potential gives the adiabatic invariance results for mechanical systems that we discussed first. A square-well potential with slowly varying width corresponds to a particle bouncing back and forth between a stationary and a moving wall. The adi-

abatic invariant is the area of the rectangle described by the particle in (q, p) space (see section 2.9.3.1). Physically, the particle energy changes because the particle is bouncing off a moving wall and on each bounce comes away with a different magnitude of velocity than it had going in. The statistical analog is a one-dimensional gas and the adiabatic invariance is represented by the conservation of pV^{γ}. The wave system may be understood as a gas of quanta (for electromagnetic waves, a gas of photons), and a quantum changes energy in bouncing off a moving wall due to the Doppler shift. The de Broglie relation $E = \hbar\omega$ for a free massless quantum shows that the action E/ω is the number of quanta times $\hbar$.

The "normal modes" for a free wave are infinite plane waves. These have infinite action, but there is a sense in which they have a well-defined action *density* (i.e. action per unit of volume). Such a notion becomes asymptotically precise when we study eikonal waves, which are plane waves with slowly varying amplitude and wave-vector. Asymptotically, it begins to make sense to think of the energy and action of an eikonal wave as being made up of additive contributions from the different regions of space. In chapter 8 we will show how the asymptotic wave energy density and wave action density arise asymptotically. We obtain the evolution of the wave action density in time. In the presence of slowly varying potentials, the wave energy density varies but the wave action density evolves as a locally conserved quantity. The pseudo-force on the medium due to the giving and taking of wave energy is the ponderomotive force. These various relations give another connection between our subjects.

We have seen that the action in Hamiltonian dynamics and the entropy in

thermodynamic systems play very analogous roles. It is interesting that these are the two concepts that often give students the most trouble when learning mechanics and thermodynamics. Both are adiabatic invariants under slow variations of the parameters of a system. Because of this, they both give rise to the pseudoforces that are a characteristic consequence of eliminating degrees of freedom. They are both intimately related to a geometrical symplectic structure on the corresponding phase spaces. The action motivates us to introduce *canonically conjugate* momentum variables to the configuration space variables of a mechanical system. The entropy motivates us to introduce *thermodynamically* conjugate variables to the mechanical observables (like energy, volume, particle number, etc.) of a thermodynamic system.

1.4.6. Action, Entropy and Asymptotics

We know, however, that both classical mechanics and thermodynamics are asymptotic theories that are limiting approximations to quantum mechanics and statistical mechanics respectively. Can we understand the geometrical structure of these limiting theories as arising from the asymptotic process? Do the key quantities: action and entropy, have any natural meaning in the detailed underlying theories? The answers to both questions are resoundingly affirmative and much of this thesis is devoted to ferreting them out. The essential idea is that quantum mechanics associates to every path a wave amplitude whose phase is the action of that path. This very general action agrees with the classical mechanical action on the paths that represent real classical motions. These paths are defined asymptotically via the method of stationary phase and satisfy the *principle of least action*.

Statistical mechanics (as we formulate it here) associates a weight to each probability distribution. This is an exponential whose exponent is the *information entropy* of the probability distribution and is defined as $-\int p \log p \, dx$. This very general entropy agrees with the thermodynamic entropy on the distributions that represent real equilibrium thermodynamic states. These distributions are defined asymptotically via the method of steepest descents and satisfy the *principle of maximum entropy*.

1.4.7. Steepest Descents, Stationary Phase, and Averaging

The two main mathematical theorems which allow the asymptotic simplification are the method of *stationary phase* for the oscillatory and wave systems, and the method of *steepest descents* for the statistical systems. These methods connect line integrals in the complex plane of functions with an asymptotic parameter, to expressions that asymptotically depend only on the function in the neighborhood of certain special points. Dependence on the full details is reduced to dependence on only certain special features. Where does the extra eliminated information go? Its contribution is to things that have no long-term effect on the degrees of freedom we are interested in. In each of the domains studied here, we perform some kind of *averaging* that eliminates the features which have no long-term contribution. Physically, we are often interested in the interactions of our system with other systems (like ourselves) that respond only to these averages and so they are really the quantities of interest. The only way a fast degree of freedom can contribute is for it to have a long-term effect. This can come about only if its many contributions

add coherently. This in turn, can happen if the fast degrees of freedom repeatedly come back close to the same state and so give a net contribution to the slow scale behavior. In that case, our system has an approximate symmetry that maps one fast excursion to the next one.

1.4.7.1. Resonance

The regions of phase space where fast contributions add coherently are called *resonances* in oscillatory systems and *correlations* in statistical ones. Long-term effects are represented by phenomena whose frequency lies in a region near zero (and with appropriate scaling, the width of this region approaches zero in the asymptotic limit, leading to the omnipresent delta-functions in frequency). Nonlinearities can make use of rationally related frequencies to create near-zero frequency effects (the dynamics is then defined on tori whose orbits do not cover their surfaces densely); these effects are known as *resonant interactions*. One can even get them linearly if it is possible to simply add the frequencies to get zero (since then there is an oscillatory coordinate with zero frequency). These resonant effects are responsible for most of our knowledge of phenomena on time scales smaller than the one on which we normally operate (eg. spectroscopy in atoms, nuclei, particles, etc.).

1.4.8. The Key Examples in Mechanics, Waves, and Statistics

To give bearings, let us list some of the key examples whose features exemplify the ideas we wish to explore. There is overlap in the techniques of the three domains (and it is this overlap that we are particularly interested in), but roughly most models can be categorized as being one of the three types. In mechanics we have systems with exact symmetries, such as a planar particle in a rotationally symmetric potential, to which we can apply Noether's theorem. We also have systems with approximate symmetries to which the method of averaging, Lie transforms, Kruskal's method, and other perturbation techniques we shall discuss can be applied. Examples include: gyromotion, oscillation center motion, oscillatorily stabilized systems, and the interesting variety of systems with adiabatic invariants.

In wave theory, we have all the situations with short wavelength waves, including plasma waves, propagation of light rays, elastic waves, quantum mechanics, and the various nonlinear plasma and fluid waves. The methods of analysis include the classical WKB theory, its geometric extension by Maslov to handle caustics, and its extension to nonlinear waves due to Whitham.

There are two types of statistical models. One type is concerned with equilibrium systems. This includes models of thermodynamics using either the maximum entropy formalism or Khinchin's approach via the central limit theorem. This type also includes Brownian motion and fluctuation theory which we relate to the renormalization group. The other type of model studies non-equilibrium situations. These include the vast majority of plasma systems. The particular models of interest here are: the BBGKY hierarchy, Bogoliubov's derivation of the Boltzmann

equation, the fluctuation-dissipation theorem, the Onsager relations, and the techniques used in Landau damping and quasilinear theory. Hamiltonian structures are lurking behind virtually every aspect of the physics of these systems and will therefore be a prime consideration in our exploration.

1.4.9. Mechanical Systems

The original example of a Hamiltonian system is given by the dynamics of point particles in a potential. In plasma physics, one is interested in *charged* particles and their Hamiltonian dynamics in electric and magnetic fields. The three nontrivial but tractable situations one often needs to study are: the gyration of a particle in a magnetic field, the motion of a particle in an electromagnetic wave, and the scattering of particles off one another (usually all three are present at the same time but certain aspects dominate the particle's behavior). When there are exact symmetries, we may simplify the system by using reduction as we have discussed. Usually, however, systems have only approximate symmetries.

1.4.10. Mechanical Syetems: Separation of Time Scales

The asymptotics in particle mechanics usually arises from a separation of time scales. One set of degrees of freedom may have dynamics that is considerably faster than the others. In this case, the effect of the fast degrees of freedom on the slow ones tends to be close to the average effect and the effect of the slow degrees of freedom on the fast ones tends to be almost as if we were holding them fixed. If we introduce asymptotics, we may obtain in asymptotic expansion a model of our

system which has completely separated the time-scales of the slow and fast degrees of freedom. We may often understand this separation as being due to reduction by an approximate symmetry (which gets better and better asymptotically). As we have discussed, we then get a model of the slow system whose dynamics is altered by the presence of new pseudopotentials in the Hamiltonian and magnetic terms in the Poisson brackets. These remains of the forgotten fast degrees of freedom give rise to the new physics introduced by asymptotics.

1.4.10.1. Gyromotion and Asymptotics

In the case of a particle in a magnetic field, we let the asymptotics move the particle faster and faster around its gyration loops in comparison to the rate of motion of the center of these loops. The asymptotic theory of guiding centers, which forgets the fast gyromotion and describes only the slow drift of the loops, has the new concepts of: $E \times B$ drift, curvature drift, polarization drift, magnetic moment as an adiabatic invariant, bouncing from a magnetic mirror, etc. None of these concepts makes precise sense for the physical system as it appears in nature. Nonetheless, they have been extremely important in the design and understanding of plasma devices and represent truly new *physical* notions in the given setting. To make the model of these useful concepts precise, we *must* introduce asymptotics. Note that this is far more significant than the usual idea of perturbation theory as giving better and better approximations to some underlying exact model. While it does that, I believe its real importance is in the *conceptual* advances it allows one to make in understanding and utilizing physics.

One often sees (particularly in older literature) the stated goal of perturbation theory as convergent series and the concomitant lament that most physically useful series are only asymptotic. From the conceptual viewpoint we emphasize here, convergence is irrelevant. We never use more than a finite number of terms of our series (usually the first order terms suffice to give the new physics), the underlying model equations we are approximating are based on concepts that themselves are only asymptotic, and convergence is a complex analytic concept requiring apparently physically irrelevant complex structures in our models. Usually we only require our transformations to be smooth and a smooth, non-analytic transformation can destroy the covergence of most series (a classical construction due to Borel uses C^∞-bump functions to construct a smooth function whose derivatives at a given point are equal to the elements of any arbitrary sequence of real numbers; using such a function to change coordinates will convert an analytic function whose first derivative is non-zero into one whose derivatives are any desired sequence; we need only choose one which grows fast enough to prevent convergence of the Taylor expansion at any radius). What I feel is important for theoretical understanding is not the actual *numbers* but rather the physical *tendencies*. Of course in applications one wants numbers, but asymptotic series often yield a given accuracy of approximation to the exact solution with many fewer terms than any corresponding convergent series.

1.4.10.2. Oscillation Centers and Ponderomotive Forces

In the case of a charged particle moving under the influence of a high-frequency wave whose amplitude slowly varies in space, we introduce the asymptotic concept of the oscillation center. As the wave oscillates, the particle feels a force first in one direction and then in the other, causing it to oscillate as well. If the wave were spatially uniform, then the excursions to each side would exactly cancel each other, yielding no net average effect. In a non-uniform wave, the particle feels a bigger force in the region of larger amplitude and so is pushed more away from such regions than into them. The average push behaves like a force on the particle pushing it away from higher wave amplitude regions. The reduced description gives the dynamics of the center of oscillation and includes a remnant of the fast motion through this so-called *ponderomotive* force. Both the guiding center equations and the ponderomotive equations are Hamiltonian.

The second chapter shows how to use reduction by a circle symmetry in conjunction with the method of averaging, to obtain the $E \times B$ drift dynamics as a Hamiltonian system. The higher order theory, based on the ideas of Kruskal, is discussed in chapter 5, and the relation to the pioneering work on the Hamiltonian structure of gyromotion due to Robert Littlejohn is discussed in chapter 6. Many other systems fit into the setting of these asymptotic methods, and the physical concepts that come out of the asymptotics can be quite striking. Systems with constraints can exhibit phenomena analogous to ponderomotive effects, due to oscillations in the constrained direction. One may understand the oscillatory stabilization of systems like the shaking inverted pendulum, the shaking inverted cup

of fluid, or the r f stabilized unstable MHD modes, as the effect of the asymptotic pseudoforces.

1.4.10.3. Geometric Perturbation Theory

To make these notions precise, we have developed a systematic geometric framework for asymptotology based on notions from the mathematical theory of *jets*. Because we are interested in systems whose phase spaces are manifolds and we particularly want to study the asymptotic behavior of geometrically intricate Lie group symmetries, we develop the various perturbation theories in the coordinate-free language of modern differential geometry. While often useful (particularly in explicit calculations), coordinates are unphysical and often obscure the underlying content of physical theories. The movement in the mathematical physics community to re-express the fundamental physical notions without arbitrary coordinate systems has been very successful in mechanics and has succeeded in dramatically simplifying some of the classical results (eg. Liouville's theorem, Noether's theorem, Darboux's theorem, action-angle variables, etc.), clarifying the essential structure of the theory (especially the extreme importance of symplectic geometry), and obtaining many new ideas and results (eg. the extension to infinite dimensions, Noether's theorem for arbitrary Lie groups, the KAM theorem, chaotic dynamics, etc.).

Our formulation of perturbation theory is in the spirit of this movement, and so we summarize some of its key ideas in chapter 2. In that chapter we also give the non-Hamiltonian aspects of this geometric approach to non-secular perturbation theory and sketch the Hamiltonian results. The Hamiltonian results are given in

detail in chapter 4. There we show that the non-secular perturbation dynamics is itself Hamiltonian when the underlying dynamics is. The Poisson bracket has a somewhat unexpected form, in that it pairs the lowest order variables with the highest order, the next to lowest with the next to highest and so on. We show that it is the natural structure in five different ways, each of which sheds a different light on the relationship between the perturbed and unperturbed systems. The relationship between symmetry and perturbations is also given in that chapter, and the operation of performing reduction is shown to commute with the operation of performing a perturbative analysis. Chapter 5 uses the previous non-secular results to do secular perturbation theory. We develop a new technique based on these ideas which is simpler to apply in practice. Its application to guiding center motion is given in chapter 6.

The limitations of this and all general secular perturbation theories do not appear to be well known in the physics community. We therefore give examples and explanations of why the time of validity of this and other theories is only of order $1/\epsilon$, even though the accuracy over this time can be to all orders in ϵ. We also discuss the case of many fast frequencies and the concomitant resonances. Finally, we end chapter 5 with the case in which the fast motion is ergodic on the energy surface and begin to make the connections with statistical mechanics.

1.4.11. Averaging in Statistical Systems

There is an interesting way in which the need to average over an intermediate scale becomes apparent in both the statistical and wave systems. Consider the common and very useful result that says that independent measurements of a definite quantity representative of a complex system will be distributed as a Gaussian. The very general argument for this assumes only that small errors from the many parts of the system will contribute additively to the error in the measured quantity. Regardless of the how the individual errors are distributed (under some very weak constraints), the central limit theorem tells us that their sum will be distributed as a Gaussian. Let us try to understand how one applies this statement operationally and so see that a specific type of averaging is required to make sense of the notion of Gaussianity. Let us consider an actual experiment where we have made, say, 1000 measurements of some quantity (so as not to obscure the argument, let us assume that the results of our measurement are precise real numbers). Taken directly, our measured probability distribution is a sum of δ-functions, one for each measured value (no Gaussianity here!). In practice, we "bin" the measurements, i.e. we make a histogram of the number of measurements that fall into each of a set of intervals that partition the space of measurement values. If we make the bins too small, we get the problem of the δ-distributions, i.e. widely spaced bins with one measurement each. If we make the bins too large, then all measurements fall into a single bin. We see something that approximates a Gaussian only if we bin on an intermediate scale. This is defined using an asymptotic parameter given by the number of measurements.

as the number of measurements gets larger, we can make our bins get smaller and it is in this asymptotic sense that we say our distribution becomes Gaussian. If we make N measurements, it is easy to see that the mean spacing between measured values goes as $1/N$. If we bin on a scale that goes to zero more slowly than this, like $1/\sqrt{N}$, then asymptotically there will be an infinite number of measurements in each bin. The law of large numbers tells us that asymptotically the number of measurements falling into each bin will agree (with probability one) with the number expected from the Gaussian distribution. Furthermore, since the width of the bins is going to zero, we get arbitrarily fine accuracy. Also note that, while with any given number of measurements the observed distribution may be changed by altering the binning, as long as the binning is on an intermediate scale, the asymptotics is bin independent. Furthermore, the range of bin choices that yield values close to the asymptotic result gets wider as N gets larger. Another way to think of the asymptotics is in terms of the convergence of the delta-function distributions *representing the measurements, to the smooth asymptotic distribution*. From this perspective, the binning procedure defines a topology on the space of distributions.

It is only in this asymptotic sense that the whole notion of a probability density makes physical sense at all. There are many other situations in which a similar asymptotic averaging is behind the models. In plasma physics we define a distribution function on phase space by averaging the Klimontovich δ-distribution to obtain the smooth Vlasov distribution (i.e. we bin the particles). We make precise sense of the manipulations we are allowed to perform on this distribution only by considering an asymptotics which chops the particles into finer and finer pieces. Similar

averaging is behind fluid descriptions and macroscopic electrodynamics. In each of these subjects we often come across δ-distributed quantities (eg. point masses and charges). These are to be interpreted in terms of the asymptotics. The real situation is not at the asymptotic limit, but the object of interest is small on the large scale (and we scale it so it gets ever smaller asymptotically) and has finite mass (or charge, etc.) on the small. We then develop consistent rules that are asymptotically valid and these lead to the familiar calculus of δ-functions and other singular distributions. Similarly, real fluid velocity functions (which are the result of an average over a macroscopically small region with a large number of particles) cannot validly have wavelengths shorter than or on the molecular scale. We model the velocity evolution by nonlinear partial differential equations, however, that can (and do) excite arbitrarily short wavelength Fourier components. If these ever become important, the separation of scales fundamental to our model has broken down and the model becomes invalid.

1.4.11.1. Matched Asymptotics

A fundamental technique of singular perturbation theory is to insert models which include the physics of the small scale in regions of breakdown. For example, in very high shear regions of a fluid, one need to include more kinetic effects than are represented in the simple Navier-Stokes model. The two important cases arise when these singular regions (with asymptotically small scale physics) are: 1) localized and get smaller with the asymptotics and 2) spread over open regions but have a local homogeneity. In the first case, one inserts a so-called boundary layer, which

is infinite on the small scale, but infinitesimal on the large scale (i.e. its size scales asymptotically between the inner and the outer scale). One then performs the inner and outer analyses separately, with the boundary condition arising from the matching of the asympotics of the inner region at infinity with the outer region at zero. The case where the small scale effects are not localized, but are homogeneous, leads to the eikonal techniques we shall discuss next, in which we have slowly modulated fast behavior which is regular on the small scale at each large scale point (eg. local plane waves).

1.4.12. Averaging and Local Fourier Transforms

The idea of averaging over an intermediate scale that arose in trying to understand the notion of probability density also arises in trying to define the frequency or wavelength of a slowly varying wave. When we talk about a wave having a certain wavelength or a sound having a certain frequency, we are always talking about a *local* Fourier transform. No sound lasts forever and we usually are not interested in the properties of a wave in the next galaxy. What do we mean by a local Fourier transform? Operationally we usually work with the Fourier transform of a signal multiplied by a window function.

$$F(t,\omega) = \int_{-\infty}^{\infty} f(t-\tau)W(\tau)e^{-i\omega\tau}d\tau. \tag{1.2}$$

We might have a spectrum analyser that works on a finite segment of the signal (and smoothly damps the on and off edges of the window to minimize spurious edge effects). As in the case of binning, the shape and size of the window function $W(\tau)$

affects the resulting local Fourier transform. If we make the window too wide, we get contributions from parts of the signal we aren't intersted in (eg. the spectrum of the news broadcast after the concert) If we make the window too small, we don't sample enough wavelengths (perhaps not even one) to get a good fix on the frequency (turning the radio on and off quickly results in a click with all frequencies). If we are trying to say that a frequency (or frequencies) is present at a given time (as in describing a piece of music in music notation), then that frequency should be distinctive To make precise the manipulations allowed, we embed the given wave in an asymptotic family, which makes the frequency of interest more and more distinct as a parameter approaches its limit. We make the number of wavelengths occuring in a region of significant wavelength change become infinite by making the scale length longer for fixed wavelength, the wavelength shorter for fixed scale length, or both. We introduce a short scale given by the wavelength and a long scale given by the scale length defined so that asymptotically their ratio vanishes. The operations and concepts we are allowed to use in describing our wave are those which make asymptotic sense. As the parameter approaches its limiting value, the domain of validity of such asymptotic concepts gets larger and larger. For any real situation, one must make sure that one doesn't have physics which violates the separation of scales and thus the validity of this kind of model.

The local Fourier transform above can be made precise asymptotically. We let the window function W scale with the asympotics between the fast and slow scales. On the slow scale it looks more and more like a δ-function and in the asymptotic calculus we may treat it as such. On the fast scale it looks more and more like

a constant and so the local Fourier transform asymptotically looks locally like the Fourier transform whose useful relations then carry over to the asymptotic calculus.

1.4.12.1. Scales in Differentiation

Many other areas of study have this same large vs. small scale dichotomy. Robert Littlejohn suggested the example of optimal algorithms for numerical differentiation (see [Stoer and Burlirsh, 1980]). Assume we have some function represented in a computer as an algorithm that can calculate values to some accuracy. What is the optimal way to calculate its derivative at a point numerically? We can evaluate the function at two nearby points, take the difference and divide by the distance between the points. How do we choose this distance? If it is too wide then we won't get the derivative at the point of interest, but rather something averaged over a region in which the function may have signifcant change. If it is too small, then the division by a small number blow up the errors until they are arbitrarily large. Given some criterion of goodness in these two respects, there is an optimal distance. We identify the asymptotic limit with increasing the numerical accuracy of the calculation. If we take the distance between the points of evaluation to shrink on an intermediate scale (i.e. they become infinitely close as far as the variation of the function is concerned, but are far enough apart that the division algorithm becomes infinitely accurate asymptotically) then in the limit we get the actual derivative. This idea may even be used to define the derivative (and is essentially the same as the mathematical definition, given that reals are defined as limits of rationals and the arithmetic operations are defined in terms of the precise

operations on rationals). The very same issues are relevant in any physical situation in which we take a derivative, and this kind of approximation is almost always what we really mean (eg. the strain is the derivative of the displacement in elasticity, but displacement as a smooth function is an asymptotic idea as the molecules of material become more and more numerous).

1.4.13. Symplectic Asymptotics in Thermodynamics and Mechanics

The last part of the thesis discusses the relationship between eikonal wave asymptotics and the particle number asymptotics of statistical mechanics. Both theories reduce to asymptotic theories with natural symplectic structures: classical mechanics in the case of waves, and thermodynamics in the case of statistical mechanics. The description of a state is given by a submanifold of these asymptotic state spaces in both cases. It turns out that this submanifold of states always has a very special relationship with the symplectic structure (the symplectic form actually vanishes when restricted to the submanifold). Such submanifolds are called Lagrangian submanifolds and give rise to a rich theory (see section 7.1.4). The role of the Fourier transform in wave theories is played by the Laplace transform in statistical mechanics. The simplification provided by the method of stationary phase in wave mechanics is provided by the method of steepest descents in statistical mechanics. The role of the action is played by the entropy. Canonically conjugate variables correspond to thermodynamically conjugate variables. The analog of the Heisenberg uncertainty principle in the asymptotic limit is the fact that the dispersion tensors of thermodynamically conjugate variables are inverses of each other.

Non-degenerate critical points give the rays in ...ve mechanics and the Gaussians in statistical mechanics.

Heisenberg, and Fourier before him prevent us from localizing states in x, k space. But, by letting k go to infinity, we can make the *relative* dispersions in both x and k go to zero (by relative dispersion in k we mean $\Delta k/k$ where Δk is the absolute dispersion). So by rescaling (i.e. going to slow variables) we get states whose local Fourier transform is asymptotically a δ-function in phase space.

Given a definite mean energy U (or any other extensive quantity from the underlying mechanical system) the maximum entropy formalism gives an exponential distribution parameterized by the conjugate variable (inverse temperature β in the case of energy). A definite temperature corresponds to a (canonical) distribution of energies. A definite energy corresponds to a distribution of temperatures. We may use Bayesian statistics to see that the temperature and energy distributions are related by essentially a Laplace transform. The analog of Heisenberg's principle says that we cannot localize a distribution in U, β space. If we let U go to infinity, however, we can make both *relative* dispersions go to zero. By rescaling, we get an asymptotic delta-function on thermodynamic phase space.

Eikonal waves have k's at each x that fit together into a Lagrangian submanifold. This manifold is locally the graph of the differential of wave phase, which for mechanical systems is the action. If we view k as the base coordinate, this manifold is the graph of the differential of a function of k that is the Legendre transform of the phase. This Legendre transform results from applying stationary phase to the Fourier transform of our eikonal wave. Asymptotically we obtain the Legen-

dre relation between functions of x and of k because only the stationary points contribute.

The state of an equilibrium statistical mechanical system is restricted to lie in a Lagrangian submanifold in thermodynamic state space. For simplicity, let us just consider (E, β) space (where we have introduced the inverse temperature $\beta = 1/T$). This manifold is locally the graph of the differential of the entropy S as a function of U. In other words, only those points in (U, β) space which satisfy

$$\beta = \frac{\partial S}{\partial U} \tag{1.3}$$

correspond to thermodynamic states. To write this as the graph of the differential of a function of β, we again introduce the Legendre transform. This Legendre transform results from applying steepest descents to the Laplace transform of our distribution function. Again only stationary points contribute asymptotically. The usual Legendre transforms of thermodynamics are of the energy instead of the entropy. These give the Helmholz and Gibbs free energies and the enthalpy. They arise from taking the entropy instead of the energy as the observable extensive variable in phase space. The two pictures are related to a large one containing the extensive and conjugate intensive variables and the entropy. This has a natural contact structure given by the first law of thermodynamics (conservation of energy). The equation of state is a Legendre submanifold with respect to this contact structure and turns into a Lagrangian submanifold when we project onto either of the natural symplectic thermodynamic state spaces.

We may understand the relation between the underlying infinite dimensional wave and probability density spaces and the resulting symplectic phase and ther-

mo s namic spaces better using a theorem about Lagrangian manifolds. The details of this theory are given in section 16.7.3. We are given a fibration of one manifold over another (i.e. a projection so that the inverse images of points all look the same) and a function S on the first manifold. The graph of the differential dS is a Lagrangian submanifold of the cotangent bundle of the first manifold. The push-forward of this submanifold to the cotangent bundle of the second manifold consists of solutions to a constrained variational principle, that is, it is the push-forward of extrema of S *restricted to each of the fibers*. From a more general result, one may show that this resulting pushforward is Lagrangian iff all of S's critical points are non-degenerate (i.e. $\delta^2 S$ is invertible). The more general formulation is: if we are given a map from one space to another one, then a Lagrangian submanifold of the cotangent bundle of the first space pushes forward to a Lagrangian submanifold in the cotangent bundle of the second space if and only if it intersects the pull back of the second cotangent bundle transversally. In the case of a fibration and a Lagrangian submanifold given by the graph of the differential of a function, the extrema of the function restricted to each fiber represent the intersection with the pull-back and so push forward to a Lagrangian submanifold (non-degeneracy corresponds here to transversal intersection). For the case of waves, we consider the space of paths originating on some source region. The image space is $\Re^3$ and the projection sends a path to its final endpoint. The fiber over a point in $\Re^3$ consists of all paths that end at that point. For the function S on path space, we take the action along each path. In the Feynman path integral formalism, this function is the phase associated with each path. The differential dS defines a Lagrangian

submanifold in the cotangent bundle of path space. Doing the Feynman integral gives us a wave on $\mathbb{R}^3$ whose phase is determined at each point by stationary phase to be that of the paths of extremal action, where we consider variations restricted to the fiber (i.e. to paths that end at the point of interest). This is exactly the pushforward of the Lagrangian submanifold dS and gives the Lagrangian submanifold in the space of (x, k)'s (i.e., the cotangent bundle of $\mathbb{R}^3$) which represents an eikonal wave. For statistical mechanics, our first space is the space of probability distributions on phase space. The entropy of a distribution (defined by the integral of $-p \log p$) is a function on this space. The map to a space of observables given by the mean value is a fibration. The maximum entropy formalism says we extremize the entropy on the fiber (i.e., all distributions with the given mean values) and the theorem says that this determines a Lagrangian submanifold in the cotangent bundle of the observables (i.e. thermodynamic state space).

In wave dynamics, we start with a theory of arbitrary waves, and specialize it to the important special case of eikonal waves. To make the ideas of this specialization precise, we introduce an eikonal parameter that specifies the separation of scales between the local wavelength and its slow variation. We may either think of the wavelength going to zero for given scale length, or (as I prefer, since the actual wavelength often has physics in it) the scale length going to infinity for given wavelength. The waves we want to study the dynamics of are viewed as asymptotic families and we utilize the asymptotology to introduce new physical concepts which apply only approximately to the real waves, but show the *tendency* of the behavior as the separation of scales becomes greater. The key simplifying idea in the calculus

of asymptotic waves is the method of stationary phase. Integrals (such as those in the description of wave propagation) which really depend on the entire state of the wave, asymptotically depend on the state only in the local neighborhood of a point. The key asymptotic information about the wave at a point is the wave-vector k and we obtain "rays" in (k, x) space, to which the dependency of a portion of a wave is restricted asymptotically. This is a notion which is precisely defined by the spatial Fourier transform in the case of plane waves, but that has only asymptotic meaning for our eikonal waves. The asymptotic local Fourier transform makes precise the notion of a wave whose wave-vector spectrum depends on x.

1.5. A Hundred Further Questions, Conjectures, and Suggestions

And time for all the works and days of hands
That lift and drop a question on your plate;
Time for you and time for me,
And time yet for a hundred indecisions,
And for a hundred visions and revisions,
Before the taking of a toast and tea.

-- from *The Love Song of J. Alfred Prufrock* by T. S. Eliot

This section presents a number of questions and suggestions for further work related to the topics covered in this thesis. Some of them appear to be fairly straightforward and some appear to be quite difficult. They are presented in the order the subjects appear in the thesis and we make reference to the relevant sections for each question. **1.** Answer the question posed in section 2.1: Does the KdV Poisson bracket naturally arise from the one for the Boussinesq equations?

2. Is there a natural way of working directly with the germs of paths, and can one use this to get information on exponentially small effects in the perturbation parameter ϵ (such as tunneling)? (section 2.3.2)

3. Does Newton's approach to centrifugal force, outlined in section 2.6.3, extend to any other situations?

4. Find physical systems that utilize the KKS symplectic structures on the spaces of measured loops and measured Lagrangian submanifolds introduced in section 2.7.10 (see for example the eikonal wave systems discussed in chapter 9).

5. Use the symplectic structure on some of the coadjoint orbits of the group of

symplectomorphisms not discussed in section 2.7.10 (eg. some of the "water-bag" models in plasma physics naturally live in these orbits).

6. *Apply the J-jet structures of chapter 2 to a variety of problems.*

7. Can one extend the Arnold stability method from fixed points to periodic orbits (or even orbits which limit on compact objects such as tori) by adding Casimirs to make the Hamiltonian quadratically maximal or minimal on the orbit in question?

8. Richard Montgomery has shown that there is a natural Poisson structure on the space of 2-jets of a Poisson manifold (see section 4.8.6.5).The construction given in section 2.8.1 "explains" the fact that the linearization of a symplectic Hamiltonian system about a fixed point is Hamiltonian. Can we apply the same construction to linearize a Poisson Hamiltonian system about a fixed point? (To linearize at a point in a symplectic leaf, one need only add Casimirs to eliminate any linear piece in the Hamiltonian, take the Poisson structure at the fixed point and the quadratic piece of the Hamiltonian. To see that there are problems near bones, just consider the Lie Poisson bracket on the dual of the Lie algebra of the rotation group. This Poisson structure vanishes at the origin, yet the linearized dynamics can be non-trivial. The approach suggested here via the jet brackets may give a solution even at the bones.)

9. Apply the linearized structures of section 2.8.1 and the previous question to any of the numerous physical systems in which one linearizes about a fixed point (*eg. the linearized* Vlasov equation about thermal equilibrium, linear wave theory in fluid mechanics or elasticity, linear surface waves, etc.).

10. Apply the linearized Hamiltonian structure about a given orbit (as opposed to a fixed point) described in section 2.8.1 to physical situations (eg. study the linear evolution of small perturbations about a nonlinear wave solution. Jerry Marsden is applying this to general relativity).

11. Use the same techniques as in the previous question and section 2.8.1 to understand the Hamiltonian nature of the lift of Hamiltonian dynamics to the symplectic frame bundle which has arisen in Robert Littlejohn's and Yukkei Hui's work on extending coherent states by including metaplectic corrections (see also section 7.1.3.2).

12. As in section 2.8.1, extend the previous four questions to Jth order structures (as opposed to just linearizing) and apply to examples where higher order effects are important.

13. Use the approach to oscillatory stabilization discussed in section 2.9.3.2 to study various r.f. plasma stabilization schemes in a Hamiltonian manner.

14. Use the connection between averaging and reduction to treat more problems like that discussed in section 2.10 in a geometric way.

15. Use the method of averaging given in section 2.10 to find the "pseudopotential" felt by a deforming elastic body due to the presence of elastic waves (i.e. the elastic analog of ponderomotive forces- as we deform the body the wave spectrum changes, causing energy to be transfered between the body and the waves).

16. Use the method of averaging given in section 2.10 to study the Stokes' drift of a fluid particle under the influence of surface waves (fluid particles move in approximate circles in the presence of small amplitude surface gravity waves in

a fluid: the circles are big near the surface and get smaller with depth; a particle then moves on a larger circle near the top than near the bottom of its orbit, just as in gyromotion the resultant particle path has a drift parallel to the surface of the fluid).

17. Use the connection between averaging and reduction given in section 2.10 to prove the $K - \lambda$ theorem relating the linear susceptibility of a plasma to waves and the ponderomotive pseudo-potential in the plasma dynamics due to the waves after reduction.

18. The explicit calculation technique of section 2.10 utilized a section of the circle bundle defined by the approximate symmetry. Develop explicit techniques for treating nontrivial bundles (these are covered by the abstract theory). An example might involve a perturbed rigid body where the circle action is rotation about a given body-fixed axis the nontrivial projection to the orbit space is the Hopf map.

19. Can one treat the conversion from a Lagrangian to an Eulerian description of a fluid with the concomitant introduction of convective terms as an example of the change of reference frame operations introduced in chapter 3?

20. (A generalization of the previous question): Can one extend the results of chapter 3 into a general theory connecting the process of reduction and the process of changing reference frames (this is carried out explicitly in section 3.1.6 for the Coriolis force)?

21. Apply the methods of section 3.1.6 to develop natural Poisson structures for a variety of rotating systems. Debbie Lewis is currently studying the infinite dimensional fluid dynamics of rotating liquid drops (such as stars or nuclei from

this perspective. Another example of interest might be a cold non-neutral electron plasma in a cylindrically symmetric "tin can" with axial magnetic field the electrons $E \times B$ drift around the axis making a naturally rotating reference frame.

22. Use the Hamiltonian structure introduced into the method of variation of parameters in section 3.2 to redo in a geometric way physical derivations based on it (eg. the usual derivation of Fermi's golden rule in quantum mechanics).

23. Does the natural symplectic structure introduced in section 3.2.1 on the group of (compactly supported) canonical transformations of a symplectic phase space have any other physical applications?

24. The perspective on Jth order non-singular perturbation theory that is taken in chapter 4 is to view the perturbation dynamics as an ordinary vector field on a perturbation extended phase space (i.e. the jet space). Another perfectly valid perspective is to introduce J-jets of vector fields on ordinary phase space, whose flow is a "J-jet of a *diffeomorphism*" and so on for the rest of the objects in a theory. Pursue this alternate route and redo the various calculations of chapter 4 this way. Are there advantages or disadvantages to one or the other of these two perspectives?

25. Does the binary notation for iterated tangent bundles which was so convenient in section 4.5 have any other natural applications?

26. The technique of going from the path space bracket to the jet bracket introduced in section 4.6.1 by quotienting out the degenerate piece is very powerful. Are there other applications of this? In this example when one tried to push forward the original bracket along the projection, one obtained products of delta functions

and other very singular things. Similar singular brackets have been found by Bruce Boghosian in attempting to find a Hamiltonian linearization for the Poisson-Vlasov equation. Can one use a similar solution in that situation?

27. The sheet quotient spaces introduced in section 4.7 extend to Jth order a construction given to first order in [Kijowski and Tulczyjew, 1979]. In that reference they extend this construction in another direction: by making the sheet parameter be larger than one-dimensional. (In their work they were interested in time as the parameter and the extension is to space-time and field theories). Can one apply the extension to J-th order derivatives given here in the field context?

28. As in the previous question, can one introduce more than one parameter but now treating them all as perturbation parameters. Thus one should redo all calculations in chapter 4 replacing paths by higher dimensional submanifolds, and path jets by jets of maps of these higher dimensional submanifolds into phase space. Use this extension to treat perturbation theories with more than one small parameter. In particular this should give a context in which to explore various relative scalings of multiple parameters. (All the constructions including jet groups and Lie-algebras appear to extend in this way.)

29. Can one extend the perturbation structures developed in chapter 4 to Poisson manifolds, as opposed to symplectic manifolds? (This is important for application to many of the physical systems of interest.)

30. In particular do the jet spaces of Poisson manifolds inherit a Poisson structure from the Poisson structure on the iterated tangent bundles discussed in section 4.8.6.5?

31. As asked in section 4.8.6.5, does the Lie Poisson bracket on the dual of the Lie algebra of $J-jets$ of paths in a Lie algebra g agree with some prescription for extending g^*'s Lie Pois bracket to the space of its $J-jets$? (This is important since many physical systems have Hamiltonian structures derived from Lie-Poisson brackets.)

32. Can one explicitly write down the "magnetic terms" in the symplectic structures on the jet coadjoint orbits defined in section 4.8.7?

33. What is the relationship between the Lie algebra of jets of paths in a Lie algebra (which arises in perturbation theory) and the Lie algebra of jets of functions on the dual of the Lie algebra which is described in section 4.8.7 (and is useful in many contexts such as geometric quantization)?

34. Use the extension of Kruskal's perturbation technique given in section 5.2 to analyse a variety of physical systems.

35. Implement the explicit algorithm given in section 5.2 on a symbolic manipulation program such as MACSYMA or SMP and carry out any of the calculations of the last question to arbitrarily high order.

36. Combine the new Kruskal method with its well defined operations with a change of coordinates (as used in Lie transforms) to obtain a precise method that is easy to carry out by hand. (section 5.2)

37. Can one find a coordinate-free interpretation of two-timing or the method of multiple scales? Are there situations to which this method applies which do not fall under Kruskal's method? (section 5.4.1.2)

38. Reexamine the attention given to the $1/t$ validity of Kruskal's method in

various situations where it is used (for example in the variation of the adiabatic invariant in magnetic mirrors). If the adiabatic invariant appears to be invariant for times longer than $1/\epsilon$ find out why and develop a new perturbation method based on the reason. (section 5.4.2)

39. Find physical situations with more than one fast frequency in which Arnold's notion of "almost adiabatic invariant" (discussed in section 5.4.3) is really physically appropriate and apply the method of averaging keeping track of the measure of the "trapped particles".

40. Study [Ott, 1979] and [Kubo et al., 1965] and develop a precise form of the argument sketched in section 5.4.4. Ott claims to be able to get an expression for the variation of the adiabatic invariant. Can his argument be made invariantly and what are the limitations on its validity?

41. Section 5.4.5 shows that the reduction for finite perturbations may not be unique. What are the physical consequences of this fact and what does it say about the result of choosing different asymptotic scalings away from the unperturbed system?

42. Chapter 6 discusses the importance of making the unperturbed system consist of only periodic orbits when one wants to do singular perturbation theory. Many derivations in the literature do not have such an unperturbed system (some do not even have unperturbed dynamics). Can these systems be converted (say by a change of coordinates) to systems where secular perturbation theory is valid? If not, is the perturbation valid for time $1/\epsilon$? If it is valid, is there a fundamentally new perturbation method hidden in the derivation?

43. Extend the analysis of gyromotion in chapter 6 to three-dimensional, time-varying magnetic geometries.

44. Can one understand the Hamiltonian structure of gyrokinetic equations as reduction by a "gauged" circle action? (The way the formal structure of gyrokinetics arises from the single particle picture in chapter 6 is very reminiscent of the way the eikonal theory of chapter 8 is related to strictly periodic waves.)

45. Section 7.1 sketches heuristically the connection between Kruskal's secular perturbation theory and WKB theory. Can this connection be made precise using the coordinate-free formulation of chapter 5?

46. Develop an analog of the local Fourier transform of section 7.1.1.3 that is appropriate for nonlinear wave systems (i.e. given the wave family, reproduce the expression in terms of periodic solutions with slowly varying parameters).

47. Apply the variational approach to the Heisenberg uncertainty principle given in section 7.1.3.1 to other inequalities (and so get new insights into them and perhaps new results).

48. Does the approach to coherent states in terms of momentum maps given in section 7.1.3.2 extend to larger groups (eg. the Heisenberg group semidirect product the metaplectic group) to give approximations to the wave dynamics which are better than classical mechanics (eg. include the dispersion of the Gaussians in addition to motion of their centers in phase space)?

49. The approach to coherent states given in section 7.1.3.2 does not need the base state to be a Gaussian. Can anything be gained by using "coherent states" which are the orbits of states other than Gaussians under the Heisenberg group?

50. Can one use the approach of section 7.1.3.2 to get a theory analogous to coherent states for nonlinear equations?

51. Does Maslov's theory extend to nonlinear waves? (section 7.1.4)

52. Can one treat tunneling in a geometric fashion? Does it make sense to treat wave phase space as a complex manifold and so treat evanescent waves? (chapter 7)

53. Can dissipation be incorporated into the geometric WKB picture? (chapter 7)

54. What is the time of validity of the WKB approximation? Is it $1/\epsilon$ as conjectured in section 7.3?

55. Are there asymptotic theories with asymptotic validity for longer times than WKB, perhaps in special situations? (section 7.3)

56. How does WKB theory's finite time of validity relate to infinite time concepts such as the eigenvalue spectrum and quantum chaos? (section 7.3)

57. Are there physical situations in which the example of dispersion given in section 7.3 that is not accounted for by WKB plays an important physical role?

58. Is there an analog of geometric diffraction theory for nonlinear eikonal waves? (section 7.3)

59. Apply the Hamiltonian techniques of chapter 8 to a physical system expressed in terms of a non-canonical Poisson bracket (such as the Maxwell-Vlasov system) to obtain results not available with Whitham's averaged Lagrangian technique. Can the inelegant V's of section 8.2.2 be expunged? (As Weinstein asks: "Ain't there no eleganter way to do it?")

60. What is the analog of caustics for nonlinear waves? Are they classified by something analogous to catastrophe theory? What are the corresponding "special functions" analagous to the Airy function or Pearcey's function? (Chapter 8)

61. Make the argument given in section 8.2.2.1 relating degenerate symplectic structures to Poisson manifolds with Casimirs into a general theory (a very similar situation occurs in the example of guiding center motion given in chapter 5: In a Poisson formulation we have a degenerate bracket perturbed by a canonical bracket. In a symplectic formulation we have a degenerate symplectic structure perturbed by a canonical one. In the Poisson case we get zero order dynamics but no unique choice of Hamiltonian or symmetry generator. In the symplectic case we get a unique Hamiltonian but no zero order dynamics. The first order piece somehow defines a natural association between the degenerate zero order structures since the bracket and symplectic structure are nondegenerate and are inverses of one another when ϵ doesn't vanish.)

62. Use the Lie Poisson structure obtained in chapter 9 to study a system where wave action density on phase space is important (for examples see [Dewar, 1972b]

63. Find examples of dissipative systems that sit inside Hamiltonian systems as invariant submanifolds as suggested in section 10.1 (John David Crawford and Bruce Boghosian have suggested that the dissipation in Landau damping is of this type)

64. Is it ever useful to convert dissipative systems to Hamiltonian systems with more variables as shown in chapter 10? (This does allow one to use variational

formulations for instance?

65. Find an example of the misapplication of Lie transforms in the literature in the same vein as section 10.3.2, demonstrate the unphysical consequences and correct the analysis.

66. Recently, Alan Weinstein has shown that the Poisson system given in section 10.4 is a special case of a very general construction. Given a Lie group acting on a Poisson manifold, consider the action on the product of that manifold with the cotangent bundle of the group (where the action on the cotangent bundle is the lift of left translation). One obtains a new bracket on the orbit space of this action by reduction. When the group is the real line acting by the flow of a vector field, this reduces to the example in 10.4. Use the more general brackets to understand other physical situations.

67. Can one use the (very clever) techniques used in Gromov's proof to get any insights into the physics of the projected area situations of chapter 11?

68. Can the intuitive argument given in section 11.2 for Gromov's theorem based on the uncertainty principle be made precise?

69. Can one make the connection between Bogoliubov's derivation of the Boltzmann equation via BBGKY and projected measures that is suggested in section 11.4 precise?

70. Can one design a better particle accelerator based on the abstract construction for shrinking projected measure given in section 11.4.2?

71. Can one place any constraints on the projected measures of regions under canonical transformations? For example, the construction in 11.4.2 for shrinking

measure in one direction of projection stretches it considerably in the other direction of projection. It is almost true that the sum of the information entropies of the projected measures is minimized for a subset which is a product. Is there some true statement of this type which would place classical constraints on one's uncertainty in measuring pairs of projected quantities?

72. Can you find finite dimensional situations where the time-reversible almost attractor introduced in chapter 12 is directly responsible for irreverible asymptotic behavior in a time-reversal symmetric situation?

73. Implement the approach to the three-wave interaction of eikonal waves discussed in section 13.3.6. Does this invalidate the usual approach in terms of infinite plane waves?

74. Can one explicitly analyze the Landau damping equations using the notion of an almost attractor in an infinite dimensional Hamiltonian system? (section *13.3.3*)

75. Section 13.3.5 introduced a very general mechanism for obtaining dissipative behavior from resonant Hamiltonian systems. Can one find an abstract setting in which the precise characteristics of systems exhibiting this phenomenon are stated? Can one analyse Landau damping with these techniques?

76. Many infinite dimensional linear physical systems have a continuous spectrum making analysis difficult (especially bifurcation theory). The "eigenfunctions" are often singular (for example the van Kampen modes in the Poisson-Vlasov system). Whether the spectrum is continuous or not is physically detectable only after an infinite time. Physically one often is not interested in "normal" modes but

rather eikonal behavior. The driven harmonic oscillator of section 13.3.5 has an infinite response to the resonant mode, making analysis intricate and non-physical. The eikonal analysis of section 13.3.5 does not suffer this defect. Can one apply a similar analysis to other problems of this type?

77. Can one analyze the dissipation due to bulk viscosity in a manner analogous to that used in section 13.4? (Bulk viscosity results from the time lag for the rotational degrees of freedom a gas to reach equilibrium with the linear degrees of freedom under compression. Imagine a gas in a cylinder with a piston which you quickly move in and out. As you push down the gas has a higher pressure than it would in equilibrium since less energy is in the rotational degrees of freedom than should be. When you pull the piston out, the pressure is less than in equilibrium because more of the energy is in rotational degrees of freedom than should be. The net result is that the work done on the downstroke is greater than the work returned on the upstroke: the gas absorbs energy and therefore appears dissipative. A very similar kind of thing is behind the dissipation due to the string.)

78. Can one use the method of Melnikov to find horseshoes in situations of plasma physics other than the situation studied in section 14.4 (eg. other magnetic geometries, particle driven by electromagnetic wave, wave dynamics, soliton dynamics)?

79. Can one use the known statistical properties of horseshoes studied in ergodic theory to explicitly do a statistical study of gyrokinetics near magnetic field reversals? (section 14.4)

80. Does the phenomenon of reinsertion in area preserving homoclinic maps

explain the complex behavior discovered in Holmes and Whitley, 1984 as one goes from dissipative to area preserving homoclinic tangencies? (section 14.5)

81. Can you find universal structure in homoclinic tangles and homoclinic bifurcations using renormalization group techniques? (chapters 14 and 15) There are many renormalization type mappings at work in homoclinic dynamics. The "cross-section" of a piece of stable or unstable manifold is a Cantor set (with corresponding scaling properties) and homoclinic tangencies occur when the stable and unstable manifolds' Cantor sets intersect. Each homoclinic tangency causes an infinite number of period doublings for which we know there is universal behavior.)

82. Study the fractals that arise in nature using renormalization. For example, a classic Cantor set (remove the middle third of an interval recursively) corresponds to a periodic orbit as we continuously blow up the scale. A fat Cantor set (eg. remove middle third, then middle ninth, then middle 27th, etc.; this type of fractal occurs as the set of parameter values at which hump map orbits are chaotic after criticality) asymptotically approaches the stable fixed point represented by a solid line. An undernourished Cantor set (eg. remove middle third, middle square root of 3rd, middle cube root of 3rd,etc.) approaches the stable fixed point representing a single point in empty space. The set obtained by removing the rationals from the unit interval and separating the two sides around each rational p/q by a distance $1/q^3$ (the KAM tori lie in phase space like the points in this set in an arbitrary perturbation of a Hamiltonian system away from criticality) scales at each point according to terms in the continued fraction expansion of the point (Liouville numbers which are easy to approximate by rationals approach a fixed point representing

an isolated point in empty space, numbers which are hard to approximate by rationals (such as Diophantine numbers) asymptote to the line segment fixed point, other numbers hop around in a complex manner and lead to chaotic renormalization orbits." (chapter 15)

83. Can one get estimates on the rate of approach to Gaussianity in central limit situations from the eigenvalues of the stable fixed point in the renormalization approach of section 15.3?

84. Can one use the renormalization group approach given in section 15.3 to derive the Boltzmann factor $e^{-\beta E}$ using renormalization?

85. Similarly, can one derive the standard probability distributions other than Gaussians (such as Lorentzians, Zipfians, Bradford's distribution, Lotka's distribution, Pareto distributions, $1/f$ noise, log normal distributions and the other "long tail" distributions so important in modern statistical physics) using renormalization but with a renormalization operator constrained by other criteria than having a prescribed normalization, mean, and dispersion? Do the renormalization eigenvalues give scaling rates for these examples? (section 15.3)

86. Some of the distributions of the previous question will correspond to unstable fixed points. Can we understand physical systems with statistics which behave according to one of these distributions for a long range of scales but asymptotically behaves differently, in terms of the renormalization orbit starting near the stable manifold of the corresponding fixed point, approaching the fixed point for a long time, and finally feeling the effect of the unstable manifold and getting attracted to a more stable distribution at some scale (i.e. after some number of applications

of the renormalization operator)? Can we use the relative values of the renormalization eigenvalues to predict the range of scales which are described by the fixed point? (An example of this type of behavior occurs in any of the extremely large number of systems well described by $1/f$ noise, this distribution is not normalizable and so must really turn into another distribution at some scale.) (section 15.3)

87. *Find "poor man" approaches to other renormalization calculations,* such as area-preserving period doubling, breakdown of circle maps, breakdown of KAM tori. (section 15.4)

88. Use the new formulation of maximum entropy given in section 16.2.3 in terms of an integral over all probability distributions to get new insights into statistical mechanics, just as the path integral formulation gave new insights into quantum mechanics [Schulman, 1981].

89. Nothing in the maximum entropy formulation (section 16.2) either requires or disallows time-dependence of the studied quantities. [Jaynes, 1983] derives some aspects of non-equilibrium thermodynamics from the maximum entropy criterion. Can one use this derivation and the symplectic structures introduced in chapter 16 to obtain a symplectic analysis of non-equilibrium thermodynamics?

90. The Onsager relations have been discussed in terms of Lagrangian submanifolds in [Abraham and Marsden, 1978]. Can this discussion be given a fundamental basis using the ideas of the last question?

91. Allan Kaufman has discovered many non-equilibrium systems whose evolution is governed by a Lie Poisson bracket plus a so-called dissipative bracket which is symmetric. The generator of the dynamics becomes the Hamiltonian plus the en-

tropy and in some cases the action plus the entropy, which may be connected with the deep relations between these quantities discussed in chapter 16) Can one derive the dissipative bracket formulation using maximum entropy? (There appears to be a perturbation expansion in the strength of dissipation involved since the Onsager relations appear naturally.)

92. Can one introduce "super-reduction" which starts with a large-dimensional underlying system, performs reduction by an exact symmetry in some variables, reduction by an approximate symmetry in some others, and reduction by a "statistical symmetry" (via maximum entropy) in some others, and end up with the standard physical models in plasma physics, gas dynamics and fluid mechanics?

93. Based on the results of the last few questions, can one obtain a geometric formulation of the fluctuation-dissipation theorem? Is this a statistical analog of the $K - \chi$ theorem?

94. Can one connect the approach of the last few questions with the test particle theorem and find a treatment of this technique as a systematic perturbation technique?

95. There is a very "symplectic" looking reciprocity that arises from the test-particle theory: there is a detailed balance between Cerenkov radiation and Landau damping, between synchrotron emission and cyclotron damping, and between Bremsstrahlung and collisional damping. Can one understand this reciprocity within a general theory?

96. Is there a precise mechanical statement and proof of Szilard's resolution of the Maxwell demon paradox? His argument showed that in many situations

one's measuring device uses up more entropy in deciding where a particle is than can be reclaimed on the basis of that knowledge. (For example, Maxwell's demon might observe the scattering of a non-equilibrium photon to detect fast particles approaching his door — the entropy lost in scattering the photon is more than the entropy gained in forcing the fast particle to be on one side of the partition.) This argument is very similar to the maximum entropy version of the second law of thermodynamics given in section 16.4. Can the connection be made precise? What is the relation to projected area concepts discussed in chapter 11? What is the connection with quantum measurement limitations?

97. Is Young's inequality for the Legendre transform (section 16.5) related to the Heisenberg uncertainty principle (section 7.1.3.1) in the eikonal limit?

98. How much of the geometric theory of first order phase transitions given in section 16.8.2 extends to the more difficult situations where renormalization theory is required (eg. critical points)?

99. *Apply the extension of Maxwell's equal area rule for first order phase transitions* given in section 16.8.3 to nontrivial problems.

100. Is there a deep reason for the remarkable parity between the asymptotic structure of eikonal wave theory and statistical mechanics as presented in section 16.9?

PART I:

MECHANICS

"...he couched his discussion in the most sophisticated language known to physics, not all weighed down by lumps of data like Joule's heavy-handed laboratory reports, nor confined to the primitive numerical equivalents of Mayer, but in the graceful, taut, and lissome differential equations of classical dynamics..."

Gillespie referring to Helmholz [Hirsch, 1984]

Chapter 2: Survey of Geometric Perturbation Theory

"I completed my course in engineering and I would like to try to explain the effect of this engineering training on me. Previously, I was interested only in exact equations. It seemed to me that if one worked with approximations there was an intolerable ugliness in ones's work and I very much wanted to preserve mathematical beauty. Well, the engineering training which I received did teach me to tolerate approximations and I was able to see that even theories based upon approximations could have a considerable amount of beauty in them"—P.A.M. Dirac (p. 112 of [Dirac, 1977])

2.1. Historical Background

In this chapter we survey ideas from the rest of the thesis, particularly chapter 4, intuitively and heuristically. In chapter 4 we assume a background in geometric mechanics and give detailed proofs. Here we will give the flavor of the structures and develop the needed background material. We will state results and indicate why *they are true without detailed proof*. We begin with some introductory remarks, discuss a geometric picture for non-singular perturbation theory, introduce the needed

Hamiltonian mechanics including the crucial process of reduction in the presence of symmetry, describe the Hamiltonian structure of non-singular perturbation theory, and close with some discussion of these ideas in connection with the method of averaging. This chapter is an expanded version of the paper [Omohundro, 1984b].

It is of interest to list the seminal ideas that form the background of the present work. In 1808, Lagrange introduced the description of the dynamics of celestial bodies in terms of what we today call Hamilton's equations ([Lagrange, 1808] and [Weinstein, 1981]). His motivation was the reduction of the enormous labor involved in a straightforward perturbation analysis, which required tedious computations to be performed on each component of the dynamical vector field, to manipulations of a single function: the Hamiltonian. The description in terms of Lagrange brackets led to several other benefits. Lagrange showed that the value of the Hamiltonian and the structure of the brackets were both invariant under the dynamics, leading to a useful check of the complex calculations (which at that time were of course done by hand). In addition, he was able to show that the invariance of the Hamiltonian could be used to prove the stability of certain equilibria. As the century progressed, Hamiltonian mechanics was refined and the connections with variational principles and optics were made. By the turn of the century Poincaré (as in [Poincaré, 1892]) had developed very powerful Hamiltonian perturbation methods utilizing generating functions, introduced the notion of asymptotic expansion, and begun the geometric and topological approach to dynamics. In 1918 Emmy Noether made the connection between symmetries and conserved quantities [Noether, 1918]. The development of quantum mechanics rested heavily on the Hamiltonian framework, as

in Dirac, 1958.] by analogy with optics and served to put Hamiltonian structures firmly at the center of the modern formulation of fundamental physics [Landau and Lifshitz, 1960, 1981]. During the 1960's the coordinate-free description of Hamiltonian mechanics in terms of symplectic geometry was developed, as described in Hermann, 1968], Souriau, 1970a], [Abraham and Marsden, 1978] and [Arnold, 1978]. About this time the method of Lie transforms greatly simplified Hamiltonian perturbation theory [Cary, 1981]. The 1970's saw enormous developments in the geometric approach to mechanics and largely as a result of these, an ever wider range of physical systems have been described in Hamiltonian terms. Some examples are: quantum mechanics: [Chernoff and Marsden, 1974], fluid mechanics: [Morrison and Greene, 1980], [Marsden and Weinstein, 1983], and [Marsden, Ratiu, and Weinstein, 1984]. Maxwell's equations: [Pauli, 1933] and [Marsden and Weinstein, 1982], the Maxwell-Vlasov and Poisson-Vlasov equations of plasma physics: [Morrison, 1980], [Marsden and Weinstein, 1982], and [Kaufman, 1982], relativistic plasma waves: [Kaufman and Holm, 1984], gyrokinetic models: [Kaufman and Boghosian, 1984], elasticity theory: [Marsden and Hughes, 1983] and [Holm and Kupershmidt, 1984a], general relativity: [Marsden and Hughes, 1983], magnetohydrodynamics: [Morrison and Greene, 1980] and [Holm and Kupershmidt, 1984a], multi-fluid plasmas: [Spencer and Kaufman, 1982] and [Holm and Kupershmidt, 1984a], chromohydrodynamics: [Gibbons, Holm and Kupershmidt, 1982], superfluids and superconductors: [Holm and Kupershmidt, 1984b], the Korteweg de Vries equation: [Faddeev and Zakharov, 1971], etc. These developments have shed light on the underlying symmetry structure of these theories, have yielded improved

stability results based on Arnold's stability method [Holm, Marsden, Ratiu, and Weinstein, 1984], and have given insight into the reasons for the integrability of certain systems [Guillemin and Sternberg, 1984].

For the most part, however, these structures describe fundamental underlying models in the various fields. In actual applications we almost always make numerous approximations which may or may not respect the underlying Hamiltonian structure. It is folklore within the particle physics community and elsewhere that perturbation methods which respect the underlying symmetries and conservation laws yield much better approximations to the actual system than those which do not. It is of interest, then, to try to do perturbation theory within the Hamiltonian framework and to obtain structures relevant to the approximate system. One may thus hope to understand the relation between the structures of systems which are limiting cases of known systems (eg. does the KdV Poisson bracket arise naturally from that of the Boussinesq equation?) [Olver, 1984]. The history of Hamiltonian mechanics is inextricably tied to perturbation methods. For the most part, though, the Hamiltonian structure was used to simplify the perturbation method and the geometric structure of the perturbation method itself was not explored. We have found in several examples that taking this structure into account leads to simplifications (as in the problem of guiding center motion discussed later in this chapter) and to deeper insight into the approximate system (as for the modulational equations for waves in the eikonal limit studied in [Omohundro, 1984c] and chapter 8).

We have therefore been engaged in a program of investigating the Hamiltonian structure of the various perturbation theories used in practice. In this chapter

we describe the geometry of a Hamiltonian structure for non-singular perturbation theory applied to Hamiltonian systems on symplectic manifolds and the connection with singular perturbation techniques based on the method of averaging. Chapter 5 discusses a singular perturbation technique based on a method introduced by Kruskal

2.2. Geometric Perturbation Theory

In this section we will place perturbation theory into the context of the geometric dynamics that has proven so fruitful in recent years. We will give intuitive discussions of the geometric concepts of dynamics and explicitly put non-singular general first-order perturbation theory into this framework. The next section will do the same for higher order perturbation theory and chapters 4, 5, and 6 will focus on Hamiltonian and singular perturbations.

2.2.1. Manifolds

The modern setting for describing an evolving system is that of a a dynamical system. The state of the system is represented by a point in a manifold M. A *manifold* is a space which locally looks like Euclidean space and in which there is a notion of derivative (for more details see [Abraham, Marsden, and Ratiu, 1983] p. 122). Globally a manifold may be connected together in a non-trivial way, as occurs in the examples of the sphere and the torus.

Many of the standard systems studied in physics have state spaces that are naturally manifolds and have apparent singularities when one tries to model them as Euclidean spaces. A simple example of this is given by the rigid body. The standard description of the configuration of a rigid body utilizes the Euler angles. This description is fraught with singular behavior (eg. more than one set of Euler angles can describe the same configuration at extremes of the angles). The singularities are an artifact of the description and do not correspond to anything in the

physical system. We will see that similar things happen in the infinite-dimensional configuration spaces of fluids and plasmas.

2.2.2. Dynamical Systems

"Among all mathematical disciplines the theory of differential equations is the most important... It furnishes the explanation of all those elementary manifestations of nature which involve time."—Sophus Lie (1895) [Hirsch, 1984]

If you know where you are in a state space, a dynamical law tells you where you're going.

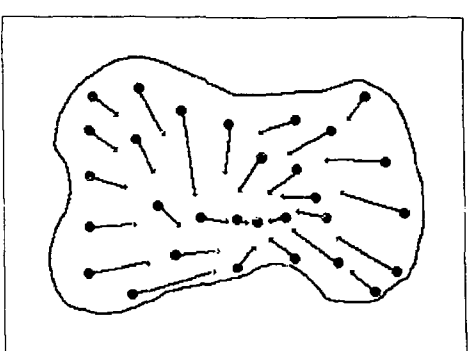

Figure 2.1: A dynamical system

A dynamical law is represented by a vector field on the manifold of states. A *dynamical system* is a manifold with a vector field defined on it, as in figure (2.1) (see Abraham, Marsden and Ratiu, 1983, p. 184). In coordinates, the dynamical vector field is described by a set of first order O.D.E.'s, one for each coordinate. Typical dynamical systems with state spaces of three dimensions or greater have *chaotic* dynamical behavior with extremely complicated trajectories. In many cases one can actually prove that there is no exact description of the solution curves in closed form (see [Guckenheimer and Holmes, 1983]). If the evolution simplifies, then there is some physically relevant special feature, such as a symmetry, which causes the simplification.

2.2.3. Perturbation Theory

In important physical applications, we often find ourselves close to a system which simplifies, and we are interested in the effect of our deviation from it. We express this deviation in terms of the small parameter ϵ.

In many physical situations we are faced with an apparently different problem in which we have but a single dynamical system and we are interested in solutions whose *initial conditions* are close to a known equilibrium point. The parameter ϵ expresses the distance of the initial condition from the equilibrium point. One common situation of this type, which appears repeatedly in plasma physics, has thermal equilibrium as the equilibrium point and studies the time evolution of deviations of initial size ϵ from it.

We may easily convert this type of problem with asymptotic *initial conditions* in an unsolvable dynamical system, to an asymptotic *system* whose limiting case is solvable. This transformation is commonly made by expanding the dynamical vector field about the equilibrium and re-expressing the dynamics in terms of scaled quantities. This leads to a linear system at the lowest order, which is often solvable.

Because thermal equilibrium is stable, this lowest order linear evolution is given by (possibly damped) oscillating normal modes (i.e. there can be no unstable modes in thermal equilibrium). These often take the form of travelling waves. The second order terms in ϵ represent two-wave coupling, the third order terms represent three-wave coupling and so on.

To keep a simple model in mind as we proceed, let us focus on the excitation of a single mode which nonlinearly couples only to itself. This reduces to a system of the form

$$\ddot{u} = f(u). \tag{2.1}$$

If $u = 0$ is to be the equilibrium solution of interest, then we have $f(0) = 0$. Let us assume that f is an odd function in u, since this is a common occurance. To write our equation as a dynamical system, we introduce $v \equiv \dot{u}$ and so obtain the system

$$\dot{u} = v \qquad \dot{v} = f(u). \tag{2.2}$$

We want to study the evolution of small values of u and v, so we choose an asymptotic initial condition:

$$u(t = 0) = \epsilon u^0 \qquad v(t = 0) = \epsilon v^0. \tag{2.3}$$

We now have a hard problem with an *initial condition* that asymptotically approaches an intitial condition whose evolution we can solve for. We want to convert this to a family of *problems* that asymptotically approach one which we can solve. Let us introduce scaled variables:

$$U \equiv \frac{u}{\epsilon} \qquad V \equiv \frac{v}{\epsilon}. \tag{2.4}$$

The initial condition in terms of U and V is constant:

$$U(t=0) = u^0 \qquad V(t=0) = v^0. \tag{2.5}$$

The dynamical system in the new variables has become an asymptotic family:

$$\dot{U} = V \qquad \dot{V} = \frac{f(\epsilon U)}{\epsilon}. \tag{2.6}$$

The limiting system as $\epsilon \to 0$ is

$$\dot{U} = V \qquad \dot{V} = f'(0) \cdot U, \tag{2.7}$$

which is linear. If we express f as an asymptotic series (and remember that f is odd), then we see that our system is asymptotically equivalent to

$$\dot{U} = V \qquad \dot{V} = f'(0) \cdot U + \frac{1}{6}\epsilon^2 f'''(0)\ U^3 + \ldots. \tag{2.8}$$

If we redefine ϵ, u, and v and assume that higher order terms vanish, then this system reduces to the Duffing equations for a nonlinear spring:

$$\dot{u} = v \qquad \dot{v} = -u - \epsilon u^3, \tag{2.9}$$

with initial conditions that are independent of ϵ, say for example

$$u(t=0) = a \qquad v(t=0) = 0. \tag{2.10}$$

We shall use this system as an illustrative example as we proceed. Let us return now to the abstract setting which we have just motivated. The class of problems which require study of a small neighborhood of an equilibrium leads to *linear* zeroth order equations. Examples like gyromotion have nonlinear zeroth order equations which are nonetheless solvable due to symmetries.

In general, we are given a dynamical system of the form

$$\dot{x} = X_0 + \epsilon X_1 + \frac{\epsilon^2}{2!} X_2 + \cdots \tag{2.11}$$

in terms of the vector fields X_i with initial conditions described by

$$x(\epsilon, t = 0) = y(\epsilon). \tag{2.12}$$

We attempt to express the solution as an asymptotic series in ϵ:

$$x(t) = x_0(t) + \epsilon x_1(t) + \frac{\epsilon^2}{2!} x_2(t) + \cdots. \tag{2.13}$$

Choosing coordinates x^a $(1 \leq a \leq N)$ in a local patch of the state space manifold and plugging this assumed asymptotic form into the equation of motion gives

$$\begin{aligned} \dot{x}_0^a + \epsilon \dot{x}_1^a + \frac{\epsilon^2}{2!} \dot{x}_2^a + \cdots = & X_0^a(x_0 + \epsilon x_1 + \frac{\epsilon^2}{2!} x_2 + \cdots) + \\ & + \epsilon X_1^a(x_0 + \epsilon x_1 + \frac{\epsilon^2}{2!} x_2 + \cdots) + \\ & + \frac{\epsilon^2}{2!} X_2^a(x_0 + \epsilon x_1 + \frac{\epsilon^2}{2!} x_2 + \cdots) + \cdots \end{aligned} \tag{2.14}$$

Asymptotic expansions are unique (see for example [de Bruijn, 1981i), so we can

equate coefficients of equal powers of ϵ to get equations for x_0, x_1, x_2.

$$\begin{aligned}
\dot{x}_0^a &= X_0^a(x_0) \\
\dot{x}_1^a &= \sum_{b=1}^{N} \frac{\partial X_0^a}{\partial x^b}(x_0) \cdot x_1^b + X_1^a(x_0) \\
\dot{x}_2^a &= \sum_{b,c=1}^{N} \frac{\partial^2 X_0^a}{\partial x^b \partial x^c}(x_0) x_1^b x_1^c + \sum_{b=1}^{N} \frac{\partial X_0^a}{\partial x^b}(x_0) \quad x_2^b \\
&\quad + 2\sum_{b=1}^{N} \frac{\partial X_1^a}{\partial x^b}(x_0) \cdot x_1^b + X_2^a(x_0) \\
&\vdots
\end{aligned} \tag{2.15}$$

If $y(\epsilon) = y_0 + \epsilon y_1 + \frac{\epsilon^2}{2} y_2 + \cdots$ is an asymptotic expansion for the initial condition $y(\epsilon)$, then the initial conditions for these equations are

$$x_0(t=0) = y_0, \quad x_1(t=0) = y_1, \quad \ldots \tag{2.16}$$

The Duffing equations yield the following equations by this prescription:

$$\begin{aligned}
\dot{u}_0 &= v_0 \qquad \dot{v}_0 = -u_0 \\
\dot{u}_1 &= v_1 \qquad \dot{v}_1 = -u_1 - u_0^3 \\
\dot{u}_2 &= v_2 \qquad \dot{v}_2 = -u_2 - 6u_0^2 u_1 \\
&\vdots
\end{aligned} \tag{2.17}$$

with initial conditions given by

$$\begin{aligned}
u_0(t=0) &= a \qquad v_0(t=0) = 0 \\
u_1(t=0) &= 0 \qquad v_1(t=0) = 0 \\
u_2(t=0) &= 0 \qquad v_2(t=0) = 0 \\
&\vdots
\end{aligned} \tag{2.18}$$

If the Duffing model is itself a truncation of a system with a more general nonlinearity, then it may not be meaningful to carry out the perturbation analysis to too many orders.

These equations immediately raise a number of questions. They are defined in terms of physically irrelevant coordinates; is the perturbation structure independent of these coordinates? If the original equations are Hamiltonian, are these equations? In Jth order perturbation theory, how are we to interpret this evolution of many variables $x_0, x_1, \ldots, x_J$? One goal of this work is to answer these questions.

2.2.4. First Order Perturbation Equations

Let us turn to the geometric interpretation of these equations. It is easiest to understand the *first order* perturbation equations.

For the Duffing example, the first order system is

$$\begin{aligned} \dot{u}_0 &= v_0 \qquad \dot{v}_0 = -u_0 \\ \dot{u}_1 &= v_1 \qquad \dot{v}_1 = -u_1 - u_0^3 \\ u_0(t=0) &= a \qquad v_0(t=0) = 0 \\ u_1(t=0) &= 0 \qquad v_1(t=0) = 0. \end{aligned} \tag{2.19}$$

In general, the first order equations have the form

$$\begin{aligned} \dot{x}_0^a &= X_0^a(x_0) \\ \dot{x}_1^a &= \sum_{b=1}^{N} \frac{\partial X_0^a}{\partial x^b}(x_0)\ x_1^b + X_1^a(x_0) \\ x_0(t=0) &= y_0 \qquad x_1(t=0) = y_1. \end{aligned} \tag{2.20}$$

We would like to determine the geometric nature of the quantities x_0 and x_1. To understand what we mean by this, let us recall the relationship between geometric quantities and coordinates.

2.2 5. Functions, Covectors, and Cotangent Bundles

A function on a manifold is an intrinsically defined thing; it assigns a real number to each point of the manifold. A coordinate system on a region of an N dimensional manifold is a collection of N real valued-functions $x^1,\ldots,x^N$ defined on that region, whose differentials are linearly independent at each point. In these coordinates, the gradient of a function is a collection of N numbers: the derivatives with respect to each of the x^a. Geometrically, however, it is wrong to think of these as just real numbers, because they change if we change our coordinate system. For example, if we choose coordinates whose values at each point of the region are twice those of $x^1,\ldots,x^N$ then the components of the gradient of a function are halved We introduce a geometric object whose relationship to the manifold at a given point is like that of the differential of a function and we call it a *covector* or *one-form* (see [Abraham, Marsden, and Ratiu, 1983] p.286). In this context, the gradient is usually refered to as the differential of the function. The collection of all covectors at a point is defined to be the *cotangent space* at that point and the collection of all cotangent spaces taken together form the *cotangent bundle* (see [Abraham, Marsden, and Ratiu, 1983] p. 285).

2.2.6. Vectors and Tangent Bundles

Similarly, the values of the components of a *vector* at a point are doubled when we double the values of the coordinates. All vectors at a point taken together form the *tangent space* at that point and all tangent spaces taken together form the *tangent bundle* TM of M (see [Abraham, Marsden, and Ratiu, 1983] p. 150).

Pictorially, we think of a vector as a little arrow whose end is at the point of interest as in figure (2.2a). A covector may be thought of as a pair of parallel planes representing local level sets of a function whose gradient is that covector ([Misner, Thorne, and Wheeler, 1973] and [Burke, 1980]) as in figure (2.2b). The distance between planes gets *smaller* as the gradient gets larger so that the amount of a vector starting at the first plane that is cut off by the second plane is independent of the scale (and so defines an invariant pairing between vectors and covectors). Covectors have been referred to as lasagna vectors because of this picture (Jim Napolitano, private communication).

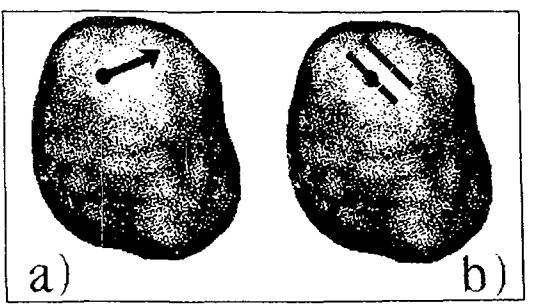

Figure 2.2: a) A picture of a vector, b) A picture of a covector.

Vectors and covectors are different objects when we consider more than one coordinate system, even though they both have N components in any given system. If we have two curves in a manifold through a given point, in any coordinate system

we may determine if the curves go through the point at the same rate and in the same direction (i.e. if they are tangent to first order). It turns out that this determination is independent of which coordinates are used. In differential geometry, one usually defines a tangent vector at a point to be an equivalence class of curves which are tangent to first order. The invariant pairing between vectors and covectors is then used to define covectors as elements of the dual space to the tangent space at a point. The dual space V^* of a vector space V is the vector space of linear functions on V. Our interest here will be to find out whether the quantities $x_0^a, \ldots, x_j^a$ for $1 \leq a \leq N$ have any geometric structure that is independent of a given coordinate system

2.2.7. The State Space for First Order Perturbation Theory

Intuitively, the first order quantity x_1 represents a small deviation from the unperturbed quantity x_0. Because x_0 can vary over the whole manifold M, we expect it to represent a point in the manifold. As ϵ gets smaller, $x_0 + \epsilon x_1$ approaches the point x_0. The variable x_1 measures the first order rate of approach to x_0. Two different paths in the manifold approaching the point x_0 as ϵ approaches zero have the same x_1 if and only if they are tangent at x_0. This, however, is the defining criterion for a vector at the point x_0. We thus expect x_1 to lie in the tangent space to M over the point x_0. The (x_0, x_1) dynamics then takes place in the tangent bundle TM. We will describe this dynamics on TM intrinsically in terms of vector fields derived from $X(\epsilon)$ on M.

2.2.8. Flows and Derivatives

The solution of a system of O.D.E.'s tells us the state at each time t of a system which began with each initial condition. Geometrically, this is a mapping of M to itself for each t. If the solution doesn't run off the manifold, then the uniqueness and smoothness of solutions with given initial conditions tells us that this map is a diffeomorphism (i.e. a smooth, 1-1, onto map with smooth inverse, [Abraham, Marsden, and Ratiu, 1983] p. 102). This one-parameter family of diffeomorphisms labelled by t is called the flow of the dynamical vector field (see [Abraham, Marsden, and Ratiu, 1983] p. 185). As ϵ varies, the corresponding flows of $X(\epsilon)$ will vary. Perturbation theory describes that variation. Any time we have a mapping f from one manifold to another, we may define the derivative map Tf, called the tangent of f (see [Abraham, Marsden, and Ratiu, 1983] p. 153). This is a map that takes the tangent bundle of the first manifold to the tangent bundle of the second. It describes how infinitesimal perturbations at a point are sent to infinitesimal perturbations at the image point. In coordinates, it acts on the tangent space at a point via the Jacobian matrix of f at that point.

2.2.9. Dynamics for First Order Perturbation Theory

Let us denote the flow of the unperturbed vector field X_0 by $x_0(t)$. $x_0(t, y_0)$ is the point to which y_0 has flowed in time t under X_0. A small perturbation in M from a given orbit will evolve under X_0 according to the derivative of this flow: $Tx_0(t)$. This derivative is itself a flow on the manifold TM. The vector field of

where it is the flow may be written

$$\tilde{X}_0 \equiv \left.\frac{d}{dt}\right|_{t=0} T x_0(t). \tag{2.21}$$

$\tilde{X}_0$ is a vector field on TM, defined without recourse to coordinates, that represents the effect of the unperturbed flow on perturbed orbits. In coordinates, $\tilde{X}_0$ has components given by

$$\begin{aligned} \dot{x}_0^a &= X_0^a(x_0) \\ \dot{x}_1^a &= \sum_{b=1}^{N} \frac{\partial X_0^a}{\partial x^b}(x_0) \cdot x_1^b. \end{aligned} \tag{2.22}$$

These dynamical equations represent exactly that part of the perturbation equations (2.20) which depends on X_0.

For the Duffing example, these equations give

$$\begin{aligned} \dot{u}_0 &= v_0 \qquad \dot{v}_0 = -u_0 \\ \dot{u}_1 &= v_1 \qquad \dot{v}_1 = -u_1. \end{aligned} \tag{2.23}$$

The last two of these describe the evolution under the zero order equations of a little perturbation along (u_1, v_1). We now see that (u_1, v_1) gives the coordinates of a tangent vector based at (u_0, v_0). The (u_1, v_1) equations are of the same form as the (u_0, v_0) equations. This is because the zero order system for the Duffing oscillator is linear and the derivative of a linear map is the identity when we identify the linear space with its tangent space.

The part of the dynamics which depends on X_1 may also be defined intrinsically. For any $x \in M$ and $v \in T_x M$, we define

$$\tilde{X}_1(x, v) \equiv \left.\frac{d}{dt}\right|_{t=0} (v - t\, X_1(x)). \tag{2.24}$$

The $\tilde{X}_1$ dynamics for the Duffing system is

$$\begin{aligned} \dot{u}_0 &= 0 \qquad \dot{v}_0 = 0 \\ \dot{u}_1 &= 0 \qquad \dot{v}_1 = -u_0^3. \end{aligned} \tag{2.25}$$

This represents the additional effect of the perturbed equation without including the effect of the unperturbed system.

The entire first order perturbation dynamics on TM is given by

$$\tilde{X}_0 + \tilde{X}_1. \tag{2.26}$$

We have therefore succeeded in finding a geometric, coordinate-free interpretation for first order perturbation theory.

2.3. The Geometry of Jth Order Perturbation Theory

We now would like to extend this picture to higher orders. The geometric object that arises is called a *jet*. To understand the setting, we discuss a number of relevant spaces.

2.3.1. The Path Space

How are we to think of the exact equation for the evolution of an ϵ-dependent point $x(\epsilon)$ under ϵ-dependent evolution equation $\dot{x}(\epsilon) = X(\epsilon, x)$ with ϵ-dependent initial conditions $y(\epsilon)$? It is useful to think of the ϵ-dependent point $x(\epsilon)$ as a curve in the space $I \times M$, where I is the interval (say $[0, 1]$) in which ϵ takes its values (as in figure (2.3)). We shall call such curves *paths* (as this is the standard mathematical terminology). Moving along a path corresponds to varying the asymptotic parameter ϵ.

If we think of $x(\epsilon)$ as a map from I to M, then the curve is the graph of this map. The dynamical vector field $X(\epsilon)$ naturally lives on $I \times M$ and its I component is zero everywhere. The flow of $X(\epsilon)$ on $I \times M$ takes paths to paths by letting each point of a path move with the flow as in figure (2.4). Our initial conditions are represented by paths (if they are independent of ϵ then they are straight lines). The true dynamics takes paths to paths. Even if the intial conditions are ϵ-independent, the ϵ dependent dynamics bends the path over as in figure (2.5).

Thus we really should think of our dynamics as living on the infinite dimensional *path space*

$$P_1 M \equiv \left\{ \text{space of all paths } p : I \to I \times M \text{ of the form } p : \epsilon \mapsto (\epsilon, x(\epsilon)) \right\}, \quad (2.27)$$

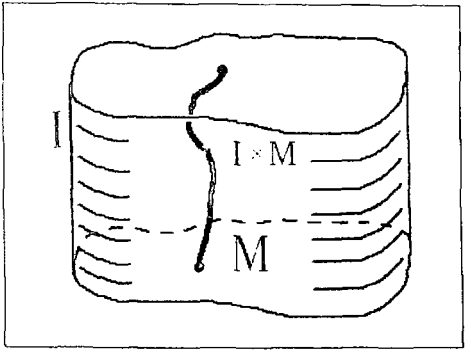

Figure 2.3: Curve in $I \times M$.

where, as before, $I = [0, 1]$. For the Duffing example, this is the space of curves in (u, v, ϵ) space that project diffeomorphically onto ϵ. Each curve represents $u(\epsilon), v(\epsilon)$ for ϵ in I.

This space projects naturally onto

$$P_0M \equiv \left\{ \text{equivalence classes in } P_1M \text{ where } p_1 \sim p_2 \text{ iff } p_1(0) = p_2(0) \right\}. \quad (2.28)$$

The projection sends a curve to its $\epsilon = 0$ endpoint which represents the point about which the perturbation is taken. P_0M is naturally isomorphic to M and represents the domain of the unperturbed dynamics. The equivalence classes forget all *perturbation information* and *only remember* behavior at $\epsilon = 0$ as shown in figure (2.6).

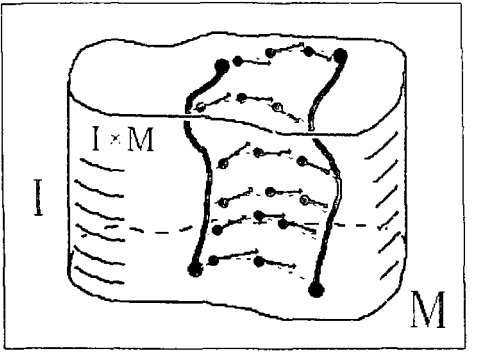

Figure 2.4: A path moving under the influence of a vector field $X(\epsilon)$.

2.3.1.1. Spaces of Shorter Paths

We are interested in spaces through which this projection of actual to unperturbed dynamics factors (i.e. spaces which are the image of a projection from the first space and the domain of a projection to the second space, such that the composition of these two projections gives the original projection from the first to the second space). Perturbation theory tries to study behavior infinitesimally close to $\epsilon = 0$ without actually getting there. For each $0 \leq \alpha \leq 1$ we may define

$$P_\alpha M \equiv \Big\{ \text{equivalence classes in } P_1 M \text{ where } p_1 \sim p_2 \text{ iff } p_1(\epsilon) = p_2(\epsilon)\ \forall\ 0 \leq \epsilon \leq \alpha \Big\}. \tag{2.29}$$

These allow us to consider more and more restricted domains of ϵ, but there is always

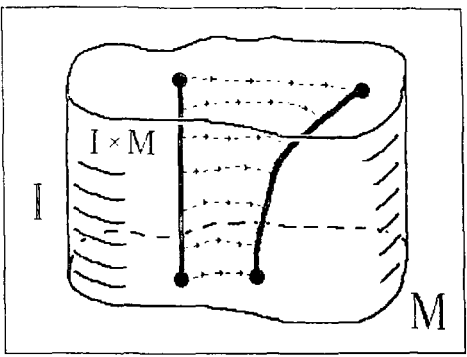

Figure 2.5: An ϵ-independent initial condition becoming ϵ-dependent.

a continuum of ϵ's to traverse before reaching $\epsilon = 0$. For each $1 \geq \alpha_1 \geq \alpha_2 \geq 0$ we have the natural maps

$$P_1 M \to P_{\alpha_1} M \to P_{\alpha_2} M \to P_0 M. \tag{2.30}$$

2.3.2. The Space of Germs of Paths

We are interested in structure between "even the smallest $P_\alpha M$ with $\alpha \neq 0$" and $P_0 M$. We may introduce germs of paths:

$$GM \equiv \Big\{ \text{equivalence classes in } P_1 M \text{ where } p_1 \sim p_2 \text{ iff} \\ \exists \alpha_{12} > 0 \text{ such that } p_1(\epsilon) = p_2(\epsilon) \; \forall \; 0 \leq \epsilon \leq \alpha_{12} \Big\}. \tag{2.31}$$

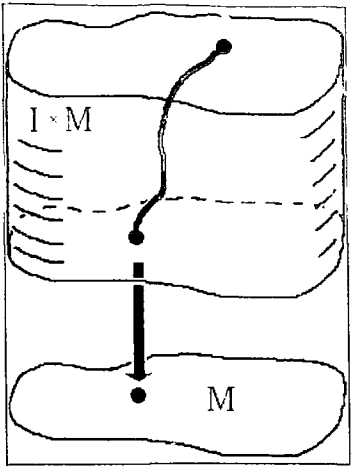

Figure 2.6: The projection of a path to its $\epsilon = 0$ endpoint.

For any $\alpha > 0$ we have $P_\alpha M \to GM \to P_0 M$. The germs capture behavior closer to $\epsilon = 0$ than any given ϵ, but still contain much more information than perturbation theory gives us (germs depend on features of functions in a little neighborhood that may not be captured in a Taylor series.)

2.3.3. The Space of Jets of Paths

Finally we may introduce spaces of jets of paths at $\epsilon = 0$ with integer $1 \leq J \leq \infty$:

$$JM = \Big\{ \text{equivalence classes in } P_1M \text{ where } p_1 \sim p_2 \text{ iff} \\ \forall\, C^\infty \text{ functions } f \text{ on } I \times M \text{ we have} \\ \left.\frac{\partial^i}{\partial \epsilon^i}\right|_{\epsilon=0} f(p_1(\epsilon)) = \left.\frac{\partial^i}{\partial \epsilon^i}\right|_{\epsilon=0} f(p_2(\epsilon)) \qquad \text{for } 0 \leq i \leq J \Big\}. \tag{2.32}$$

Notice that ∞M is the space of infinite formal power series by this definition. The space of J-jets gives the first J terms in a Taylor expansion of the curve around $\epsilon = 0$ in any coordinate system. Clearly,

$$GM \to \infty M \to IM \to JM \to P_0M \qquad \text{for } I > J. \tag{2.33}$$

Thus the jets focus on information closer to $\epsilon = 0$ than even the germs.

In the Duffing example, the J-jet of $(u(\epsilon), v(\epsilon))$ consists of the values of the first J derivatives with respect to ϵ of u and v. We called these $u_0, v_0, u_1, v_1, \ldots u_J, v_J$.

2.3.3.1. Coordinates on the Jet Space

If x^a for $1 \leq a \leq N$ are coordinates on $M \approx P_0M \approx 0M$, then we may introduce coordinates $\{x_0^a, x_1^a, \ldots, x_J^a\}$ for $0 \leq J \leq \infty$ on JM to represent the equivalence class of the curve:

$$x_0^a + \epsilon x_1^a + \frac{\epsilon^2}{2!} x_2^a + \cdots + \frac{\epsilon^J}{J!} x_J^a \tag{2.34}$$

in $I \times M$ (near $\epsilon = 0$ this won't leave the chart on which the x^a are defined).

The claim here is that JM represents geometrically the perturbation quantities $x, \dot{x}, \ddot{x}, \ldots$. It may seem strange to go through the infinite dimensional space P_1M to get to it, but we shall see (especially when looking at the Hamiltonian structure) that it organizes and simplifies the structures of interest. It is a completely intrinsic and natural (or in modern parlance: functorial) operation to go from the original dynamical manifold M to the path space P_1M to the jet space JM. We shall now show that the dynamics on M also induces natural dynamics on P_1M and then projects from there down to JM where it is the perturbation dynamics we are interested in. Later we will see that a Hamiltonian structure on M leads to Hamiltonian structures on P_1M and JM.

The dynamics $\dot{x} = X(\epsilon, x)$ takes elements of P_1M to other elements of P_1M and in fact takes equivalence classes to equivalence classes for each of P_0M, GM, ∞M, JM, and M. This is what allows us to obtain an induced dynamics on each of these spaces. To determine this dynamics explicitly, we must understand what a tangent vector on each space is.

2.3.4. Tangent Vectors to Path Space

Intuitively, a vector represents a little perturbation to a point. We define it precisely as an equivalence class of tangent curves, where the curve represents the direction of perturbation and the equivalence class ensures that only the first order motion is reflected in the tangent vector. A point in the path space P_1M represents a path in $I \times M$. A small perturbation of this point represents a nearby path. Each point of the path is perturbed a little bit and we are interested in the first order

perturbation. We therefore expect a tangent vector to a point in path space to be a vector field along the corresponding path as in figure (2.7). By a vector field *along* a path, we mean a choice of tangent vector to M at each point of the path. In general, a vector field along a map f from a manifold M to a manifold N, is a smooth choice of image-space tangent vector in $T_{f(x)}N$ for each point x in the source space M.

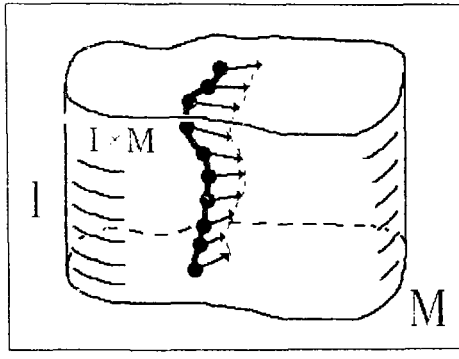

Figure 2.7: A small perturbation of a path in $I \times M$ is given by a vector field along that path.

A curve $p(\gamma)$ in P_1M parameterized by γ defines a curve $p(\epsilon, \gamma)$ for each ϵ through $p(\epsilon, \gamma = 0)$ in $I \times M$. The equivalence class of curves in P_1M defining a vector thus reduces to an equivalence class of curves in M for each ϵ. We may identify a tangent vector to p in P_1M with a field of vectors over p in $I \times M$ such

that each vector has no $\frac{\partial}{\partial \epsilon}$ component. For $p \in P_1 M$ a vector $\tilde{V} \in T_p(P_1 M)$ is a map

$$\tilde{V} : I \to I \times TM \tag{2.35}$$

taking $\epsilon \mapsto (\epsilon, V(\epsilon))$ where $V(\epsilon) \in T_{p(\epsilon)}M$ (i. e. a path in TM)

For the Duffing example, a tangent vector based at the path $(u(\epsilon), v(\epsilon))$ has the form $(\delta u(\epsilon), \delta v(\epsilon))$ and represents the vector field

$$\delta u(\epsilon) \left.\frac{\partial}{\partial u}\right|_{(u(\epsilon),v(\epsilon))} + \delta v(\epsilon) \left.\frac{\partial}{\partial v}\right|_{(u(\epsilon),v(\epsilon))} \tag{2.36}$$

along the path $(u(\epsilon), v(\epsilon))$.

2.3.5. Tangent Vectors to the Quotient Spaces

The tangent spaces to the quotient spaces are defined by taking the derivatives of the natural projections. Because $P_0 M \approx M$, we see that $TI_0 M \approx TM$. Because the first jet space is isomorphic to the tangent bundle: $1M \approx TM$, we see that $T1M \approx TTM$. Thus the first order perturbation space $1M$ is naturally TM and the dynamical evolution is given by a vector field on TM as we saw in section 2.2.9.

2.3.5.1. Coordinates on the Tangent Bundle to the Jet Space

As with all tangent bundles, TJM has a natural coordinate chart, derived from the coordinates $\{x_0^a, \ldots, x_J^a\}$, $1 \le a \le N$ on JM defined earlier. We obtain coordinates $\{x_0^a, \ldots, x_J^a, v_0^a, \ldots, v_J^a\}$ by writing the corresponding vector as

$$\sum_{a=1}^{N} v_0^a \frac{\partial}{\partial x_0^a} + \cdots + v_J^a \frac{\partial}{\partial x_J^a}. \tag{2.37}$$

We would like to know to which set of components $\{v_0^a, \ldots, v_J^a\}$, the equivalence class of a vector $V(\epsilon)$ on P_1M corresponds. For the Duffing example, a tangent vector to the space of J-jets looks like

$$\delta u_0 \frac{\partial}{\partial u_0} + \delta v_0 \frac{\partial}{\partial v_0} + \ldots + \delta u_J \frac{\partial}{\partial u_J} + \delta v_J \frac{\partial}{\partial v_J}. \tag{2.38}$$

2.3.5.2. Coordinate Relation Between Path and Jet Vectors

To the path $x^a(\epsilon)$ representing a point in P_1M corresponds the point coordinatized by

$$x_k^a = \left.\frac{\partial^k}{\partial \epsilon^k}\right|_{\epsilon=0} x^a(\epsilon), \qquad 1 \le a \le N, \quad 0 \le k \le J \tag{2.39}$$

in JM. To the curve of paths $x^a(\epsilon, \gamma)$ in P_1M corresponds the curve

$$x_k^a(\gamma) = \left.\frac{\partial^k}{\partial \epsilon^k}\right|_{\epsilon=0} x^a(\epsilon, \gamma), \qquad 1 \le a \le N, \quad 0 \le k \le J \tag{2.40}$$

in JM. The vector tangent to this curve in TP_1M has coordinates

$$V^a(\epsilon) = \left.\frac{\partial}{\partial \gamma}\right|_{\gamma=0} x^a(\epsilon, \gamma). \tag{2.41}$$

In TJM this corresponds to

$$\begin{aligned} v_k^a &= \left.\frac{\partial}{\partial \gamma}\right|_{\gamma=0} \left(\left.\frac{\partial^k}{\partial \epsilon^k}\right|_{\epsilon=0} x^a(\epsilon, \gamma) \right) \\ &= \left.\frac{\partial^k}{\partial \epsilon^k}\right|_{\epsilon=0} \left(\left.\frac{\partial}{\partial \gamma}\right|_{\gamma=0} x^a(\epsilon, \gamma) \right) \\ &= \left.\frac{\partial^k}{\partial \epsilon^k}\right|_{\epsilon=0} V^a(\epsilon). \end{aligned} \tag{2.42}$$

For the Duffing equation, the tangent vector represented by $\{\delta v(\epsilon), \delta v(\epsilon)\}$ projects to the jet space vector

$$\delta u_0 \frac{\partial}{\partial u_0} + \delta v_0 \frac{\partial}{\partial v_0} + \cdots + \delta u_J \frac{\partial}{\partial u_J} + \delta v_J \frac{\partial}{\partial v_J} \tag{2.43}$$

where

$$\begin{aligned} \delta u_0 &= \delta u(0) \qquad \delta v_0 = \delta v(0) \\ \delta u_1 &= \frac{\partial}{\partial \epsilon} \delta u(\epsilon)|_{\epsilon=0} \qquad \delta v_1 = \frac{\partial}{\partial \epsilon} \delta v(\epsilon)|_{\epsilon=0} \\ &\vdots \\ \delta u_J &= \frac{\partial^J}{\partial \epsilon^J} \delta u(\epsilon)|_{\epsilon=0} \qquad \delta v_J = \frac{\partial^J}{\partial \epsilon^J} \delta v(\epsilon)|_{\epsilon=0} \end{aligned} \tag{2.44}$$

2.3.6. Dynamics on Path Space

Let us now consider the effect of the dynamics $\dot{x} = X(\epsilon, x)$ on paths. This lifts to a vector field on $P_1 M$ given by

$$\tilde{X} \quad \text{where} \quad \tilde{X}(p) : \epsilon \mapsto X(\epsilon, p(\epsilon)). \tag{2.45}$$

This is the path space dynamical vector field. Note that this vector field is of a very special type and not every vector field on $P_1 M$ can arise in this way. For the Duffing example, the path space vector field $\{\delta u(\epsilon), \delta v(\epsilon)\}|_{u,v}$ at the point (u, v) is equal to

$$\delta u(\epsilon) = v(\epsilon) \qquad \delta v(\epsilon) = -u(\epsilon) - \epsilon u^3(\epsilon). \tag{2.46}$$

2.3.7. Dynamics on Jet Space

In coordinates, the corresponding vector field on JP is

$$V_k^a(x_0, \ldots, x_J) = \left.\frac{\partial^k}{\partial \epsilon^k}\right|_{\epsilon=0} X^a(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J)$$

$$\frac{\partial^k}{\partial \epsilon^k}\Big|_{\epsilon=0} X^a(\epsilon, x_0) + k \left.\frac{\partial^{k-1}}{\partial \epsilon^{k-1}}\right|_{\epsilon=0} \sum_{b=1}^{N} \frac{\partial}{\partial x_0^b} X^a(\epsilon, x_0) x_1^b + \cdots, \tag{2.47}$$

which is exactly the perturbation dynamics up to order J obtained in equations (2.15)!

For example, we calculate that

$$\begin{aligned}
V_0^a(x_1, \ldots, x_J) &= X^a(0, x_0) \\
V_1^a(x_0, \ldots, x_J) &= \left.\frac{\partial}{\partial \epsilon}\right|_{\epsilon=0} X^a(\epsilon, x_0) + \sum_{b=1}^{N} \frac{\partial}{\partial x_0^b} X^a(0, x_0) x_1^b \\
V_2^a(x_0, \ldots, x_J) &= \left.\frac{\partial^2}{\partial \epsilon^2}\right|_{\epsilon=0} X^a(\epsilon, x_0) \\
&\quad + 2 \left.\frac{\partial}{\partial \epsilon}\right|_{\epsilon=0} \sum_{b=1}^{N} \frac{\partial}{\partial x_0^b} X^a(0, x_0) x_1^b \\
&\quad + \sum_{b,c=1}^{N} \frac{\partial^2}{\partial x_0^b} X^a(0, x_0) x_1^b x_1^c + \sum_{b=1}^{N} \frac{\partial}{\partial x_0^b} X^a(0, x_0) x_2^b \\
&\vdots
\end{aligned} \tag{2.48}$$

In the Duffing example, we see that the jet vector field

$$\delta u_0 \frac{\partial}{\partial u_0} + \delta v_0 \frac{\partial}{\partial v_0} + \cdots + \delta u_J \frac{\partial}{\partial u_J} + \delta v_J \frac{\partial}{\partial v_J} \tag{2.49}$$

has the components:

$$\delta u_0 = \delta u(0) = v(0)$$

$$\delta v_0 = \delta v(0) = -u(0)$$

$$\begin{aligned}\delta u_1 &= \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \delta u(\epsilon) + \frac{\partial}{\partial u}\bigg|_{\epsilon=0} (\delta u(0))u_1 + \frac{\partial}{\partial v}\bigg|_{\epsilon=0} (\delta u(0))v_1 \\ &= 0 + 0 + 1 \cdot v_1 = v_1(0)\end{aligned} \tag{2.50}$$

$$\begin{aligned}\delta v_1 &= \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \delta v(\epsilon) + \frac{\partial}{\partial u}\bigg|_{\epsilon=0} (\delta v(0))u_1 + \frac{\partial}{\partial v}\bigg|_{\epsilon=0} (\delta v(0))v_1 \\ &= -u^3(0) + (-1 - 3\epsilon u^2)|_{\epsilon=0} u_1 = -u^3(0) - u_1\end{aligned}$$

$$\vdots$$

We have thus found the natural geometric setting for Jth order perturbation theory in a certain jet bundle. The picture of the dynamics of paths in $I \times M$ is an extremely fruitful one. One can prove that the solution of the perturbation equations (2.15) really is the asymptotic expansion of the true solution just by noting that they are the equations of evolution of the jets of the paths evolving under the true dynamics. The coordinates in which the dynamics are expressed are irrelevant as regards the perturbation dynamics and therefore we can do perturbation theory on manifolds and in infinite dimensions as is required for many physical systems. Next we will review modern Hamiltonian mechanics and indicate why the perturbation dynamics is Hamiltonian in a natural way if the unperturbed dynamics is.

2.4. Geometric Hamiltonian Mechanics

"The next morning, I hurried along to one of the libraries as soon as it was open and then I looked up Poisson brackets in Whittaker's 'Analytic Dynamics' and I found that they were just what I needed"—P. A. M. Dirac in [Dirac, 1977] p. 122.

The evolution of mechanical systems is traditionally described in terms of generalized coordinates q_i and their conjugate momenta p_i. One introduces the Hamiltonian function

$$H(q_1, \ldots, q_n, p_1, \ldots, p_n) \tag{2.51}$$

and the Poisson bracket

$$\{f, g\} = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right) \tag{2.52}$$

of two functions of q_i and p_i. Any observable f evolves according to the evolution equation

$$\dot{f} = \{f, H\}. \tag{2.53}$$

For the Duffing example, we may use the Poisson bracket

$$\{f, g\} = \frac{\partial f}{\partial u} \frac{\partial g}{\partial v} - \frac{\partial f}{\partial v} \frac{\partial g}{\partial u}, \tag{2.54}$$

the Hamiltonian

$$H = \frac{v^2}{2} + \frac{u^2}{2} + \epsilon \frac{u^4}{4}. \tag{2.55}$$

Together these give rise to the correct equations:

$$\dot{u} = \{u, H\} = \frac{\partial H}{\partial v} = v \tag{2.56}$$

and:

$$\dot{v} = \{v, H\} = -\frac{\partial H}{\partial u} = -u - \epsilon u^3. \tag{2.57}$$

2.4.1. Poisson Manifolds

For a detailed description of the modern approach to mechanics see references [Abraham and Marsden, 1978], [Arnold, 1978], and [Marsden, 1981]. The modern perspective regards the particular coordinates p_i and q_i as physically irrelevant. Just as general relativity isolates the physically relevant essence of local coordinates in a metric tensor, modern classical mechanics views the Poisson bracket structure (not necessarily expressed in any coordinate system) as the physical entity. Just as physics in spacetime is invariant under transformations that preserve the metric, physics in phase space is invariant under the *canonical transformations* which preserve the Poisson bracket. In the modern viewpoint one proceeds axiomatically and does not require canonical coordinates. Dynamics occurs on a *Poisson manifold*. This is a manifold of states with a Poisson bracket defined on it. From this viewpoint a Poisson bracket is a bilinear map from pairs of functions to functions which makes the space of functions into a Lie algebra and acts on products as a derivative does:

$$\begin{aligned}
&\text{I. Bilinearity:} && \{af_1 + bf_2, g\} = a\{f_1, g\} + b\{f_2, g\} \\
&\text{II. Anti-symmetry:} && \{f, g\} = -\{g, f\} \\
&\text{III. Jacobi's identity:} && \{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0 \\
&\text{IV. Derivation property:} && \{f, gh\} = \{f, g\}h + \{f, h\}g
\end{aligned} \tag{2.58}$$

2.4.2. Hamiltonians and Hamiltonian Vector Fields

The *Hamiltonian* is a function on the Poisson manifold. The evolution of local coordinates z^i is obtained from a Hamiltonian H and the Poisson bracket $\{,\}$ via

$$\dot{z}^i = X_H \cdot z^i = \{z^i, H\}. \tag{2.59}$$

X_H is the *Hamiltonian vector field* associated with H and defines a dynamical system. The fourth property of a Poisson bracket implies the useful expression

$$\{f, g\} = \sum_{i,j} \frac{\partial f}{\partial z^i} \{z^i, z^j\} \frac{\partial g}{\partial z^j}. \tag{2.60}$$

Thus the Poisson bracket is equivalent to an antisymmetric contravariant two-tensor

$$J^{ij} \equiv \{z^i, z^j\}. \tag{2.61}$$

2.4.3. Symplectic Manifolds

If this tensor is nondegenerate, its inverse $\omega \equiv J^{-1}$ is a closed, nondegenerate two-form called a *symplectic structure*. In this case our Poisson manifold is known as a *symplectic manifold* (see [Abraham, Marsden, and Ratiu, 1983] p. 463). The terminology is due to Herman Weyl. The symplectic group is related to "line complexes" in projective geometry and so ω was originally refered to as a complex structure. There is another object that naturally has this name in the study of the complex analysis of many variables, however. To eliminate this confusion, Weyl took the Latin roots *com* and *plex* and converted them to their Greek equivalents *sym* and *plectic*.

2.4.4. Symplectic Leaves and Bones and Casimir Functions

Because we do not require nondegeneracy, a Poisson manifold is a more general notion than a symplectic manifold. If J is degenerate, then there are directions in phase space in which no Hamiltonian vector field can point. The available directions lie tangent to submanifolds which fill out the Poisson manifold and on which J is nondegenerate. The highest dimensional of these form a foliation of their union and so are known as *symplectic leaves*. The only prior usage of the term symplectic in English is to describe a small bone in the head of a fish. Because Poisson is French for fish, the lower dimensional symplectic submanifolds are sometimes known as *symplectic bones* (these notions were introduced in [Weinstein, 1983a]). Together, the symplectic leaves and the symplectic bones fill out the Poisson manifold, and any Hamiltonian dynamics is restricted to lie on a single bone or leaf. Any function which is constant on each bone and leaf Poisson commutes with every other function. Any function which Poisson commutes with every function is automatically a constant of the motion, regardless of the Hamiltonian and is called a *Casimir function*.

2.4.5. The Natural Symplectic Structure on Cotangent Spaces

A natural symplectic manifold arises from each Lagrangian mechanical system on a configuration space C. The Lagrangian L lives on the tangent bundle TC (velocities being tangent to the curves of motion in configuration space are naturally tangent vectors). Hamiltonian mechanics takes place on the cotangent bundle T^*C (momenta, being derivatives of L with respect to velocity, are naturally dual to

velocities and thus are covectors). At each point x of C, the set of velocity vectors v in T_xC are identified with corresponding momenta p in T_x^*C by a Legendre map defined by L (the classical concept of a Legendre map is described in modern terms in section 16.5.11):

$$p = \frac{\partial L}{\partial v}. \tag{2.62}$$

The Hamiltonian is a function on the cotangent bundle which on each cotangent fiber is equal to the Legendre transform (described in section 16.5.5) of the the Lagrangian restricted to the corresponding fiber of the tangent bundle.

T^*C has the natural symplectic structure

$$\omega = -d\theta \tag{2.63}$$

where θ is an intrinsically defined one-form on T^*C (see [Abraham, Marsden, and Ratiu, 1983] p. 465). θ must pair with a tangent vector v in TT^*C based at (x, α) in T^*C to give a real number. To define this we use the natural projection

$$\pi : T^*C \to C \tag{2.64}$$

which takes a covector to its basepoint in C. The differential of π sends TT^*C to TC and may be applied to v to get a vector tangent to C. α is a one-form on C and may be applied to this vector. Let us define the pairing of θ with v to be the pairing of α with the image of v under $T\pi$:

$$\theta(v) = \alpha(T\pi v). \tag{2.65}$$

In coordinates q^a on C, this takes the form

$$\theta = p_a dq^a \tag{2.66}$$

and leads to the symplectic form

$$\omega = dq^a \wedge dp_a. \tag{2.67}$$

This construction generalizes the usual structure in terms of canonical p's and q's to configuration spaces which are manifolds. Symmetry is responsible for the simplified systems about which we perturb and plays an intimate role in our geometric theory. We therefore next introduce some key modern ideas and basic examples relating to Hamiltonian symmetry.

2.5. Hamiltonian Systems with Symmetry

Perhaps the central advantageous feature of systems with a Hamiltonian structure is a generalization of *Noether's theorem* relating symmetries to conserved quantities. Noether considered symmetries of the Lagrangian under transformations of configuration space. One may introduce generalized coordinates $(q_1, \ldots, q_n)$ where $(q_2, \ldots, q_n)$ are constant under the symmetry transformation and q_1 varies with the transformation. For example we might take the configuration space to be ordinary Euclidean 3-space where the action of the symmetry is translation in the x direction, and utilize the coordinates

$$q_1 = x, \quad q_2 = y, \quad q_3 = z. \tag{2.68}$$

That L is invariant means that it does not depend on q_1, i.e. q_1 is an ignorable coordinate. The Euler-Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = 0 \tag{2.69}$$

show that in this case the momentum

$$p_1 = \frac{\partial L}{\partial \dot{q}_1} \tag{2.70}$$

conjugate to q_1 is actually a constant of the motion.

2.5.1. Generalized Noether's Theorem

By going to a Hamiltonian description in terms of Poisson brackets we may extend Noether's theorem in a fundamental way. We may consider a one-parameter symmetry transformation of the whole phase space as opposed to just configuration space. If this transformation preserves the Hamiltonian and the Poisson bracket (i.e. is a canonical transformation) then it is associated with a conserved quantity. We will see that this extension of Noether's theorem is essential in the case of gyromotion and in other examples.

One-parameter families of canonical transformations of this type may be represented as the "time" s evolution generated by some function J, treated momentarily as a Hamiltonian. Parametrizing our transformation by s and labelling points in phase space by $\underline{z}$, the solution $\underline{z}(s)$ of

$$\frac{d\underline{z}}{ds} = \{\underline{z}, J\} \qquad \underline{z}(s=0) = \underline{z}_0 \tag{2.71}$$

is the family of canonical transformations generated by J.

If the transformations generated by J are symmetries of H then

$$\begin{aligned} \frac{dH}{ds} = 0 &= \sum_i \frac{\partial H}{\partial z^i}\frac{dz^i}{ds} \\ &= \sum_i \frac{\partial H}{\partial z^i}\{z^i, J\} \\ &= \{H, J\} \\ &= -\{J, H\} \\ &= -\dot{J}. \end{aligned} \tag{2.72}$$

So J *is a conserved quantity.*

2.5.2. Circle Actions

We now consider the case in which the solutions of

$$\frac{d\underline{z}}{ds} = \{\underline{z}, J\} \tag{2.73}$$

are all closed curves with the same period We will call these closed orbits *loops*. The symmetry transformation is then said to be a *circle action* on phase space.

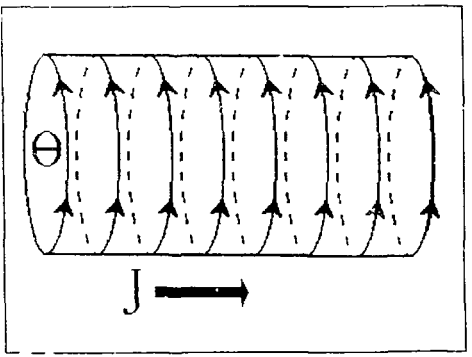

Figure 2.8: The circle action on J, θ phase space.

For example we might consider rotation by θ in J, θ space. In this case phase space looks like a cylinder (as shown in figure (2.8). The Poisson bracket is

$$\{f, g\} = \frac{\partial f}{\partial \theta}\frac{\partial g}{\partial J} - \frac{\partial f}{\partial J}\frac{\partial g}{\partial \theta}. \tag{2.74}$$

J generates the dynamics

$$\begin{aligned} \frac{d\theta}{ds} &= \{\theta, J\} = 1 \\ \frac{dJ}{ds} &= 0, \end{aligned} \tag{2.75}$$

which just rotates the cylinder.

2.5.3. Reduction by a Circle Action

In studying the dynamics of a Hamiltonian H symmetric under a circle action generated by J, we may make two simplifications which together comprise the *process of reduction*. This procedure was defined in [Marsden and Weinstein, 1974] in a more general setting that we will describe shortly. The process unifies many previously known techniques for simplifying specific examples of Hamiltonian systems.

2.5.3.1. The Reduced Phase Space

1. Because J is a constant of the motion, the surface J =constant in phase space is invariant under the dynamics and so we may restrict attention to it.

2. The symmetry property of H implies that if we take a solution curve $z(t)$ of the equation $\dot{z} = \{z, H\}$ and let it evolve for a "time" s under the dynamics $\dot{z} = \{z, J\}$ then we obtain another solution curve of $\dot{z} = \{z, H\}$. In fact the dynamics of H takes an entire loop into other entire loops.

The dynamics around loops is easy to solve for because

$$\theta = \frac{\partial H}{\partial J}. \tag{2.76}$$

Notice that θ is not uniquely defined but $\dot{\theta}$ is. We are interested in the problem of finding the dynamics from loop to loop. We want to project the original dynamics on phase space P down to a space P/S^1 whose points represent whole loops in P. Let us call P/S^1, the *space of loops* and $\pi : P \to P/S^1$, the projection mapping loops in P to points in P/S^1. For example, when $P = J, \theta$ space the projection mapping takes J, θ to J. This projection is shown in figure (2.9). Thus the second simplification is to consider dynamics on the space of loops P/S^1.

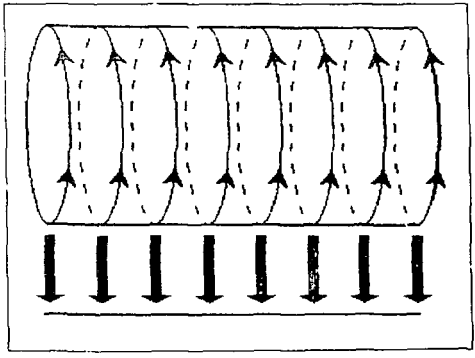

Figure 2.9: The projection of J, θ space to the space of loops

Performing both of these operations— restricting to J =constant and considering the space of loops— leaves us with a space,

$$R \equiv P/S^1|_{J=\text{constant}} \tag{2.77}$$

of two dimensions less than P, called the *reduced space*.

We have seen that the dynamics on P naturally determines dynamics on R. The key importance of R is that *R's dynamics is itself Hamiltonian*. For this statement to make sense we need to find a Hamiltonian and a Poisson bracket on R. These are the so-called *reduced Hamiltonian* and *reduced Poisson bracket*.

2.5.3.2. The Reduced Hamiltonian

The original Hamiltonian H on P is constant on loops by the symmetry condition. We may take the value of the reduced Hamiltonian at a point of R to be the value of H on the corresponding loop in P.

2.5.3.3. The Reduced Poisson Bracket

To obtain the reduced Poisson bracket of two functions f and g on R, we consider any two functions $\hat{f}$ and $\hat{g}$ on P which are constant on loops and agree with f and g when restricted to J =constant and projected by π to R. The Poisson bracket on P of $\hat{f}$ and $\hat{g}$ will be constant on loops and its value on J =constant will be independent of how $\hat{f}$ and $\hat{g}$ were extended as functions on J (because they are constant on loops: $\{\hat{f}, J\} = 0$ and $\{\hat{g}, J\} = 0$ so $\{\hat{f}, \hat{g}\}$ is independent of $\partial\hat{f}/\partial J$ and $\partial\hat{g}/\partial J$). Thus the value of the reduced Poisson bracket on R of f and g is the value on the corresponding loop in P of the Poisson bracket of any two extensions $\hat{f}, g$ that are constant on loops.

2.5.3.4. Coordinate Calculation of the Reduced Space

In examples we often introduce a coordinate θ describing the position on a loop. We may then treat P/S^1 as the set $\theta = 0$ (at least locally). In this case R is the subset θ =constant, J =constant of P. The reduced Hamiltonian on R is just the value of H on this subset of P. To calculate the value of the Poisson bracket of two functions on P on this surface, we need only their first derivatives there.

If the functions are constant on loops (i.e. independent of θ), then the derivative $\partial/\partial\theta$ is zero. The dependence on J is irrelevant, so we may take the derivative $\partial/\partial J$ to be zero. Plugging these two expressions into the Poisson bracket on P gives us the expression for the reduced Poisson bracket on R.

2.6. Example: Centrifugal Force

We consider a particle on a two-dimensional plane moving in a rotationally symmetric potential. The phase space is then $T^*\Re^2$ with coordinates (x, y, p_x, p_y). The Poisson bracket is the canonical one:

$$\{f, g\} = \frac{\partial f}{\partial x}\frac{\partial g}{\partial p_x} - \frac{\partial f}{\partial p_x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial p_y} - \frac{\partial f}{\partial p_y}\frac{\partial g}{\partial y}. \tag{2.78}$$

The Hamiltonian is taken to be

$$H = \frac{1}{2m}(p_x^2 + p_y^2) + V\left(\sqrt{x^2 + y^2}\right). \tag{2.79}$$

The symmetry on phase space is given by the evolution of the equations

$$\begin{aligned} \frac{dx}{ds} &= -y & \frac{dy}{ds} &= x \\ \frac{dp_x}{ds} &= -p_y & \frac{dp_y}{ds} &= p_x \end{aligned} \tag{2.80}$$

We may think of a point in phase space as a point in the plane (x, y) with a vector attached (p_x, p_y). The action of the symmetry is to rotate the plane about the origin, vector and all as shown in figure (2.10).

2.6.1. Angular Momentum Generates Rotations

The Hamiltonian depends only on the radial distance and the magnitude of the momentum vector and so clearly remains invariant under this rotation. The rotation is a canonical transformation with generator J satisfying

$$\frac{df}{ds} = \{f, J\} = x\frac{\partial f}{\partial y} - y\frac{\partial f}{\partial x} + p_x\frac{\partial f}{\partial p_y} - p_y\frac{\partial f}{\partial p_x} \tag{2.81}$$

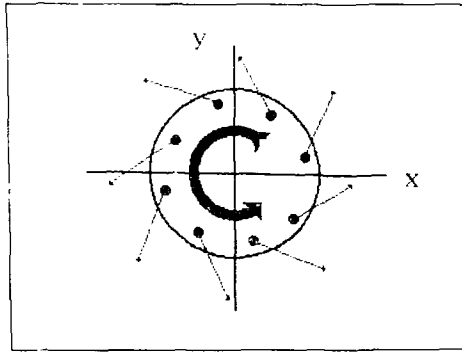

Figure 2.10: The action of the circle symmetry on momentum vectors.

for any f. Taking $f = x, y, p_x, p_y$ gives

$$\frac{\partial J}{\partial p_x} = -y \qquad \frac{\partial J}{\partial p_y} = x \qquad \frac{\partial J}{\partial x} = p_y \qquad \frac{\partial J}{\partial y} = -p_x. \tag{2.82}$$

Thus we see that the generator is $J = xp_y - yp_x$, i.e. the angular momentum. We may label a loop by the value of x, p_x, and p_y when $y = 0$ and $x \geq 0$. J on this subset is just xp_y. These then form coordinates on the space of loops P/S^1. Notice that the value of x on a loop when $y = 0$ defines the radial coordinate. For clarity, we will introduce the coordinate functions r and p_r on the reduced space for each J. A loop is labeled by r if $x = r$ when $y = 0$ on that loop. A loop is labeled by p_r if $p_x = p_r$ when $y = 0$ on that loop.

2.6.2. The Reduced Space and Bracket

To get the reduced space we set J to the constant value μ. To get the reduced bracket, it is easiest to envision the two-dimensional reduced space with coordinates (r, p_r) as a submanifold of the four-dimensional total space with coordinates (x, p_x, y, p_y) defined by:

$$\begin{aligned} x &= r \qquad p_x = p_r \\ y &= 0 \qquad p_y = \frac{\mu}{r}. \end{aligned} \tag{2.83}$$

This submanifold intersects each loop transversely in exactly one point since a loop has $y = 0$ with $x \geq 0$ at only one point. It also lies entirely in the level surface $J = \mu$ since

$$J = xp_y - yp_x = r\left(\frac{\mu}{r}\right) - 0 \cdot p_r = \mu. \tag{2.84}$$

We want to know the reduced Poisson bracket of functions $f(r, p_r)$ and $g(r, p_r)$ of r and p_r. As in section 2.5.3.3, we define the bracket of two such functions by introducing $\hat{f}(x, p_x, y, p_y)$ and $\hat{g}(x, p_x, y, p_y)$ on the four-dimensional space which are equal to f and g when restricted to the two-dimensional submanifold and which are constant on loops. We take the four-dimensional Poisson bracket of $\hat{f}$ and $\hat{g}$. This bracket is also guaranteed to be constant on loops. The restriction of the bracket of $\hat{f}$ and $\hat{g}$ to the (r, p_r) submanifold is defined to be the bracket of f and g. Since we only need the full bracket on the submanifold, we don't really need to know $\hat{f}$ and $\hat{g}$ everywhere; we need only their derivatives in each direction at points of the submanifold. We calculate these derivatives as follows. Derivatives along the submanifold are the same as for $\hat{f}$ and $\hat{g}$:

$$\frac{\partial \hat{f}}{\partial x} = \frac{\partial f}{\partial r} \tag{2.85}$$

and

$$\frac{\partial \hat{f}}{\partial p_z} = \frac{\partial f}{\partial p_r} \tag{2.86}$$

with similar expressions for g. Derivatives of $\hat{f}$ and $\hat{g}$ along the loops (and so transverse to the submanifold) must be zero since $\hat{f}$ and $\hat{g}$ are constant on loops. This implies that on the manifold

$$\begin{aligned}\left.\frac{\partial \hat{f}}{\partial \hat{\theta}}\right|_{\text{submanifold}} &= 0 \\ &= x\frac{\partial \hat{f}}{\partial y} - y\frac{\partial \hat{f}}{\partial x} + p_x\frac{\partial \hat{f}}{\partial p_y} - p_y\frac{\partial \hat{f}}{\partial p_x} \\ &= r\frac{\partial \hat{f}}{\partial y} - 0\cdot\frac{\partial f}{\partial r} + p_r\frac{\partial \hat{f}}{\partial p_y} - \frac{\mu}{r}\frac{\partial f}{\partial p_r}.\end{aligned} \tag{2.87}$$

Finally, we may extend $\hat{f}$ in the fourth direction in any way we like because the result is independent of this extension. For simplicity, let us choose

$$\frac{\partial \hat{f}}{\partial p_y} = 0. \tag{2.88}$$

This leads to

$$\frac{\partial \hat{f}}{\partial y} = \frac{\mu}{r^2}\frac{\partial f}{\partial p_r}. \tag{2.89}$$

Now we can calculate the reduced bracket:

$$\begin{aligned}\{f,g\}_R(r,p_r) &= \{\hat{f},\hat{g}\}(x=r, y=0, p_x=p_r, p_y=\frac{\mu}{r}) \\ &= \frac{\partial \hat{f}}{\partial x}\frac{\partial \hat{g}}{\partial p_x} - \frac{\partial \hat{f}}{\partial p_x}\frac{\partial \hat{g}}{\partial x} + \frac{\partial \hat{f}}{\partial y}\frac{\partial \hat{g}}{\partial p_y} - \frac{\partial \hat{f}}{\partial p_y}\frac{\partial \hat{g}}{\partial y} \\ &= \frac{\partial f}{\partial r}\frac{\partial g}{\partial p_r} - \frac{\partial f}{\partial p_r}\frac{\partial g}{\partial r} + \frac{\mu}{r^2}\frac{\partial f}{\partial p_r}\cdot 0 - 0\cdot\frac{\mu}{r^2}\frac{\partial g}{\partial p_r} \\ &= \frac{\partial f}{\partial r}\frac{\partial g}{\partial p_r} - \frac{\partial f}{\partial p_r}\frac{\partial g}{\partial r}.\end{aligned} \tag{2.90}$$

The reduced bracket in this case is just the canonical bracket on r, p_r space.

2.6 3. The Reduced Hamiltonian Gives Centrifugal Force

The reduced Hamiltonian is obtained by restricting the original Hamiltonian submanifold and is given by

$$H = \frac{p_r^2}{2m} + \left(\frac{\mu^2}{2m} \frac{1}{r^2} + V(r) \right). \tag{2.91}$$

Note the effective potential $\mu^2/2mr^2$ due to reduction, that represents the centrifugal force.

As an aside, it is interesting to note that Newton derived the centrifugal force in a more "kinetic" way which is related to averaging in statistical gas models [Brush, 1983]. He envisioned the motion of a particle along a circular path as the limit of motions on paths on inscribed polygons as the number of sides becomes infinite. Imagine a circular frictionless billiard table and a billiard ball which reflects from the wall in each traversal about the edge of the table. Equivalently a mass attached to a string whose other end is fixed (e.g., a tether ball) can undergo polygonal motion where the string is fully extended only at the vertices. The centrifugal force is just the average radial momentum transfer per unit of time. It turns out that this is independent of the number of bounces. Newton considered the average force for an inscribed square. Each impact transfers $2mv/\sqrt{2}$ units of momentum. The total radial momentum transfer is then $4\sqrt{2}mV$. The distance travelled by the ball in one traversal is $4\sqrt{2}r$. The time this takes is $v/(4\sqrt{2}r)$. The average radial momentum transfer per unit of time is then

$$4\sqrt{2}mv \left(\frac{v}{4\sqrt{2}r} \right) = \frac{mv^2}{r} \tag{2.92}$$

which is the centrifugal force. (The same result is even easier to obtain with just two bounces. The momentum transfer is $4mv$ and the time is $\pi/4r$.) The first correct account of centrifugal force was actually given by Huygens, who is better known for his ideas on wave motion.

2.7. Higher Dimensional Symmetries

"The best and latest mathematical methods to appear on the market have been used whenever possible. In doing this many an old and trusted favorite of the older generation has been forsaken, as I deemed it best not to hand dull and worn-out tools down to the next generation." [Thirring, 1978, p. iv]

Quite often physical systems are blessed with more than one dimension of symmetry. In keeping with the philosophy of not making unphysical choices, it is natural to consider the process of reduction in the presence of an arbitrary Lie group of symmetry. This program was carried out in [Marsden and Weinstein, 1974].

A Lie group is a group which is also a manifold, such that the group operations respect the smoothness structure. A good example to keep in mind is the group of rotations of three-dimensional Euclidean space. This group is denoted by $SO(3)$. It may be thought of as the space of 3 by 3 matrices (this is the 3 in $SO(3)$) which are orthogonal (this is the O in $SO(3)$) and have unit determinant (this is the S, which stands for special). The group multiplication is matrix multiplication, and the manifold structure arises from thinking of $SO(3)$ as a submanifold of the nine-dimensional Euclidean space of all 3×3 matrices.

2.7.1. Hamiltonian Symmetry

A Hamiltonian system with symmetry consists of a Poisson manifold M, a Hamiltonian H, and a group G that acts on M so as to preserve both H and the Poisson bracket $\{,\}$. The tangent space of G at its identity may be identified with the Lie algebra g of the group and represents group elements infinitesimally close

to the identity. The action of an infinitesimal element of G on M perturbs each point of M by an infinitesimal amount. Thus each element v of the Lie algebra of G naturally determines a vector field on M. The action on M of the one-dimensional subgroup to which v is tangent, is given by the flow of this vector field. That the group action preserves the Poisson bracket, implies that this vector field is actually Hamiltonian.

For example, M might be the canonical phase space of a system of particles in a central potential (such as planets around a sun). H is then symmetric under the action of $SO(3)$ which rotates the positions of the particles and the directions of their momenta. The Lie algebra $so(3)$ represents infinitesimal rotations and gives rise to a vector field on M. Rotation by a finite angle is the result of flowing along this vector field for a finite time.

2.7.2. The Momentum Map

We may therefore associate to each Lie algebra element v, a Hamiltonian function which generates its corresponding vector field (at least locally). If G is n-dimensional, and we pick a basis for g, then the group action gives us n corresponding Hamiltonian functions on M. So as not to prefer one basis over another, we collect these n numbers at each point of M into a vector. This vector naturally pairs with an element of g (to give the value of the function which generates the action of that element) and so the collection of n Hamiltonians is a vector in the *dual* of the Lie algebra g^* at each point of M. Thus with every Hamiltonian group action of G on M, there is a natural map called the momentum map from M to

g^* which collects together the generators of the infinitesimal action of G on M (see [Abraham and Marsden, 1978] p. 276)

2.7.2.1. Linear and Angular Momentum as Momentum Maps

For a mechanical system in R^3 which is translation invariant, the momentum map associates with each point in phase space the total linear momentum of the system in that state. If the Hamiltonian is rotationally symmetric, the momentum map gives the total angular momentum in each state (so the angular momentum isn't naturally a vector in R^3; rather it takes its values in the dual of the Lie algebra of the rotation group $so(3)^*$). When we talked about reduction in the one dimensional case above, the generator J of the action was the momentum map.

2.7.3. Non-commutativity as the Obstruction to Reduction

Does reduction work for higher dimensional symmetries? If the group is commutative, we may apply the one-dimensional procedure repeatedly to eliminate two dimensions of phase space for each dimension of symmetry. If we are able to eliminate all dimensions of phase space in this way, the system is called *integrable* (see [Abraham and Marsden, 1978] p. 393). If the group orbits are bounded, then one can prove that the group is a torus in this case, assuming that the "periods" are constant (see [Arnold, 1978] p. 271). Locally we may define angle variables on the toroidal group orbits and the corresponding action variables form the momentum map. Recall that there were two steps in the reduction of systems with one dimension of symmetry, each of which eliminated one dimension of the phase space

and that either could be performed first. One was to restrict to a level set of the generating function and the other was to drop down to the orbit space (space of loops). For non-commutative groups, we may again perform either of these two operations, but each gets in the way of subsequently performing the other. The main issue here is that while the Hamiltonian is invariant under the group action, the momentum map is not. Consider the example of a mechanical system in a spherically symmetric potential so that the rotation group acts on phase space as a symmetry and the momentum map is the total angular momentum. While the energy is left unchanged as we rotate the state, the angular momentum is rotated just like a vector in R^3. This action of $SO(3)$ on the dual of its Lie algebra is known as the *coadjoint action*.

2.7.4. The Adjoint and Coadjoint Actions

Let us digress a bit on the structure of Lie groups to make this point clearer. We will use the rotation group $SO(3)$ as an example. As shown in the diagram in figure (2.11), every Lie group has three natural actions on itself. If h is an element of G, then we may multiply on the left by h to get the action $L_h \cdot$, we may multiply on the right by h^{-1} (the inverse is chosen so that $R_{fh} = R_f\ R_h$) to get R_h, and coojugate by h (i.e. $c \mapsto hch^{-1}$) to get the action $AD_h \cdot$.

Let us consider these three actions for $SO(3)$ when h is a rotation by $\pi/2$ about the $\hat{z}$ axis (say, clockwise looking down $\hat{z}$). Let the actions be applied to c, a rotation by $\pi/2$ about the $\hat{x}$ axis. $L_h \cdot c = hc$ means first rotate by c and then by h. c sends $\hat{x}$ to $\hat{x}$ which h sends to $-\hat{y}$, c sends $\hat{y}$ to $-\hat{z}$ which h sends to $-\hat{z}$,

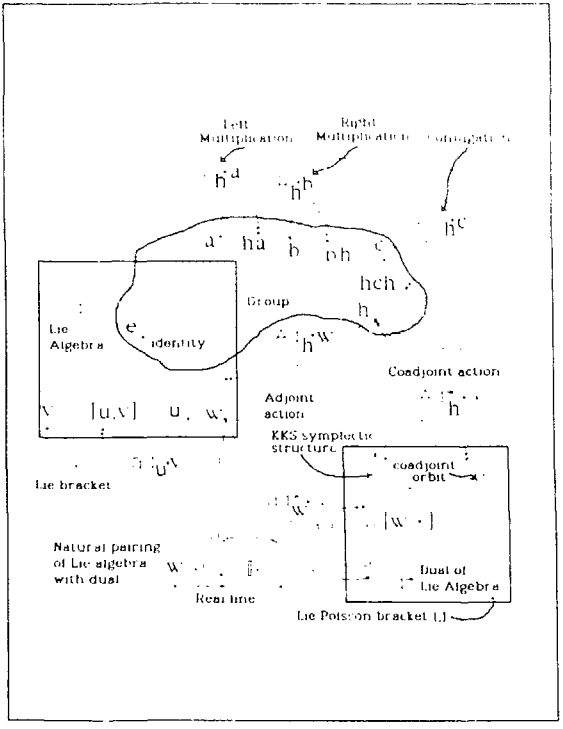

Figure 2.11: Some Natural Group Actions

and c sends $\hat{z}$ to $\hat{y}$ which h sends to $\hat{x}$. The net result is a rotation about the axis containing $\hat{x} + \hat{y} + \hat{z}$ by an angle of $2\pi/3$.

$R_h \cdot c = ch^{-1}$ means rotate by h^{-1} and then by c. h^{-1} sends $\hat{x}$ to $\hat{y}$ which c sends to $\hat{z}$, h^{-1} sends $\hat{y}$ to $-\hat{x}$ which c sends to $-\hat{x}$, and h^{-1} sends $\hat{z}$ to $\hat{z}$ which c sends to $\hat{y}$. The net result is a rotation about the axis defined by $\hat{x} - \hat{y} - \hat{z}$ by $2\pi/3$.

$AD_h \cdot c = h \cdot c \cdot h^{-1}$ means first do h^{-1}, then c, and then h. h^{-1} sends $\hat{x}$ to $\hat{y}$ which c sends to $-\hat{z}$ which h sends to $-\hat{z}$, h^{-1} sends $\hat{y}$ to $-\hat{x}$ which c sends to $-\hat{x}$ which h sends to $\hat{y}$, and h^{-1} sends $\hat{z}$ to $\hat{z}$ which c sends to $\hat{y}$ which h sends to $\hat{x}$. The net result is a rotation by $\pi/2$ about the $-\hat{y}$ axis. It is no accident that this is where h sends $\hat{x}$ which is c's axis.

Conjugation captures the noncommutativity of the group that is at issue here. $AD_h \cdot$ leaves the identity invariant (since $h \cdot e \cdot h^{-1} = e$). We may therefore take the derivative of $AD_h \cdot$ at the identity to get a linear map from the Lie algebra to itself denoted $Ad_h \cdot$. Ad is actually a representation of G on its Lie algebra and is sometimes called the fundamental representation. For rotations, Ad_h takes an infinitesimal rotation about an axis v to an infinitesimal rotation about the axis $h \cdot v$.

If we take the derivative of $Ad_h \cdot$ in the h variable, we get an action ad of the Lie algebra on itself. The action of an element $u \in g$ is none other than Lie bracket with u, i.e.:

$$ad_u \cdot v = [u, v]. \tag{2.93}$$

For rotations, this action is given by the cross product. An infinitesimal rotation about axis u followed by an infinitesimal rotation about axis v differs from first

rotating about v and then about u by an infinitesimal rotation about the axis u × v.

We have seen that the dual of the Lie algebra g^* plays an important role in Hamiltonian symmetries. Any time you have a linear transformation L acting on a vector space V, you can define its adjoint L^* acting on V^* by requiring that

$$\langle L^* \lambda, v \rangle = \langle \alpha, Lv \rangle. \tag{2.94}$$

The adjoint of Ad_h is called the coadjoint action of G on the dual of its Lie algebra g^* and is written Ad^*_h. The action of the rotation group on angular momenta that we discovered above is an example of this. The rotation h acts on the angular momentum α to give an angular momentum $Ad^*_h \cdot \alpha$ which is rotated by h.

2.7.4.1. Equivariance of the Momentum Map

One usually requires that a momentum mapping be *equivariant* as in this example (see [Abraham and Marsden, 1978] p. 269). This means that the value of the momentum map varies as the group acts on the phase space according to the coadoint action:

$$J(g \cdot x) = Ad^*_g \cdot J(x). \tag{2.95}$$

The interpretation of this for the rotation group is as follows: x is a point in phase space. g is a rotation (i.e. an element of $SO(3)$). $g \cdot x$ is the point in phase space we get to by applying the rotation g (for a mechanical system, $g \cdot x$ just has each of its position and momentum variables rotated by g). $J(g \cdot x)$ is the angular momentum of the rotated state. It is naturally an element of the 3-dimensional dual of the Lie algebra of $SO(3)$. $J(x)$ is the angular momentum of the original state. Ad^*_g

is the coadjoint action of the rotation g on the dual of the Lie algebra of $SO(3)$. Here this action just rotates the angular momentum vector by g. That the action is equivariant says that the angular momentum of the system rotated by g is the rotation by g of the original angular momentum.

2.7.5. Multidimensional Reduction using a Coadjoint Isotropy Subgroup

Let us now try to mimic the reduction procedure in this noncommutative case. First we restrict attention to the subset of phase space

$$J = \mu, \tag{2.96}$$

where μ is a constant element of $\mathfrak{g}^*$. The dynamics restricts to this subset because J is a constant of the motion. The whole group G does not act on this subset however, because a general element of G will change the value of J. The subgroup of G which leaves μ invariant under the coadjoint action (known as μ's isotropy subgroup G_μ) will act on this subset, and we may drop the dynamics down to its orbit space. The resulting space,

$$J^{-1}(\mu)/G_\mu, \tag{2.97}$$

has a natural symplectic structure, and the Hamiltonian restricted to it generates the projected dynamics. For the rotation group example, we restrict to states with a given total angular momentum (eliminating 3 dimensions) and then forget about the angle of rotation about the axis defined by that angular momentum (eliminating one more). The result is a phase space of four dimensions lower than we started with

2.7.6. Multidimensional Reduction using Coadjoint Orbits

We may obtain the same result in another way. Consider the orbit of a particular element μ of the dual of the Lie algebra under the coadjoint action. This *coadjoint orbit* O_μ has a natural *symplectic structure* which we will discuss momentarily (see [Abraham and Marsden, 1978] p. 302). For the rotation group the coadjoint orbits are spheres of constant total angular momentum (and the origin). This is because the rotation group acts on elements of the dual of the Lie algebra by rotating them, and the set of all vectors reachable by rotation from a given vector is a sphere. The orbit space of M modulo G has a natural Poisson structure (the bracket of G invariant functions is G invariant) which is not typically symplectic. The symplectic leaves of this structure project onto the coadjoint orbits under the momentum map. The inverse image of a whole coadjoint orbit under the momentum map is invariant under the group action on M. The orbit space

$$J^{-1}O_\mu/G \tag{2.98}$$

is the same reduced space we constructed above. For the rotation group this consists of restricting to states with a given total magnitude of angular momentum and then modding out by the whole rotation group.

2.7.7. The Lie-Poisson Bracket and Group Configuration Spaces

An important example of reduction applies to mechanical systems whose configuration space is the symmetry group itself. We will see that the free rigid body and the perfect fluid are examples of this type in the next two sections, a fact first

discussed in [Arnold, 1966]. The phase space M is then T^*G and the G action is the canonical lift to T^*G of left or right multiplication. The G orbits have one point in each cotangent fiber (i.e. the group action associates with a given momentum in a given configuration exactly one momentum in each other configuration) and so we may identify the orbit space with the cotangent space at the identity, i.e. the dual of the Lie algebra. If the group identity represents a reference configuration, then we may use the group action to identify momenta in an arbitrary configuration with momenta in the reference. If we apply the process of reduction to this setting, this shows that the orbit space of G acting on T^*G may be identified with the dual of the Lie algebra g^*. If we forget about the configuration and identify momenta with momenta at the reference, then our phase space becomes the space of momenta at the reference.

The momentum map for the group action is then the identity. The coadjoint orbits receive a natural symplectic structure, being the reduced spaces. These symplectic structures are known as Kirillov-Kostant-Souriou (KKS) symplectic structures. If we just consider the orbit space T^*G/G, then we obtain a natural Poisson bracket on g^* already known to Sophus Lie ([Weinstein, 1983b]) and so called the Lie Poisson bracket. Explicitly it is

$$\{f,g\}(\alpha) = \langle \alpha, [\frac{\delta f}{\delta \alpha}, \frac{\delta g}{\delta \alpha}] \rangle, \tag{2.99}$$

where $\alpha \in g^*$, f and g are functions on g^*, $[,]$ is the Lie algebra bracket, and $\langle , \rangle$ is the natural pairing of g and g^*. This bracket is behind many of the nontrivial Poisson structures recently discovered in various areas of physics.

2.7.8. Euler's Equations for the Free Rigid Body

As an example, let us consider Euler's equations for the free rigid body (see [Abraham and Marsden, 1978] p. 311). To specify the configuration of a free rigid body, we give a reference configuration and every other configuration is uniquely specified by the element of the rotation group that acts on the reference to give that configuration. The configuration space is therefore identifiable with the group $SO(3)$ itself. As the body rotates in some manner, the representative point in $SO(3)$ moves along a curve in $SO(3)$. The angular velocity of the body (i.e. the velocity in this configuration space) in a given configuration represents the first order change in configuration as we let it evolve for a short time. Two evolution curves through a given configuration point have the same angular velocity iff they are tangent to first order. In section 2.2.6 we defined a tangent vector to be just such an equivalence class of curves. Geometrically, then, we may identify the angular velocity in a given configuration with a tangent vector to $SO(3)$ based at the point representing that configuration. Therefore the state including the angular velocity is naturally a point in $TSO(3)$.

The angular momentum is obtained by acting on the angular velocity by the moment of inertia tensor. Since the moment of inertia tensor pairs with two copies of the angular velocity to give the kinetic energy which is a scalar, we see that both of its indices are covariant (i.e. it has two slots for vectors). The angular momentum arises from filling only one of the slots and is therefore a covector (i.e a one-form). For general Lagrangian systems the momentum is defined in terms of the Lagrangian as $p = \frac{\partial L}{\partial v}$. Since L is a scalar, this again shows that p is a one-

form (which pairs with vectors to give the first-order change in L when the velocity is varied along the vector.) In section 16.5.11, this kind of map is defined as the Legendre map generated by L from the tangent space to the cotangent space. The moment of inertia tensor (and in general the bilinear kinetic energy form) plays the role of a metric on configuration space which converts velocity vectors to momentum one-forms. For example, in the simple free particle relation $p = mv$, we should view the mass as a tensor (we can see that it is not a scalar by considering multiple particles with different masses). Therefore the state including angular momentum is a point in $T^*SO(3)$. This is then the natural phase space for the rigid body.

A priori, there is no way of comparing the angular velocity or momentum in one configuration with that in another. Using the group action, however, we may push all velocities to velocities at the identity (i.e. velocities on the reference configuration) which may be identified as elements of the Lie algebra. Both left and right multiplication can bring us to the identity since they each act on the group transitively. Consider a path at the identity (for example a rotation about the z axis) to which a given element of g is tangent. Left multiplication by $h \in G$ means move along the path and then rotate by h. Thus the path is associated with the body, and we get the angular velocity in the body-fixed frame. Multiplying on the right means rotate first by h, then follow the path. The path applied is independent of the configuration of the body (described by h) and so its tangent represents angular velocity in the space-fixed frame. Similarly, left multiplication gives angular momentum in the body-fixed frame and right multiplication gives it in the space-fixed frame. At a configuration represented by $h \in G$, the map from

g to g that takes spatial angular velocity to body angular velocity is the adjoint action of h. Similarly, the map from g^* to g^* that takes spatial angular momentum to body angular momentum is the coadjoint action.

The energy depends only on the angular momentum in the body (the orientation in space is irrelevant for a free rigid body) and so the Hamiltonian on $T^*SO(3)$ is invariant under the cotangent lift of left multiplication and we are indeed in the situation described above. If we drop down to the orbit space of this left multiplication, we get a Poisson bracket and Hamiltonian on the three dimensional space of angular momenta in the body. The dynamics on this space is exactly Euler's equations. The Poisson bracket is explicitly given by

$$\{J_x, J_y\} = J_z \tag{2.100}$$

plus cylic permutations. The total angular momentum $J_x^2 + J_y^2 + J_z^2$ is a Casimir function and so is automatically conserved. The coadjoint orbits (and so the symplectic leaves and bones) are the spheres of constant total angular momentum and the origin as shown in figure (2.12). The area element on each sphere is the two-form representing the KKS symplectic strucuture.

2.7.9. Euler's Equations for a Perfect Fluid

In an exactly analogous way, we may consider the Hamiltonian structure of a perfect fluid. If we choose a reference configuration, then to get any other configuration we apply a unique diffeomorphism as in figure (2.13) (volume preserving if the fluid is incompressible). Thus the configuration space may be identified with

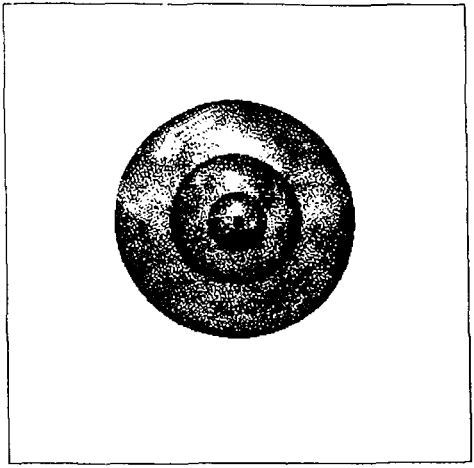

Figure 2.12: The coadjoint orbits of the rotation group.

the group of diffeomorphisms of the region in which the fluid resides. The state of the fluid plus its velocity field is represented by a point in the tangent bundle of this group. Points of the phase space represent the state of the fluid and the momentum density and so lie in the cotangent bundle of the group. Again we may identify velocities and momenta with elements of the Lie algebra and its dual by left or right multiplication. Right multiplication gives the Eulerian velocity or momentum field in space. Left multiplication gives them for material points in the reference

configuration. Here, in contrast to the rigid body case, the energy depends only on the spatial momentum (which fluid particle is where is energetically irrelevant) and so the Hamiltonian is right invariant. Dropping to the orbit space gives us dynamics for the spatial momentum density, i.e. Euler's fluid equations, in Hamiltonian form

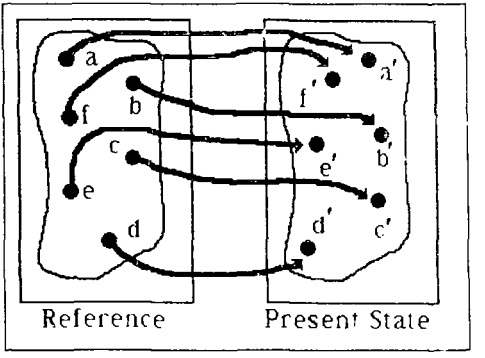

Figure 2.13: A configuration of the fluid specified by a diffeomorphism.

2.7.10. Gases and Plasmas

For gases and plasmas, the state of the system is represented by the particle distribution function on single-particle phase space. This distribution function evolves by the action of symplectomorphisms (i.e. canonical transformations) of this phase space (see [Weinstein, 1984b] for more discussion and references). The

group of symplectomorphisms has the Hamiltonian vector fields as its Lie algebra. We may identify this with the space of functions on the phase space where the Lie bracket is the Poisson bracket of functions. The dual of the Lie algebra is then densities on phase space, which we may use to describe the kinetic state of plasmas and gases. The coadjoint action just pushes the density around by the symplectomorphism. One coadjoint orbit comes from considering a delta distribution on phase space. The symplectomorphisms push it all over phase space to give a coadjoint orbit that is identifiable with the original phase space. In fact the KKS symplectic structure is exactly the original symplectic structure. This shows that every symplectic manifold is a coadjoint orbit (albeit in the dual of the Lie algebra of the infinite-dimensional group of canonical transformations of that manifold).

It is interesting to consider some of the other coadjoint orbits for this system. In chapter 9, eikonal waves are associated with certain distributions on (z, k) space whose support (i.e. the closure of the complement of the region where the region vanishes) is a Lagrangian submanifold (an N-dimensional manifold on which the symplectic structure vanishes, see section 7.1.4). The space of such distributions is a union of coadjoint orbits of the group of symplectomorphisms.

Any local piece of a Lagrangian submanifold can be made to coincide with a local piece of any other Lagrangian submanifold by a canonical transformation [Weinstein, 1977]. In fact, the symplectic structure of a small neighborhood of a small piece of Lagrangian submanifold is identifiable with the canonical symplectic structure of a small neighborhood of the zero-section of the cotangent bundle of the piece of Lagrangian submanifold. In other words, small regions of any Lagrangian

submanifold may be straightened out to a piece of $\mathbb{R}^N \subset \mathbb{R}^{2N}$. This means that there are no local invariants in a symplectic manifold other than dimension and no local structure associated with a Lagrangian submanifold other than the property of being Lagrangian. (As a simple example of something with no local constraints yet still obeying a global constraint, consider a water balloon. The surface of the balloon is arbitrary locally but globally must enclose a given volume.) This behavior is in great distinction to the case of Riemannian geometry where local pieces of space are invariantly characterized by various curvatures and even flat submanifolds may sit in space in locally different ways. The less rigid nature of symplectic geometry helps to give it its characteristic feel and makes the study of phase space quite different from the study of spacetime.

In section 10.1 we show that an arbitrary diffeomorphism of a manifold M may be extended to a symplectomorphism of T^*M which acts on the zero section according to the original diffeomorphism. We may use this type of map to vary the value of a distribution with Lagrangian support while leaving the support manifold invariant. There are global constraints on the image of a Lagrangian submanifold under a canonical transformation as well. In $\mathbb{R}^{2N}$ one can associate with each Lagrangian torus, the actions of each of the non-contractible loops lying in it. These actions cannot change under a symplectomorphism. In non-trivial topologies, one must talk about the change in action of a loop under a deformation because there may be no disk with the loop as its boundary (the change in action is simply the symplectic area of the region swept out by the loop under deformation). If the deformations are exact symplectomorphisms (e. g. symplectomorphisms of a

simply connected region), then the action cannot change. One may show that, other than this constraint on actions, each Lagrangian torus can be taken to every other nearby Lagrangian torus.

There are some subtleties in generalizing from a delta function at a point to delta functions defining a submanifold. In general, a distribution is a linear functional satisfying certain continuity criteria that associates a real number with each smooth function (satisfying certain vanishing criteria, see for instance [Hörmander, 1983]). We think of a distribution as something we may integrate smooth functions against. A delta function at a point $p \in M$ in a manifold just assigns to each smooth function f its value $f(p)$ at the point p. The analog of a delta function whose support is a Lagrangian submanifold would associate with each function on M its integral over the Lagrangian submanifold. This requires a measure on the Lagrangian submanifold. The space of Lagrangian supported delta-like distributions may be identified with the smooth Lagrangian embeddings of N dimensional manifolds *with measures on them*. Let us call these measured Lagrangian submanifolds. The space of measured Lagrangian tori with given actions is thus a coadjoint orbit. Similarly, the space of measured loops with given action is a coadjoint orbit and therefore a symplectic manifold.

2.8. Geometric Hamiltonian Perturbation Theory

Let us now relate this geometric Hamiltonian mechanics to the geometric perturbation theory we discussed earlier. We will see that the Jth order perturbed dynamics has a natural Hamiltonian structure if the exact dynamics does. More details on the ideas of this section are given in section 4.6.

The first thing to note is that the path space dynamics is Hamiltonian. This is not surprising if we think of the path space as a kind of direct integral of the phase spaces at each ϵ. The dynamics at different ϵ's are completely independent (except for the fact that the paths are smooth). If we had the product of only two Hamiltonian systems (instead of a continuum of them) then we would get the correct dynamics from a symplectic structure which is the sum of the pullback to the product of the individual symplectic structures and a Hamiltonian which is the sum of the pulled back Hamiltonians. Extending this construction to a continuum of multiplicands leads to the symplectic structure

$$\tilde{\omega}_p(\tilde{V}_1, \tilde{V}_2) \equiv \int_0^1 \omega_{p(\epsilon)}(V_1(\epsilon, p(\epsilon)), V_2(\epsilon, p(\epsilon))) \, d\epsilon \tag{2.101}$$

The analog of the sum of Hamiltonians is

$$\tilde{H}(p) \equiv \int_0^1 H(\epsilon, p(\epsilon)) \, d\epsilon \tag{2.102}$$

The dynamics these two generate indeed the correct path space dynamics. In the case of a product of a finite number of Hamiltonian systems, we are actually allowed to take any linear combination of the symplectic structures (instead of a straight sum), as long as no coefficient vanishes and we take the same linear combination

of Hamiltonians. If s coefficient vanishes, that factor has no dynamics. For our perturbation dynamics then, we want to ignore the region in the interval that is away from $\epsilon = 0$.

In fact if we insert the Jth derivative of a delta function into the integrals in (2.101) and (2.102) we get the correct perturbation dynamics on JM. If the Poisson bracket on M is $\{x^a, x^b\} = J^{ab}$ then the bracket on JM is

$$\{x_k^a, x_m^b\} = J^{ab}\frac{k!\,m!}{J!}\delta_{k,J-m} \tag{2.103}$$

and the Hamiltonian is

$$\tilde{H}(x_0, \ldots, x_J) \equiv \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} H(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J). \tag{2.104}$$

Together these give the correct perturbation dynamics. Notice that the 0th order variables are paired with Jth order variables, 1st order with $J-1$st order, etc.

From the above coordinate description it is not clear that this bracket is in fact intrinsic. We may show this by considering the *iterated tangent bundle* to M. The tangent bundle to a symplectic manifold has a natural symplectic structure. if ω is the structure on M, then we may use it to identify TM and T^*M. T^*M has a natural symplectic structure, which we defined in section 4. The structure on TM is obtained by pulling T^*M's back using the identification supplied by ω. This operation may be iterated to give symplectic structures on the iterated tangent bundles: TTM, $TTTM$, $TTTTM$, etc. The Jth order jets naturally embed into the Jth iterated tangent bundle. If the symplectic structure on T^JM is pulled back to JM we obtain the jet Poisson bracket in equation (2.103).

The symplectic structure on TM may be thought of as the first derivative of the original symplectic structure Kijowski and Tulczyjew 1979. The jet bracket may be thought of as the Jth derivative (whose J sheets spaced evenly in $J \cdot M$. The path dynamics projects down to the product of these sheets. We may take the product symplectic structure on JM with arbitrary coefficients in front of the individual summand symplectic structures. If these coefficients are chosen to give a nonsingular resultant symplectic structure as the sheet spacing goes to zero, we again obtain the jet symplectic structure and Hamiltonian. Such an argument given in detail in section 4.7 shows that the perturbation brackets and Hamiltonian are in essence Jth derivatives of the path structures. This approach is similar to studies of *finite differences* as *approximants of ordinary derivatives*.

2.8.1. Linearized Dynamics at a Fixed Point from Jet Bracket

We have seen that when the Poisson bracket is degenerate, non-degenerate symplectic leaves and bones are *injected* into the Poisson manifold as submanifolds. If a closed two-form is degenerate, then we *project* out the degenerate directions to obtain a symplectic manifold. The fact that the two-form is closed implies that the annihilated directions satisfy the conditions of Frobenius's theorem and so lie tangent to smooth submanifolds which we may then project along (at least locally). We have used an example of this construction above. If we insert the Jth derivative of a delta function into the path symplectic integral (2.101), we obtain a degenerate, closed two-form on the path space $P_1 M$. The projection eliminating the degenerate directions is exactly the projection from path space down to the jet

space JM. The resulting symplectic structure is the jet perturbation structure. If we have a Hamiltonian system with an invariant submanifold, we may attempt to obtain the restricted dynamics in Hamiltonian form by pulling back the symplectic structure. The resulting two-form will be closed but may not be non-degenerate. If there are more globally, we may apply the above projection. A special case of this demonstrates that the jet construction contains as a special case the linearized dynamics of a Hamiltonian system around a fixed point. We consider the 2-jet space $2M$. The submanifold of jets with base point equal to the fixed point is an invariant submanifold. Because the zero order base directions are paired with the second order directions in (2.103), restricting to a given basepoint makes the second order directions degenerate. Projecting these out leaves us with only the first order jets at the fixed point (i.e. the tangent space there). These are paired with themselves by the second order bracket according to the original symplectic structure at the fixed point. The second order Hamiltonian (2.104) gives the quadratic piece of the Taylor expansion in the x_1 variables. Together these give the linearized flow in the tangent space of the fixed point as a Hamiltonian system. The situation in Poisson manifolds is more complex [Weinstein, 1983a]. If the fixed point is in a symplectic leaf (as opposed to a bone), we take the Poisson bracket at the point, the quadratic part of the Hamiltonian in the leaf direction, and the linear part of the Hamiltonian across leaves. The bones are more difficult.

We can also consider the same approach to the evolution of J-jets based at a zero order fixed point. The zero order variables are now paired with J-th order variables. When we restrict to a given zero order point, the J-th order directions

become degenerate for the J-jet symplectic structure. When we quotient out this degeneracy we are left with only the first through $J-1$st order variables. The symplectic structure is the $J-2$ jet structure up to numerical coefficients and the Hamiltonian is the J-th derivative.

We may also use this approach to obtain a symplectic structure for linearizing about a single non-fixed unperturbed orbit. We restrict the space of 2-jets to jets that originate on the orbit of interest. There is now a $2N-1$ dimensional degenerate foliation which we may quotient by. The result is a symplectic space whose points indicate position along the unperturbed orbit, an element of the tangent space at that position, and a single quotiented 2-jet variable that pairs non-trivially with the orbit position variable. The Hamiltonian is again the second derivative of the original Hamiltonian:

$$\begin{aligned}\tilde{H}(x_0,x_1,x_2) &= \frac{d^2}{d\epsilon^2}\Big|_{\epsilon=0} H(\epsilon, x_0+\epsilon x_1+\frac{\epsilon^2}{2}x_2)\\ &= \frac{d^2}{d\epsilon^2}\Big|_{\epsilon=0}\left(H(\epsilon,x_0)+\epsilon\frac{\partial H}{\partial x_0}(\epsilon,x_0)x_1+\frac{\epsilon^2}{2}\frac{\partial^2 H}{\partial x_0^2}(\epsilon,x_0)x_1^2+\frac{\epsilon^2}{2}\frac{\partial H}{\partial x_0}x_2\right)\\ &= \frac{\partial^2 H}{\partial\epsilon^2}H(\epsilon,x_0)+\frac{\partial}{\partial x_0}\frac{\partial H}{\partial\epsilon}(\epsilon,x_0)x_1+\frac{\partial^2 H}{\partial x_0^2}x_1^2+\frac{\partial H}{\partial x_0}x_2.\end{aligned} \tag{2.105}$$

The dependence of this on x_2 is only through the term

$$\frac{\partial H}{\partial x_0}(0,x_0)\cdot x_2.$$

$\tilde{H}$ is thus constant under variations of x_2 in the annihilator subspace of dH. But these are exactly the directions symplectically orthogonal to the zero order orbit and so span the degenerate foliation. The linearized dynamics about an unperturbed orbit therefore has an invariant formulation.

2.8.2. Symmetry and Perturbation Theory

We have seen how important symmetry and its related concepts are in Hamiltonian mechanics. How do the symmetry operations intermix with the perturbation operations? A Hamiltonian G action on M lifts to both the path space PM (just push the whole path around by the group action) and the jet space JM (just push the jet around). The corresponding momentum maps are just the integral along a path of the M momentum map and the same integral with the Jth derivative of a delta function thrown in. Both are equivariant.

When considering reduction we quickly see that these groups are not of high enough dimension. A 4-dimensional phase space with a 1-dimensional symmetry drops down to 2 dimensions. The first order perturbation space has 8 dimensions. In the presence of symmetry we expect to be able to drop this down to the first order perturbation space of the 2-dimensional reduced space. The above group action can only eliminate 2 dimensions instead of the needed 4 and so we expect a larger group to act. This is indeed the case. It makes sense to multiply two paths in a group by multiplying pointwise. Thus PG is an infinite dimensional "Lie" group and its "Lie" algebra is the path space of G's Lie algebra g. PG has a Hamiltonian action on the path space PM by multiplying the point $p(\epsilon)$ by the group element $g(\epsilon)$. The momentum map sends a path in M to a path in g^* gotten by applying M's momentum map to each ϵ. In an exactly analogous way, we may define the group JG of J-jets of paths in G with Lie algebra being J-jets of paths in g. This acts in a Hamiltonian and equivariant way on the perturbation space JM. The momentum map is obtained by extending a jet to any consistent path, taking the

path momentum map to I^*g^* and dropping down to Jg^*.

The process of reduction commutes with taking the path space or jet space. The jet or path space of the reduced space is the reduced space of the jet or path space by the jet or path group.

We have seen the central importance of the dual of the Lie algebra and the coadjoint orbits with their KKS symplectic structure for physics. We have seen that any symplectic manifold may be thought of as a coadjoint orbit in the dual of the Lie algebra of some group. It turns out that if M is a coadjoint orbit in the dual of G's Lie algebra then the perturbation space JM with the jet symplectic structure are naturally a coadjoint orbit in the dual of the Lie algebra of the jet group JG and the jet bracket (2.103) is the natural KKS symplectic structure. This is shown in section 4.8.6.

These relations are at the heart of a new framework for singular Lie transform perturbation theory about which we will report in chapter 5. Here we discuss only the first order method of averaging.

2.9. The Method of Averaging for Hamiltonian Systems

Many of the interesting physical regularities we find in diverse systems are caused by the presence of processes that operate on widely separated time scales. The basic simplification that this separation entails is that the fast degrees of freedom act almost as if the slow variables are constant and the slow degrees of freedom are affected only by the average behavior of the fast variables. Bogoliubov in particular has used this separation of scales with great success in many examples. For example he obtains the Boltzmann equation from the BBGKY hierarchy of evolution equations for correlation functions by holding the 1-particle distribution functions fixed while determining the fast evolution of the higher correlations, and then substituting the result in as the collision term driving the 1-particle evolution. One makes a similar separation in calculating fluid quantities like viscosity, thermal conductivity, diffusion or electrical conductivity from an underlying kinetic description. In studying complex situations with slow, heavy nuclei and fast, light electrons in molecular and solid state physics, one often holds the nuclei fixed, calculates the electron ground state and energy as a function of the nuclei positions and then uses them to define an effective potential in which the nuclei move (this is the Born-Oppenheimer approach).

We have seen that in the presence of an exact symmetry, the symmetry directions may be completely eliminated by the process of reduction. We will now discuss how the averaging underlying the systems with separated scales can be viewed as reduction by an approximate symmetry. One often finds that the effect of "forgetting" these degrees of freedom is to introduce an amended potential into the

Hamiltonian and a "magnetic" piece to the Poisson bracket of the reduced system. In the systems with approximate symmetry, these extra terms encapsulate the new physics revealed by the averaging procedure. Earlier we saw the centrifugal force coming out as an effective potential. We give a version of the reduction procedure which begins by including the "angle of the earth" as a dynamical variable and reduces by the earth's rotation and the rotation of the system together in the next chapter. The resulting reduced space gives the centrifugal force as an amended potential in the reduced Hamiltonian and the Coriolis force as a new term in the Poisson bracket.

2.9.1. Approximate Noether's Theorem

When we introduce a perturbation which breaks a symmetry we no longer have exactly conserved quantities. It is easy to prove an "approximate Noether's theorem", however, which says that the momentum map for a slightly broken symmetry evolves slowly:

$$X_J \cdot H = \{H, J\} = \epsilon \qquad \text{implies} \qquad \dot{J} = \{J, H\} = -\epsilon. \tag{2.106}$$

2.9.2. Hamiltonian Averaging as Reduction by a Circle Action

In the special case where the unperturbed dynamics is entirely composed of periodic orbits, the action of the orbit through each point is the momentum map of a circle symmetry of the unperturbed Hamiltonian. As we turn on a perturbation which breaks this symmetry, the motion will still be primarily around the loops.

but it will slowly drift from loop to loop. Because the symmetry is broken, different points on a loop will move toward different loops. As the perturbation is made smaller, though, phase points orbit many times near a given loop before drifting away. This suggests (correctly) that the perturbation a point feels will asymptotically be the same as the average around an unperturbed loop. Because this average is the same for all points on a loop, for small perturbations entire loops drift onto other entire loops. We may therefore drop the dynamics down to the loop space. In fact one can prove that for a general (even dissipative) system where the unperturbed dynamics X_0 is entirely composed of periodic orbits, the motion of a point under the flow of $X_0 + \epsilon X_1$ projected down to the loop space remains within ϵ for a time $1/\epsilon$ of the orbit of a corresponding point on the loop space under the flow of the average of X_1 around each loop projected down [Arnold, 1983]. In the Hamiltonian case we break the circle symmetry of H_0 to get the perturbed system $H_0 + \epsilon H_1$. We average H_1 around the loops to get $\bar{H}_1$. $H_0 + \epsilon \bar{H}_1$ is again invariant under the circle action and so we may perform reduction. The reduced dynamics is the slow dynamics on the reduced space and the fact that we may restrict to a constant value of the momentum map shows that it is actually conserved to within order ϵ for time $1/\epsilon$. The momentum map for the circle group sends each point of phase space to the action of the loop it belongs to. The action of a loop is the integral of the symplectic form ω over a disc whose boundary is the loop. Since ω is invariant under a canonical transformation, so is the value of the action of a loop (this is Poincare's first integral invariant). If the dynamics was represented by a canon transformation that took loops exactly to other loops, then the action would be an

exact constant of the motion. The true dynamics takes loops to only within ϵ of another loop after time $1/\epsilon$. This says that the action of the loop a phase space point is on changes only by ϵ as we follow the phase point for time $1/\epsilon$. The action is not invariant but is *adiabatically invariant* (i.e. the error is small for longer and longer times as $\epsilon \to 0$. Kruskal has shown that there is actually a quantity which is conserved to all orders in ϵ for time $1/\epsilon$ [Kruskal, 1962] (we give a geometric formulation of this result in chapter 5). Getting results valid for times longer than $1/\epsilon$ is extremely important physically, but so far no general theory exists. Chapter 5 includes some discussion of the relevant issues here.

2.9.2.1. Averaging and the Jet Picture

Let us relate this procedure to the perturbation structures we developed in previous sections. We have an action of the circle group S^1 on M. This lifts to an action of PS^1 on PM and JS^1 on JM. The unperturbed Hamiltonian is invariant under the S^1 action on M, but the path and perturbation Hamiltonians are not invariant under PS^1 and JS^1. We would like to change the action of PS^1 on PM so as to leave the path Hamiltonian invariant and so allow reduction. Since the resulting action should still be Hamiltonian, we look for an ϵ-dependent canonical transformation of $I \times M$ which is the identity at $\epsilon = 0$ and which pushes the PS^1 action into a symmetry. The method of Lie transforms (see0 [Cary, 1981] and0 [Nayfeh, 1973] p. 200) attempts to do this at the perturbation level, letting the canonical transformation be the flow of an ϵ-dependent Hamiltonian, which is then obtained order by order. Here we need only consider the first order group action

of $1S^1 \sim TS^1$ on $1M \sim TM$. We know that the action will be perturbed so that the value of the reduced Hamiltonian is the average of the perturbed Hamiltonian around the untransformed circles. TM has twice the dimension of M. Reducing by TS^1 eliminates 4 dimensions. The resulting dynamical vector field has no unperturbed component. One may think of this as the reason for getting results good for time $1/\epsilon$ (it is the effect of the unperturbed flow on the perturbation which causes this level of secularity). In this situation it makes intrinsic sense to project the 1st order vector field down to M, where it represents the slow dynamics.

2.9.2.2. Extensions from Loops to Tori and Energy Surfaces

A loop in a 2-dimensional phase space (like an orbit of a simple harmonic oscillator) may be thought of in 3 ways. It is 1-dimensional, 1 dimension less than 2, and half of 2. Each has an important generalization to higher dimensional Hamiltonian systems. In the presence of a slowly varying Hamiltonian, we have already seen that the action of a 1-dimensional loop is conserved. There is an analogous result for half dimensional Lagrangian tori. Kubo has shown that for a system ergodic on an energy surface (which has one dimension less than phase space), the volume enclosed is adiabatically invariant under slow variation of parameters [Kubo et al., 1965]. Roughly: since the motion is ergodic, every orbit changes according to the average of the perturbation over the energy surface; thus the entire energy surface changes by the same energy and so is taken to another energy surface; but the volume enclosed by a surface is preserved under a canonical transformation by

Liouville's theorem. For a large number of degrees of freedom, this leads to the adiabatic invariance of the entropy in statistical mechanics.

2.9.3. Pseudo-Potentials and Adiabatic Invariants

The funny potentials and Poisson brackets that result from reduction contain the average effect of the fast on the slow degrees of freedom. Capturing this effect is the content of many physically useful theories. It is interesting to note that in the late nineteenth century, the idea that all potential energies were really kinetic energies of hidden or forgotten degrees of freedom was one the the main motivations for the development of kinetic theory. We may use averaging to see how this comes about.

Because of its direct relevance to the ideas presented here, let us give a long quotation from Felix Klein's historical account of nineteenth century mathematical physics. He first introduces Routh's function R which was in between the Lagrangian and Hamiltonian in that only the first m of the n configuration space variables had their velocities Legendre transformed into momenta. He continues:

"Thus the equations split into two groups, one of the Lagrangian kind and one of the Hamiltonian. For $m = 0$ the Routh function, and therewith the system of equations, is the same as in the Lagrangian case; while for $m = n$ it is the same as in the Hamiltonian.

This system of equations now acquires particular interest through certain general fundamental concepts of mechanics that are connected with it. Namely, if R

does not explicitly contain the s, (the configuration variables whose velocity variables were Legendre transformed), then we have the special case that Helmholtz (Crelle Bd. 97, 1884) has called a cyclic system and made a deep study of, and which occurs somewhat earlier in Thomson and Tait as a "cycloidal system".

In practice this case arises when one is dealing with rotating motions of rotating bodies (e.g. fly-wheels), where the angle of rotation is the "cyclic" coordinate (so that only its corresponding momentum coordinate will come in). If the rotating body is enclosed in an opaque box, then its "hidden motion" reveals nothing more than the usual behavior that the body as a whole shows as it moves in space (a top or gyroscope). In cases like this where outside influence on the motion of the fly-wheel is excluded, the momenta corresponding to the cyclic coordinates are constant.

From these facts follow some remarkable ideas on the nature of potential energy. If we assume that the kinetic energy T decomposes into a part $T(\dot{q})$ that depends only on the velocities $\dot{q}$ and into a part $T(\pi)$ that depends only on the cyclic momenta π (thus assuming that there are no terms in which velocities $\dot{q}$ are multiplied by momenta π), then the Routh function is

$$\begin{aligned} R &= T(\dot{q}) - T(\pi) - U \\ &= T(q, \dot{q}) - T(q, c) - U(q), \end{aligned} \tag{2.107}$$

if we bear in mind the dependence of all the quantities on the coordinates q and replace the constant momenta π_s by the quantities c_s. The $q_{m+1}, \ldots, q_n$ are determined from the differential equations

$$p_\alpha = -\frac{\partial T(\dot{q})}{\partial \dot{q}_\alpha}, \quad \frac{dp_\alpha}{dt} = \frac{\partial [T(q) - (U + T(c))]}{\partial q_\alpha}. \tag{2.108}$$

Thus one gets a system of formulas that exactly corresponds to a system with n degrees of freedom and whose potential energy has been increased by T(c), the kinetic energy of hidden movements. The quantities U and T(π) are both functions of q with constant coefficients; they enter into the sum only together, not separately. Hence the question arises where we in any case have no idea of the essence of the potential energy–of whether every quantity that appears in mechanics as "potential energy" is actually a kinetic energy caused by a hidden, cyclic, so-called "ignored" motion. Like a fata morgana, the possibility of a purely kinetic theory of matter appears in the distance.

This general idea was first presented in 1888 by J. J. Thomson in his book Applications of Dynamics to Physics and Chemistry *(lecture at Cambridge in 1886, then in the Philosophical Transactions 1886-87). But in special cases it had already been pursued by Wil[illegible]m Thomson (=Lord Kelvin), for example in his address to the British Association in Montreal in 1884, which he prudently titled "Steps to a kinetic theory of matter" (Math. and Phys. Papers, Volume 3, p. 366). This idea was finally worked out for closed systems in Heinrich Hertz's posthumous work of 1904,* Die Prinzipien der Mechanik *[The Principles of Mechanics]."*

2.9.3.1. Ping-pong balls and One-dimensional Gases

If we slowly move a ping pong paddle up and down from a table with a ping pong ball bouncing very rapidly between the paddle and the table, then we will feel a varying force due to the average momenta imparted due to the impacts of the ball. In phase space the ball describes a rectangle and so the action is given by

$J = 4LmV$ where L is the distance from the paddle to the table and V is the speed of the ball. Because this is invariant under slow paddle movements, the ball velocity goes as $1/L$. The momentum transferred on each impact is $2mV$ and there are $V/2L$ impacts per unit of time, so the average force felt goes like $V^2/L \sim 1/L^3$. Thus starting with no potential energy at all, we end up with a $1/L^3$ effective potential for the paddle.

It is well known that slow compression of an ideal gas keeps the quantity pV^γ constant, where the ratio of specific heats γ depends on the properties of the gas (this follows from the adiabatic invariance of the entropy). Our single particle result is exactly this requirement for a one-dimensional gas with $\gamma = 3$. Since particles do not interact in an ideal gas, it makes sense that each particle should reflect the behavior of the entire gas. (A similar result holds for radiation, where the adiabatic compression of a container containing black body radiation acts on each normal mode separately and yet the overall effect keeps the spectrum black body but at a different temperature.)

2.9.3.2. Oscillatory Stabilization

For a harmonic oscillator, the energy is the product of the action and the frequency: $H = \omega J$. If we have a weight, hanging on a string and undergoing small amplitude oscillations as we slowly pull the string, the change in pendulum energy is the change in $J\omega$. J remains constant and $\omega \sim \sqrt{g/L}$ so we feel a $1/\sqrt{L}$ potential. We get other potentials if we ask for the force we feel if we tune a guitar string as someone plays it or the acoustic pressure on the water if we fill up a

shower as someone sings in it. The effective force due to the fast degrees of freedom may sometimes stabilize an unstable fixed point of the slow system. Ordinarily an inverted pendulum is unstable and falls to the position with the weight hanging downward. If we shake the support of the pendulum periodically hard enough and fast enough, the inverted position is stabilized! An even more spectacular version of this effect occurs if you shake an inverted cup of fluid and stabilize the Rayleigh-Taylor instability which ordinarily causes the fluid to spill out (it is easiest to actualiy do the experiment with a high viscosity fluid like motor oil). The idea of RF stabilization is to stabilize unstable modes of a plasma (say in a tokamak) by bathing it in a high frequency radio wave. Some of the modern airplanes with wings in a forward facing delta are actually operated in an aerodynamically unstable regime that is stabilized by the fast dynamics of a computer controlled feedback loop. This allows for great maneuverability (since the plane would like to turn anyway!).

2.9.3.3. Multiple Space and Time Scales

Quite often it is very useful to split out the main dynamics of a system and linearize the rest, treating them as fast oscillations. Thus one takes a fluid, elastic, or plasma medium and treats its evolution as slow overall development of the background medium with fast oscillations occuring on top of it. The effect of the oscillations is to change or *renormalize* the dynamics of the background. N.G. van Kampen has called into question the usual treatments of constrained mechanical systems [van Kampen, 1983]. One usually just writes down the Lagrangian for

such a system in generalized coordinates which respect the constraints. Physically, though, one supposes that there is some large potential normal to the constraint surface. The system will execute rapid oscillation in the normal direction and slow evolution along it. If the width of the constraining potential well varies with the mechanical coordinates, then as we have seen the adiabatic invariance will give rise to a new pseudopotential which affects the mechanical motion. In a plasma we treat the slowly varying background as a dielectric medium in which waves propagate according to WKB theory. The waves affect the background (introducing a radiation pressure in the dynamics) via ponderomotive forces. If we have a charged particle in the presence of a wave with a slowly varying amplitude, the particle will oscillate back and forth with the wave. It feels more of a push in going down an amplitude gradient than in going up one, leading to an overall average force described by the *ponderomotive potential*. This kind of separation is the basis of plasma quasilinear theory. We have extended the geometric perturbation theory to some of these singular perturbation problems. Chapter 8 gives a Hamiltonian treatment of an eikonal theory for linear or nonlinear waves (which is related to the averaged Lagrangian treatment of Whitham [Whitham, 1974]). Here let us demonstrate the efficacy of a global geometric approach only with the simple example of $E \times B$ drift. A charged particle restricted to a plane with a constant perpendicular magnetic field executes perfect circles. If there is, in addition, an electric field then the radius of the circles is greater in low potential regions and smaller in high potential regions. Thus the circular orbits do not close and the particle drifts perpendicularly to the electric field. A Hamiltonian treatment of more complicated versions of this so-called guid-

ing center motion has been previously given [Littlejohn, 1983]. This work required great cleverness in the choice of physically relevant coordinates. We would like to demonstrate, in this simple version, how a coordinate free approach would lead us to the correct answer, with no previous knowledge.

2.10. Example: $E \times B$ Drift

In the simplest situation we have a charged particle in the x, y plane moving in the presence of a constant magnetic field B which points in the $\hat{z}$ direction and a small constant electric field ϵE which points in the $\hat{x}$ direction. We introduce the phase space $P \sim T^*R^2$ with coordinates (x, y, p_x, p_y) (we use mechanical momenta $p = mv$ here). The correct dynamics in the presence of a magnetic field may be described in a Hamiltonian formulation in two ways. The standard approach is to introduce the unphysical vector potential A and to work with canonical momenta $p = mv - (e/c)A$. Here we use the physical momenta and magnetic field, but a noncanonical Poisson bracket:

$$\{f, g\} = f_x g_{p_x} - f_{p_x} g_x + f_y g_{p_y} - f_{p_y} g_y + \frac{eB}{c}(f_{p_x} g_{p_y} - f_{p_y} g_{p_x}). \tag{2.109}$$

We obtain the correct dynamics in this case with the Hamiltonian

$$H = H_0 + \epsilon H_1 = \frac{1}{2m}(p_x^2 + p_y^2) - \epsilon e E x. \tag{2.110}$$

The dynamics is then

$$\begin{aligned} \dot{x} &= \frac{p_x}{m} \qquad \dot{y} = \frac{p_y}{m} \\ \dot{p}_x &= \frac{eB}{mc} p_y + \epsilon e E \qquad \dot{p}_y = -\frac{eB}{mc} p_x. \end{aligned} \tag{2.111}$$

The unperturbed situation here is just a charged particle on a plane in a constant magnetic field. Every orbit in this situation is a closed loop. Thus the unperturbed system has a circle symmetry:

$$\begin{aligned} \dot{x} &= \frac{p_x}{m} \qquad \dot{y} = \frac{p_y}{m} \\ \dot{p}_x &= \frac{eB}{mc} p_y \qquad \dot{p}_y = -\frac{eB}{mc} p_x \end{aligned} \tag{2.112}$$

The generator of this symmetry (i.e. the momentum map) is none other than the unperturbed Hamiltonian itself:

$$H_0 = \frac{1}{2m}(p_x^2 + p_y^2). \tag{2.113}$$

This is because non-relativistic motion in a constant magnetic field has the remarkable property that the period of all orbits is the same (we could introduce a normalization constant to make it 1 or 2π if desired). Let us obtain the reduced phase space and Poisson bracket for this symmetry action. First we look at the space of loops P/S^1. Each circular particle orbit has exactly one point where $p_y = 0$ and $p_x \geq 0$. We may label a loop by the values of x, y, p_x at this point. Next we restrict to the set where the momentum map is a constant: $H_0 = \alpha$. The reduced space is

$$R = P/S^1|_{H_0=\alpha} \tag{2.114}$$

and may be coordinatized by the values of x and y when $p_x = \sqrt{2m\alpha}$ and $p_y = 0$. The reduced Poisson bracket $\{,\}_\alpha$ of two functions $f(x,y)$ and $g(x,y)$ is obtained by extending them to P in such a way that

$$\left.\frac{\partial \hat{f}}{\partial p_x}\right|_{p_x=\sqrt{2m\alpha},\ p_y=0} = 0 \tag{2.115}$$

and

$$\{\hat{f}, H_0\} = 0 = \frac{\sqrt{2m\alpha}}{m}\frac{\partial \hat{f}}{\partial x} - \frac{eB}{mc}\sqrt{2m\alpha}\frac{\partial \hat{f}}{\partial p_y} \tag{2.116}$$

Thus we replace $\partial/\partial p_x$ by 0 and $\partial/\partial p_y$ by $(c/eB)\partial/\partial x$ to get

$$\{f, g\}_\alpha = \frac{c}{eB}(f_y g_x - f_x g_y). \tag{2.117}$$

Thus we see that the original spatial coordinates x and y now play the role of canonically conjugate variables in the reduced space. The factor of $1/B$ in the bracket appeared in Littlejohn's work [Littlejohn, 1983]. The full system is not invariant under our circle action. If we average the perturbation Hamiltonian H_1 around the circles, we do obtain a circle symmetric system. The average of the potential ϵeEx around a loop is just the value when $p_y = 0$. Thus the reduced averaged Hamiltonian is

$$\bar{H}_\alpha(x, y) = \alpha - \epsilon eEx. \tag{2.118}$$

The reduced averaged dynamics is then

$$\begin{aligned} \dot{x} &= \{x, \bar{H}_\alpha\}_\alpha = 0 \\ \dot{y} &= \{y, \bar{H}_\alpha\}_\alpha = \frac{c}{eB}(-\epsilon eE) = -\epsilon \frac{cE}{B}. \end{aligned} \tag{2.119}$$

This is indeed the $E \times B$ drift dynamics.

Chapter 3: Pseudo-forces and Reduction

"Philosophy is written in this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth."—Galileo in *The Assayer*

3.1. Pseudo-forces and Reduction

If you have a system that is coupled to some subdynamics and you forget about the subdynamics, the original system may evolve with new "pseudo-forces" acting. For simple mechanical systems with symmetry, these take the form of "magnetic" terms in the Poisson bracket (by analogy with a Hamiltonian description of particles in magnetic fields) and new "effective potentials" in the Hamiltonian. These forces may stabilize previously unstable dynamics, just as a free charged particle at

rest in the plane is stabilized by a perpendicular magnetic field (*magnetic stabilization*). One finds these pseudo-forces also in situations where the observer's motion is included in the dynamics. In more complex examples these forces seem to lead to "convective" terms in Eulerian continua descriptions, drifts and pondermotive forces, "forces" that bend light rays in inhomogeneous media, pressure and other thermodynamic forces in statistical mechanics, "renormalized" masses for quasi-particles, etc. Near the end of the last century, there was a school of thought that held all potential energies to be merely the kinetic energy of "hidden" degrees of freedom. This gave great impetus to the kinetic theory of matter as we discussed in section 2.9.3.1.

3.1.1. Fictitious Forces

Our goal in this chapter is to give a coordinate-free interpretation to the process of changing reference frames and to the concommitant new physical effects. We first consider the effect of a time dependent change of phase space. A single state in the new frame corresponds to a time-parametrized curve of states in the old frame (e.g., a given configuration of particles and their velocities fixed in a rotating frame corresponds to a whole circle of states in a fixed frame as time evolves). We therefore have a time-dependent identification between the original phase space and the new one. The time dependence of the identification diffeomorphism may be expressed as the flow of a vector field Y (for example, a rotating reference frame is described by the vector field we discussed in section 2.6).

Let FX_t be the flow of X for time t, and FY_t the flow of Y for time t, on M. FX_t is viewed as the dynamics and FY_t as our "changing point of view", so the dynamical evolution we observe is $FY_t \circ FX_t$. This is the flow of the time-dependent vector field

$$\tilde{X} = FY_{t*} \cdot X + Y \tag{3.1}$$

by the chain rule (the lower star means push-forward by the map FY_t, and represents the image of X at each point under the differential of FY_t). If X is Y-invariant, then

$$\tilde{X} = X + Y. \tag{3.2}$$

If X is Hamiltonian X_H and Y is a symmetry generated by J, then the combined flow has Hamiltonian

$$\tilde{H} = H + J. \tag{3.3}$$

J represents fictitious forces in the Hamiltonian due to our changing perspective.

3.1.2. Rotating Coordinates

In this section we will demonstrate these ideas on the example of a rotating coordinate system for a particle in the plane. A very important subtlety arises from the question of what the velocity and momenta are in a rotating frame. One perspective takes some given inertial frame and always talks about velocity v in that frame with corresponding momentum mv. The other perspective measures velocities with respect to the observer's coordinate system. If we do this and still define momentum as mass times velocity, then the Poisson structure will change in

general as we change coordinates. We take this approach here and will identify the new piece in the Poisson bracket as the Coriolis force. In older times the notion of a non-canonical Poisson bracket was not in widespread use. To keep the bracket definition invariant, one had to say that momentum did not change under change of reference frame (see for example p. 129 of the mechanics volume of [Landau and Lifshitz, 1960–1981]).

Let

$$H = \frac{1}{2m}(p_x^2 + p_y^2) + V(\sqrt{x^2 + y^2}) \tag{3.4}$$

be a rotationally symmetric Hamiltonian on $T^*\Re^2$ and

$$J = \omega(xp_y - yp_x) \tag{3.5}$$

generate rotations. Since H is invariant under the symmetry here,

$$\tilde{H} = H + J. \tag{3.6}$$

This captures the observer's rotating reference, but does not include the fact that due to his motion the observer will measure a different set of values for the momenta. (In some sense, the transformation thus far has given us valid orbits of the observed system but has changed which orbit we are looking at). We must change coordinates by

$$\begin{aligned} \tilde{x} &= x, \\ \tilde{y} &= y \\ \tilde{p}_x &= p_x - \omega m y \\ \tilde{p}_y &= p_y + \omega m x \end{aligned} \tag{3.7}$$

to get the momenta as seen in the rotating frame. Here we define the momentum $\tilde{p}$ to be m times the observed velocity. The observed x-component of the velocity will be the velocity v_x measured in the fixed frame minus the x-component of the velocity of the observation point in the rotating frame, which is ωy. Similarly, the observed y-component of the velocity is v_y minus $-\omega x$. The dynamics is now described by

$$\tilde{H} = \frac{1}{2m}(\tilde{p}_x^2 + \tilde{p}_y^2) + V(\sqrt{x^2 + y^2}) - \frac{\omega^2 m}{2}(\tilde{x}^2 + \tilde{y}^2) \tag{3.8}$$

with the Poisson bracket

$$\{f, g\} = \{f, g\}_{\text{canonical}} + 2\omega m \left(\frac{\partial f}{\partial \tilde{p}_x} \frac{\partial g}{\partial \tilde{p}_y} - \frac{\partial f}{\partial \tilde{p}_y} \frac{\partial g}{\partial \tilde{p}_x} \right). \tag{3.9}$$

The centrifugal term

$$-\frac{\omega^2 m}{2}(\tilde{x}^2 + \tilde{y}^2) \tag{3.10}$$

makes an effective potential or pseudo-force and the Coriolis term in the Poisson bracket is of the "magnetic" type which causes drifts.

3.1.3. Reduction of Simple Mechanical Systems

We have seen how changing coordinates can lead to pseudo-potentials in the Hamiltonian and magnetic terms in the Poisson bracket. Asymptotic systems get pseudo-forces by reduction by an approximate symmetry. Here we give the context in which pseudo-forces and magnetic terms in the Poisson bracket may be seen to arise from the process of reduction. We specialize this to circle actions in the next section. In the following 3 sections we show how changing coordinates may be

viewed as reduction of a larger space which includes the observer's state. This unifies these two sources of pseudo-forces. The argument is of necessity a bit abstract and so some readers may wish to skip the details. Let us start with the general setting.

If the phase space is T^*Q, where Q is Riemannian with metric K and the Hamiltonian is of the form

$$H = K^* + \pi^* V, \tag{3.11}$$

where we moved K to T^*Q and

$$V : Q \to \Re \tag{3.12}$$

is a real valued function on Q which we lift to T^*Q along the natural projection

$$\pi : T^*Q \to Q, \tag{3.13}$$

then we have a *simple mechanical system* (see [Abraham and Marsden, 1978] p. 341). A group action of G on Q by isometries that preserve V lifts to T^*Q to preserve H. If we reduce at $\mu \in g^*$ with the momentum map

$$J : T^*Q \to g^*, \tag{3.14}$$

the reduced space is

$$(T^*Q)_\mu = J^{-1}(\mu)/G_\mu \tag{3.15}$$

where G_μ is the isotropy subgroup of μ under the coadjoint action. We may identify this reduced space witb the cotangent bundle $T^*(Q/G_\mu)$ with a new Poisson bracket (the old one plus "magnetic terms") and a new Hamiltonian (the old one with a new "effective potential"). Using the metric, we choose a one-form α_μ on Q to behave

on vectors tangent to the orbit of G_μ in Q the way μ behaves on the corresponding elements of the Lie algebra, and to annihilate vectors perpendicular to the orbit. α_μ induces $\hat{\alpha}_\mu$ on Q/G_μ and by lifting, on $T^*(Q/G_\mu)$. The new symplectic structure on $T^*(Q/G_\mu)$ is the old one plus the magnetic piece $d\hat{\alpha}_\mu$. The new Hamiltonian is the old one (H is G invariant and so is defined on $T^*(Q/G_\mu)$) plus a pseudo-potential $K^*(\alpha_\mu(q))$ (for more details see [Marsden, 1981] p. 33).

3.1.4. Circle Actions on Simple Mechanical Systems

With the notation above, if G is a circle and the vector field generating its action on Q is ξ, then there are some simplifications. The momentum map

$$J : T^*Q \to \Re \tag{3.16}$$

may be taken to be

$$\alpha_q \mapsto \alpha_q(\xi). \tag{3.17}$$

The G-action on T^*Q is then generated by X_J. α_μ is then simply

$$\frac{\mu}{K(\xi,\xi)}K(\cdot,\xi), \tag{3.18}$$

so the effective potential is

$$\frac{\mu^2}{K(\xi,\xi)}. \tag{3.19}$$

In this case the reduced space is identifiable with $T^*(Q/S^1)$ with its modified structures.

3.1.5. Fictitious Forces as Reduction

We now describe a setup which unifies some of these constructions. Assume we are given, as above, a Riemannian Q with metric K and symmetry generated by ξ which also leaves $V : Q \to \Re$ invariant. ξ lifts to X_J, which is generated by

$$J(q,p) = p(\xi(q)) \tag{3.20}$$

and leaves $H = K^* + V$ on T^*Q invariant. Above we reduced by X_J to get a lower dimensional system. The approach to fictitious forces given in section 3.1.1 treated the rotating phase space as being of the same dimension as the fixed phase space. Here we increase the dimension by 2. Now we introduce a "rotating observer", described by a point θ of a circle S^1 and a rotation action p_θ in T^*S^1. A circle acts on this S^1 generated by $\partial/\partial\theta$ and lifts to T^*S^1 generated by p_θ. Call

$$\hat{\xi} = \frac{\partial}{\partial\theta} + \xi \tag{3.21}$$

on $S^1 \times Q$ and lift it to $X_{\hat{J}}$ on $T^*(S^1 \times Q)$ generated by

$$\hat{J} = p_\theta + J. \tag{3.22}$$

ξ rotates the configuration space without changing θ (the state of the observer), $\partial/\partial\theta$ rotates the observer without changing configuration space, $\hat{\xi}$ rotates the two together (twisting together the two circle actions), X_J rotates particle phase space alone, X_{p_θ} rotates the observer alone, and $X_{\hat{J}}$ rotates the particle phase space and the observer.

We have the Hamiltonian

$$\hat{H} = H + \frac{1}{2}p_\theta^2 \tag{3.23}$$

on $T^*(S^1 \times Q)$ which generates the real dynamics on T^*Q and rotates the observer with speed p_θ. It Poisson commutes with p_θ, J (thought of on $T^*(S^1 \times Q)$) and $\hat{J}$. We want to consider the orbit space of the X_J action, but to look at the level surfaces of p_θ (as opposed to $\hat{J}$) since we want to study a given speed of rotation and we don't want to mix up different speeds in the same phase space. We may identify

$$T^*(S^1 \times Q)|_{p_\theta = \text{constant}}/X_J \tag{3.24}$$

with T^*Q by identifying (θ, q, p_θ, p) to the $(\tilde{q}, \tilde{p})$ on the X_J orbit through it at the point $\theta = 0$. Since

$$X_H = X_H + p_\theta \frac{\partial}{\partial \theta} \tag{3.25}$$

and

$$X_J = X_J + \frac{\partial}{\partial \theta}, \tag{3.26}$$

the dynamics on $T^*(S^1 \times Q)/X_J$ identified in the above way is given by

$$X_H - p_\theta X_J. \tag{3.27}$$

The Poisson bracket of two functions is seen to be

$$\left\{f, g\right\}_{T^*(S^1 \times Q)/X_J} = \left\{f, g\right\}_{T^*Q} - X_J \cdot f \frac{\partial g}{\partial p_\theta} + \frac{\partial f}{\partial p_\theta} X_J \cdot g. \tag{3.28}$$

If we now restrict to p_θ =constant, and identify with T^*Q, we get a Hamiltonian system with the original Poisson bracket and Hamiltonian

$$H - p_\theta J, \tag{3.29}$$

showing $-p_\theta j$ to be a fictitious force.

3.1.6. Fictitious Forces with Momentum Shift as Reduction

Now we apply the ideas from section 3.1.4 to the setup in section 3.1.5. The orbit space of $\hat{\xi}$ on $S^1 \times Q$ is identifiable with Q by sending an orbit to its q value at $\theta = 0$, call it $(S^1 \times Q)/\hat{\xi}$. We would like to compare the T^*Q obtained in section 3.1.5 by quotienting $T^*(S^1 \times Q)$ by X_J and holding p_θ fixed, with the T^*Q obtained by taking the cotangent bundle

$$T^*[(S^1 \times Q)/\hat{\xi}] \tag{3.30}$$

and so obtain its relation to the original T^*Q dynamics. We use the fact that on the cotangent bundle

$$T^*[S^1 \times Q)/\hat{\xi}] \approx T^*Q \tag{3.31}$$

with metric K on Q, the momentum $\tilde{p}$ should be related to the velocity via K:

$$\tilde{p} = 2K(\cdot, \dot{q}). \tag{3.32}$$

Here $\dot{q}$ is the projection of the dynamics on T^*Q down to Q. This gives us a map

$$T^*Q \to T^*Q, \quad \text{by} \quad (q,p) \to (q,\tilde{p}) \tag{3.33}$$

where (q,p) is a representative of a point in

$$T^*(S^1 \times Q)/X_J\big|_{p_\theta = \text{constant}}$$

and $(q,\tilde{p})$ in

$$T^*[(S^1 \times Q)/\hat{\xi}].$$

Since the Hamiltonian is now $H - p_\theta J$, we see that

$$\begin{aligned}\dot{q} &= \frac{\partial}{\partial p}(H - p_\theta H) \\ &= \frac{\partial}{\partial p}(K^*(p,p) + V(q) - p_\theta p(\xi(q))) \\ &= \frac{\partial}{\partial p}K^*(p,p) - p_\theta \xi(q).\end{aligned} \tag{3.34}$$

K is quadratic, so

$$2K(\cdot, \frac{\partial}{\partial p}K^*(p,p)) = p \tag{3.35}$$

and so

$$\begin{aligned}\bar{\partial} &= 2K(\cdot, \frac{\partial H}{\partial p}) \\ &= p - p_\theta 2K(\cdot, \xi).\end{aligned} \tag{3.36}$$

In the case $Q = \Re^2$,

$$\xi = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \tag{3.37}$$

is a rotation, and

$$K = \frac{m}{2}(dx^2 + dy^2) \tag{3.38}$$

on TQ and

$$K^* = \frac{1}{2m}(p_x^2 + p_y^2) \tag{3.39}$$

on T^*Q, we see that

$$\tilde{p}_x = p_x + p_\theta m y \tag{3.40}$$

and

$$\tilde{p}_y = p_y - p_\theta m x. \tag{3.41}$$

If we identify

$$\omega = -p_\theta \tag{3.42}$$

then this is exactly the setup in section 3.1.2. Thus the Coriolis force and centrifugal force obtained in 3.1.2 are identified with magnetic terms in the Poisson bracket and a modified potential in the Hamiltonian arising from reduction.

3.2. Geometry of the Method of Variation of Parameters

Let us now consider the powerful and commonly used perturbation technique known as the method of variation of parameters. A good description of the technique may be found in [Nayfeh, 1973] on p. 59. This method defines the evolution of the perturbation in terms of quantities at the unperturbed point. We expect to be $O(1)$ away from this point in time $1/\epsilon$ (since the size of the perturbation is $O(\epsilon)$) and so cannot get a correct description on this time scale. To extend the time of validity, the orbit of the unperturbed system to which we compare the true flow must be allowed to vary. If we label unperturbed orbits by their initial conditions, then we may rewrite the perturbed orbit's drift from unperturbed orbit to orbit in terms of a drift of initial conditions. Often this slow drift may be further simplified (eg. via averaging) and results in a usable perturbation theory for time $1/\epsilon$. Let us formulate this dynamics geometrically.

Any dynamical system

$$\dot{x} = X(x) \tag{3.43}$$

on M gives rise to a natural dynamics on the group $Diff(M)$ of diffeomorphisms of M. We view X as an element of the Lie Algebra $diff(M)$ and we get the right-invariant dynamical vector field by right translation to each point of $Diff(M)$. In general, we get a *right*-invariant vector field on a group containing the vector v at the identity, by considering the first order infinitesimal *left* translation by the 1-parameter subgroup to which v is tangent. The resulting flow on $Diff(M)$ has the following interpretation: The time-t evolution of $f \in Diff(M)$ is the diffeomorphism obtained by first applying f and then letting X flow for time-t.

Now consider a perturbed vector field

$$X_0 + \epsilon X_1 \tag{3.44}$$

on M. Let $\mathcal{F}_t$ represent the time t flow of the unperturbed dynamics X_0. If $x(t)$ is to be the "evolution of initial conditions" that variation of parameters produces, then $\mathcal{F}_t(x(t))$ is a solution to the full equation

$$\frac{d}{dt}\mathcal{F}_t(x(t)) = X_0(\mathcal{F}_t(x(t))) + \epsilon X_1(\mathcal{F}_t(x(t))). \tag{3.45}$$

By definition, the left hand side is

$$= X_0(\mathcal{F}_t(x(t))) + D\mathcal{F}_t \cdot (\dot{x}(t)). \tag{3.46}$$

Thus

$$\dot{x}(t) = \epsilon D\mathcal{F}_{-t}(x(t)) \cdot X_1(\mathcal{F}_t x(t)) = \epsilon \mathcal{F}_{-t*} X_1. \tag{3.47}$$

We may view this as a time dependent evolution equation or consider dynamics on $Diff(M) \times M$ given by

$$\tilde{X} : (f, x) \mapsto (X_0 \circ f, \epsilon f_*^{-1} X_1) \tag{3.48}$$

with initial condition $(identity, x_0 + \epsilon x_1)$. If $\mathcal{F}_t$ is periodic then $\epsilon \mathcal{F}_{-t*} X_1$ is periodic and the method of averaging is to average it over a period. We discuss the method of averaging in greater detail in section 2.9.

3.2.1. Hamiltonian Variation of Parameters

If M, ω is symplectic and

$$X_{H_0} + \epsilon X_{H_1} \tag{3.49}$$

is Hamiltonian, then we may work on the group $Symp(M)$ of symplectomorphisms. If $f \in Symp(M)$ and X_{H_0} is Hamiltonian, then so is $X_{H_0} \circ f$. We may define a right invariant symplectic structure on $Symp(M)$ by

$$\tilde{\omega}_f(X_1 \circ f, X_2 \circ f) = \int_M \omega(X_1, X_2) dx. \tag{3.50}$$

We would like to find a Hamiltonian on $Symp(M)$ whose corresponding vector field at $f \in Symp(M)$ is $X_{H_0} \circ f$. We first determine the one-form obtained by inserting this into $\tilde{\omega}$:

$$\begin{aligned} \tilde{\omega}_f(X_{H_0} \circ f, X \circ f) &= \int_M \omega(X_{H_0}, X) dx \\ &= \int_M dH_0(X) dx \\ &= \int_M X(H_0) dx. \end{aligned} \tag{3.51}$$

Thus we may take the Hamiltonian on $Symp(M)$ to be the right invariant function

$$\tilde{H}_0(f) = \int H \circ f \, dx. \tag{3.52}$$

To check that this gives the correct dynamics, notice that the vector $X \circ f$ acting on this is

$$\begin{aligned} \frac{d}{d\epsilon} \tilde{H}(f + \epsilon X) &= \int \frac{d}{d\epsilon} H(f + \epsilon X) dx \\ &= \int X(H_0) dx \end{aligned} \tag{3.53}$$

as desired. Now notice that

$$\epsilon f_*^{-1} X_{H_1} = \epsilon X_{f^* H_1}. \tag{3.54}$$

So with a symplectic structure on $Symp(M) \times M$ given by

$$\tilde{\omega} \times \omega, \tag{3.55}$$

and a Hamiltonian $\tilde{H} : Symp(M) \times M \to \Re$ given by

$$\tilde{H} : (f, x) \mapsto \int_M H_0 \circ f \, dx + \epsilon H_1(f(x)), \tag{3.56}$$

we get the variation of parameter dynamics.

We may see from this that if X_{H_0} has all periodic orbits then the averaged dynamics is Hamiltonian with Hamiltonian $H_1(f(x))$ averaged over the evolution f.

Chapter 4: Hamiltonian Structures in Perturbation Theory

"Schrödinger and I both had a very strong appreciation of mathematical beauty and this dominated all our work. It was a sort of act of faith with us that any equations which describe fundamental laws of Nature must have great mathematical beauty in them. It was a very profitable religion to hold and can be considered the basis of much of our success."—P. A. M. Dirac on p. 136 of [Dirac, 1977]

4.1. Introduction

In this chapter we describe the geometry of a Hamiltonian structure for non-singular perturbation theory applied to Hamiltonian systems on symplectic manifolds. This is limited in two respects: 1) Many systems of physical interest require more sophisticated singular perturbation methods as in [Nayfeh, 1973] and [Kevorkian and Cole, 1981] and 2) the Hamiltonian structures of many systems are given in terms of more general Poisson manifolds [Weinstein, 1983a]. Some extension of the present chapter to these cases is given in later chapters. The results

in this chapter are relevant to these investigations though. Most singular perturbation methods have a nonsingular expansion underlying them. Poisson manifolds are stratified by symplectic manifolds and many of the symplectic constructions considered here are susceptible to generalization.

If we are given dynamics in the form

$$\dot{x} = X_0 + \epsilon X_1 + \frac{\epsilon^2}{2} X_2 + \cdots \tag{4.1}$$

where each of the vector fields X_i is Hamiltonian with respect to a common Poisson structure, we may attempt to express the solution as an asymptotic series in ϵ:

$$x(t) = x_0(t) + \epsilon x_1(t) + \frac{\epsilon^2}{2} x_2(t) + \cdots. \tag{4.2}$$

Plugging this form into the equation of motion and equating coefficients of powers of ϵ gives us equations for $x_0, x_1, \ldots$. The main result of this chapter is that the equations for $x_0, \ldots, x_J$ form a Hamiltonian system for any J.

These results were discussed in chapter 2 but are proved here in full detail. The background material and non-Hamiltonian perturbation structures introduced in chapter 2 will be needed in this chapter. The reader not familiar with geometric mechanics at the level of [Arnold, 1978] may find sections of this chapter rough going. Except for parts of chapter 5, the rest of the thesis is independent of the detailed derivations given here.

We study the perturbation Hamiltonian structure in five different ways, each of which sheds a different light on it. In the second section we do the case of first order perturbation theory explicitly, where it is easier to understand the structure. The extension to arbitrary order necessitates the introduction in the third section

of certain path spaces and jet bundles. The five investigative approaches then follow. 1) In the fourth section we give the desired Hamiltonian structure in local canonical coordinates. 2) The fifth section shows that this structure is coordinate independent by imbedding the jet bundle in an iterated tangent bundle. 3) The sixth section obtains the structure from a natural one on the infinite dimensional path space. 4) The seventh section shows in what sense the structure is the Jth derivative of a product structure. 5) The eighth section shows that if the original symplectic manifold is a coadjoint orbit in the dual of a Lie algebra with the Kirillov-Kostant-Souriou (KKS) Lie symplectic structure, theo the jet bracket is a KKS Lie symplectic structure for a coadjoint orbit of a certain jet group. We close with a discussion of the process of reduction applied to perturbed systems with symmetry.

4.2. First Order Hamiltonian Perturbation Theory

Let us recall the geometric structure of Hamiltonian mechanics. The phase space will be a $2N$-dimensional symplectic manifold M (many of the constructions work in infinite dimensions and many of the interesting physical examples are infinite dimensional [Marsden, 1981], but this entails technicalities which we will not consider here). This is a manifold with a distinguished closed non-degenerate 2-form ω which geometrizes the classical Lagrange bracket. The Hamiltonian is a distinguished function H on M which we take to be a function of ϵ as well. Usually we assume that $\epsilon \in I \equiv [0, 1]$ and so we can view H as a function on $I \times M$. For each value of ϵ we obtain a vector field $X(\epsilon)$ on M by the Hamiltonian prescription. This says that at a point $x_0 \in M$, $X(\epsilon, x_0)$ is the unique (since ω is non-degenerate) vector at x_0 which gives the one form $dH|_{x_0,\epsilon}$ when inserted in $\omega|_{x_0}$, i.e.

$$i_{X(\epsilon)}\omega = dH. \tag{4.3}$$

We will discuss the flow of $X(\epsilon)$ as though it were complete, though this need not be the case (a vector field is complete if solution curves don't run off the manifold in finite time). Let us assume that H may be represented in an asymptotic series as $\epsilon \to 0$ which is uniform in x, i.e.

$$H(\epsilon, x) \sim H_0(x) + \epsilon H_1(x) + \frac{\epsilon^2}{2!} H_2(x) + \cdots \tag{4.4}$$

where the first derivatives of each H_i are uniformly bounded in x. Because of the linearity in going from functions to their Hamiltonian vector fields, we also have the asymptotic expansion

$$X(\epsilon, x) \sim X_0(x) + \epsilon X_1(x) + \frac{\epsilon^2}{2!} X_2(x) + \cdots \tag{4.5}$$

as $\epsilon \to 0$. In this expression, X_ϵ represents the Hamiltonian vector field corresponding to H_ϵ. Hamilton's equations are

$$\dot{x}(\epsilon, t) = X\big(\epsilon, x(\epsilon, t)\big). \tag{4.6}$$

Let the initial conditions be given as $x(\epsilon, t = 0) = y(\epsilon)$ and assume that y has an expansion $y(\epsilon) \sim y_0 + \epsilon y_1 + (\epsilon^2/2!)y_2 + \cdots$.

As we have seen in chapter 2, non-singular perturbation theory asks for the flow representing the solution as an asymptotic series in ϵ:

$$x(\epsilon, t) \sim x_0(t) + \epsilon x_1(t) + \frac{\epsilon^2}{2!} x_2(t) + \ldots \qquad \text{as } \epsilon \to 0. \tag{4.7}$$

In chapter 2 we substituted this representation into the equations of motion, equated coefficients of equal powers of ϵ, and so obtained differential equations for $x_0, x_1, \ldots$ with initial conditions given by $x_i(t = 0) = y_i$. The solution of these equations gave us an asymptotic representation of the true solution, but in general it was non-uniform in t. To deal with times (like $\frac{1}{\epsilon}$) longer than some bounded value as $\epsilon \to 0$, we must use more sophisticated perturbation techniques such as Lie transforms or multiple time scales and so lose the generality of the problems we may treat. We discuss these *singular* or *secular* perturbation theories in chapter 5.

The goal in this chapter is to determine the geometric nature of the quantities $x_0, x_1, \ldots$ and to determine a Hamiltonian structure for their evolution equations. It is easy to relate these asymptotic expansions for the abstract vector fields and their Hamiltonians to a coordinate representation of them. If we introduce a local coordinate system $x^1, \ldots, x^{2N}$ on a chart of M, we may express

$$X(\epsilon, x) = \sum_{a=1}^{2N} X^a(\epsilon, x) \frac{\partial}{\partial x^a} \tag{4.8}$$

and

$$X_i(x) = \sum_{a=1}^{2N} X_i^a(x) \frac{\partial}{\partial x^a} \qquad \text{for} \quad 0 \leq i < \infty. \tag{4.9}$$

We will use a convention where the upper index from the beginning of the Latin alphabet (e.g. a,b,c,...) gives the coordinate on M and the lower index from the middle of the Latin alphabet (e.g. i,j,k,...) gives the order of perturbation. The components of X in this coordinate system have the asymptotic expansion

$$X^a(\epsilon, x) \sim X_0^a(x) + \epsilon X_1^a(x) + \cdots \qquad \text{as} \quad \epsilon \to 0 \quad \text{for} \quad 1 \leq a \leq 2N \tag{4.10}$$

as may easily be seen. The equations of motion are

$$\dot{x}^a = X^a(\epsilon, x) \quad \text{with} \quad x^a(t=0) = y^a \quad \text{for} \quad 1 \leq a \leq 2N. \tag{4.11}$$

We saw in chapter 2 that the first order perturbation approximation to these equations is

$$\begin{aligned} \dot{x}_0^a &= X_0^a(x_0) \\ \dot{x}_1^a &= \sum_{b=1}^{2N} \frac{\partial X_0^a}{\partial x^b}(x_0) \cdot x_1^b + X_1^a(x_0) \end{aligned} \tag{4.12}$$

with initial conditions $x_0^a(t=0) = y_0^a$ and $x_1^a(t=0) = y_1^a$. A solution $x_0(t), x_1(t)$ to these equations with the correct initial conditions will form a first order asymptotic solution $x_0(t) + \epsilon x_1(t)$ to the original equation at each fixed t (and so uniformly over bounded time intervals).

We have seen how to formulate this procedure in terms of coordinate free objects. The true dynamics takes place on M and for each t the flow $x(t)$ is a diffeomorphism of M to itself taking initial conditions to their time t evolution. X_0 and X_1 are vector fields on M. $x_0(t)$ gives the flow of the unperturbed vector

field X_0. Because x_1 represents a small perturbation to x_0 as ϵ vanishes, it lives in the tangent space to M at x_0. x_0, x_1 represent an equivalence class of flows parameterized by ϵ, where we identify flows which asymptote to the unperturbed flow with linear rate x_1 as ϵ goes to zero. If we pick a time t and an initial condition then each flow defines a curve parameterized by ϵ, which passes through $x_0(t)$ when $\epsilon = 0$. The equivalence relation on flows leads to an equivalence relation on curves through $x_0(t)$ that is exactly the defining relation for a tangent vector based at $x_0(t)$. In local coordinates on M, we see that x_0^a, x_1^b for $1 \leq a, b \leq 2N$ are coordinates on the tangent bundle TM where x_0 coordinatizes the base and x_1 the fiber over x_0.

Equations (4.12) are to be thought of as locally defining a vector field on TM. We showed in chapter 2 that they are actually coordinate independent by defining vector fields $\tilde{X}_0$ and $\tilde{X}_1$ on TM from X_0 and X_1 on M. If $x_0(t)$ is the flow of X_0 on M, then its derivative $Tx_0(t)$ defines a flow on TM. We defined the corresponding vector field on TM:

$$\tilde{X}_0 \equiv \left.\frac{d}{dt}\right|_{t=0} Tx_0(t). \tag{4.13}$$

$\tilde{X}_0$ defines the linearized flow of X_0 (see [Abraham and Marsden, 1978], page 252). In coordinates, $\tilde{X}_0$ is given by

$$\begin{aligned} \dot{x}_0^a &= X_0^a(x_0) \\ \dot{x}_1^a &= \sum_{b=1}^{2N} \frac{\partial X_0^a}{\partial x^b}(x_0) \cdot x_1^b \qquad \text{for} \quad 1 \leq a \leq 2N. \end{aligned} \tag{4.14}$$

Because the tangent space at a point is linear, it makes sense to add a vector v in TM to a tangent vector in TTM over v. Thus for $(x, v) \in TM$ we may define

$$\tilde{X}_1(x, v) = \left.\frac{d}{dt}\right|_{t=0} (v + tX_1(x)). \tag{4.15}$$

This gives a vertical vector field on the bundle TM which is constant on each fiber. In coordinates $\tilde{X}_1$ is given by

$$\begin{aligned} \dot{x}_0^a &= 0 \\ \dot{x}_1^a &= X_1^a(x_0) \end{aligned} \tag{4.16}$$

Thus the invariantly defined vector field $\tilde{X}_0 + \tilde{X}_1$ on TM gives the correct equations (4.12) in each local coordinate patch.

Let us now assume that X_0 and X_1 are Hamiltonian and investigate the Hamiltonian nature of the perturbed system $\tilde{X}_0 + \tilde{X}_1$ on TM. M carries the symplectic two-form ω and we are given Hamiltonians H_0 and H_1 such that

$$\begin{aligned} i_{X_0}\omega &= dH_0 \\ i_{X_1}\omega &= dH_1. \end{aligned} \tag{4.17}$$

TM carries a natural symplectic structure $\tilde{\omega}$ gotten by using ω to identify TM with T^*M and by pulling back T^*M's natural structure to TM ([Abraham and Marsden, 1978], page 200, problem 3.31). Because the unperturbed Hamiltonian is a map from M to the reals,

$$H_0 : M \to \Re, \tag{4.18}$$

we see that its differential,

$$dH_0 : TM \to \Re, \tag{4.19}$$

may be thought of as a function on TM which is linear on the fibers. If we denote the natural projection of TM to M by π then since H_1 is a function on M, we see that π^*H_1 is a function on TM. We shall see that $\tilde{X}_0$ and $\tilde{X}_1$ are Hamiltonian with respect to $\tilde{\omega}$ with Hamiltonians given by dH_0 and π^*H_1. Thus

$$\tilde{H} \equiv dH_0 + \pi^*H_1 \tag{4.20}$$

is the desired Hamiltonian on TM for the perturbed equations (4.12).

We demonstrate these statements in local Darboux (canonical) coordinates q^α, p^α for $1 \le \alpha \le N$ on M. The symplectic structure is

$$\omega = \sum_{\alpha=1}^{N} dq^\alpha \wedge dp^\alpha \tag{4.21}$$

with corresponding Poisson bracket:

$$\{q^\alpha, p^\beta\} = \delta^{\alpha\beta} \tag{4.22}$$

and all other combinations of q's and p's vanishing. We represent x_0 by q_0^α, p_0^α and x_1 by q_1^α, p_1^α. One can see that the natural Poisson bracket on TM corresponding to to is

$$\{q_0^\alpha, p_1^\beta\} = \delta^{\alpha\beta} \qquad \{q_1^\alpha, p_0^\beta\} = \delta^{\alpha\beta} \qquad \text{for} \quad 1 \le \alpha, \beta \le N \tag{4.23}$$

where all other combinations vanish. In these coordinates

$$\tilde{H}(q_0, p_0, q_1, p_1) = \sum_{\alpha=1}^{N} \left(\frac{\partial H_0}{\partial q_0^\alpha} q_1^\alpha + \frac{\partial H_0}{\partial p_0^\alpha} p_1^\alpha \right) + H_1(q_0, p_0). \tag{4.24}$$

The corresponding equations of evolution are

$$\begin{cases} \dot{q}_0^\alpha = \{q_0^\alpha, \tilde{H}\} = \dfrac{\partial \tilde{H}}{\partial p_1^\alpha} = \dfrac{\partial H_0}{\partial p_0^\alpha} \\ \dot{p}_0^\alpha = \{p_0^\alpha, \tilde{H}\} = -\dfrac{\partial \tilde{H}}{\partial q_1^\alpha} = -\dfrac{\partial H_0}{\partial q_0^\alpha} \\ \dot{q}_1^\alpha = \{q_1^\alpha, \tilde{H}\} = \dfrac{\partial \tilde{H}}{\partial p_0^\alpha} = \displaystyle\sum_{\beta=1}^{N} \left(q_1^\beta \frac{\partial}{\partial q_0^\beta} + p_1^\beta \frac{\partial}{\partial p_0^\beta} \right) \frac{\partial H_0}{\partial p_0^\alpha} + \frac{\partial H_1}{\partial p_0^\alpha} \\ \dot{p}_1^\alpha = \{p_1^\alpha, \tilde{H}\} = -\dfrac{\partial \tilde{H}}{\partial q_0^\alpha} = \displaystyle\sum_{\beta=1}^{N} \left(q_1^\beta \frac{\partial}{\partial q_0^\beta} + p_1^\beta \frac{\partial}{\partial p_0^\beta} \right) \left(-\frac{\partial H_0}{\partial q_0^\alpha} \right) - \frac{\partial H_1}{\partial q_0^\alpha}. \end{cases} \tag{4.25}$$

But X_0 gives the equations

$$\dot{q}_0^a = \frac{\partial H_0}{\partial p_0^a} \qquad \dot{p}_0^a = -\frac{\partial H_0}{\partial q_0^a}. \tag{4.26}$$

so we see that

$$\frac{\partial X_0}{\partial x_0} \cdot x_1 = \begin{pmatrix} \frac{\partial}{\partial q_0}\frac{\partial H_0}{\partial p_0} & \frac{\partial}{\partial p_0}\frac{\partial H_0}{\partial p_0} \\ \frac{\partial}{\partial q_0}\left(-\frac{\partial H_0}{\partial q_0}\right) & \frac{\partial}{\partial p_0}\left(-\frac{\partial H_0}{\partial p_0}\right) \end{pmatrix} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix}. \tag{4.27}$$

X_1 gives the equations

$$\dot{q}_0^a = \frac{\partial H_1}{\partial p_0^a} \qquad \dot{p}_0^a = -\frac{\partial H_1}{\partial p_0^a} \tag{4.28}$$

and $\bar{H}$ has indeed given us the desired perturbation equations (4.12).

4.3. Path and Jet Spaces

We would now like to extend the first order results of the last section to arbitrary order J. We will extend the Hamiltonian structure from the tangent bundle to the jet bundles introduced in Chapter 2.

As in that chapter, we introduce the path space:

$$PM \equiv \left\{ \text{space of all paths } p : I \to I \times M \text{ of the form } p : \epsilon \mapsto (\epsilon, x(\epsilon)) \right\} \tag{4.29}$$

and from this we define the jet spaces with integer $1 \leq J \leq \infty$:

$$\begin{aligned} JM \equiv \Big\{ & \text{equivalence classes in } P_1 M \text{ where } p_1 \sim p_2 \text{ iff} \\ & \forall\, C^\infty \text{ functions } f \text{ on } I \times M \text{ we have :} \\ & \left.\frac{\partial^i}{\partial \epsilon^i}\right|_{\epsilon=0} f(p_1(\epsilon)) = \left.\frac{\partial^i}{\partial \epsilon^i}\right|_{\epsilon=0} f(p_2(\epsilon)) \qquad \text{for } 0 \leq i \leq J \Big\}. \end{aligned} \tag{4.30}$$

If x^a for $1 \leq a \leq 2N$ are coordinates on $M \approx P_0 M \approx 0M$, then we introduce coordinates $\{x_0^a, x_1^a, \ldots, x_J^a\}$ for $0 \leq J \leq \infty$ on JM to represent the equivalence class of the curve

$$x_0^a + \epsilon x_1^a + \frac{\epsilon^2}{2!} x_2^a + \ldots + \frac{\epsilon^J}{J!} x_J^a \tag{4.31}$$

in $I \times M$ (near $\epsilon = 0$ this won't leave the chart on which the x^a are defined).

In chapter 2 we identified the tangent spaces to these and showed how ϵ-dependent dynamics on M induces dynamics on these. The induced dynamics on the jet space JM was exactly the perturbation dynamics up to Jth order. We will now assume that the dynamics on M is Hamiltonian and try to find Hamiltonian structures for the dynamics on PM and JM.

4.3.1. Path Space Symplectic Structure and Hamiltonian

The Hamiltonian structure on M lifts to one on P_1M. As before, ω is a symplectic form on M and $H(\epsilon, x)$ is a Hamiltonian. There is a natural symplectic form $\tilde{\omega}$ on P_1M. Intuitively, if we think of P_1M as a continuous product of M's corresponding to each value of ϵ, then $\tilde{\omega}$ will just be the continuous sum of the corresponding symplectic structures. At a point $p \in P_1M$ and with vectors $\tilde{V}_1, \tilde{V}_2 \in T_pP_1M$, we define

$$\tilde{\omega}_p(\tilde{V}_1, \tilde{V}_2) \equiv \int_0^1 \omega_{p(\epsilon)}(V_1(\epsilon, p(\epsilon)), V_2(\epsilon, p(\epsilon)))\, d\epsilon. \tag{4.32}$$

Similarly, we expect the Hamiltonian to be a continuous sum of the Hamiltonians for each ϵ. We define $\tilde{H}$ on P_1M as

$$\tilde{H}(p) \equiv \int_0^1 H(\epsilon, p(\epsilon))\, d\epsilon. \tag{4.33}$$

4.3.2. The Path Space Dynamics is Hamiltonian

We will now show that the Hamiltonian vector field on P_1M defined by $\tilde{H}$ and $\tilde{\omega}$ is exactly the lift $\tilde{X}_H$ of the Hamiltonian vector field X_H on M. In finite dimensions, the differential of a function pairs with a vector by taking the sum over components of the product of each component of the vector with the derivative of the function in the corresponding direction. When we consider functions on a path space, this sum turns into an integral. The differential of $\tilde{H}$ thus satisfies

$$d\tilde{H}(\tilde{V}_p) = \int_0^1 V(\epsilon, p(\epsilon)) \cdot H(\epsilon)\, d\epsilon. \tag{4.34}$$

$V\cdot$ is the directional derivative along V, with $H(\epsilon)$ viewed as a function on M and $V(\epsilon, p(\epsilon))$ viewed as a vector in TM. Let us see what $\iota_{\tilde{X}_H}\tilde{\omega}$ is:

$$\begin{aligned}\tilde{\omega}_p(\tilde{X}_H, \tilde{V}) &= \int_0^1 \omega(X_{H(\epsilon)}(p(\epsilon)), V(\epsilon, p(\epsilon)))\, d\epsilon \\ &= \int_0^1 V(\epsilon, p(\epsilon)) \cdot H(\epsilon)\, d\epsilon \\ &= d\tilde{H}(\tilde{V}).\end{aligned} \tag{4.35}$$

Thus $\tilde{X}_H$ is indeed Hamiltonian on P_1M.

In chapter 2 we saw that the dynamics on P_1M naturally projects down to the desired perturbation dynamics on JM. We would like to project the Hamiltonian and symplectic structure as well to make JM's dynamics Hamiltonian. Unfortunately, functions and forms may only be pulled back functorially and cannot be naturally pushed forward. We may write down the Poisson bracket on P_1M corresponding to $\tilde{\omega}$. Poisson brackets can sometimes be pushed forward along a projection by pulling back the bracketed functions. In this case, however, things become too singular and we would be left with "products" of delta functions. In the next section we will find a Hamiltonian structure on JM and in later sections we will relate it to the structure on P_1M.

4.4. Coordinate Description of the J-jet Structure

Let us reiterate the fundamental problem. We have defined a space of J-jets JM with coordinates $\{x_0^a, \ldots, x_J^a\}$, $1 \leq a \leq 2N$. The correct perturbation dynamics is given by the vector field with components

$$V_k^a(x_0, \ldots, x_J) = \left.\frac{\partial^k}{\partial \epsilon^k}\right|_{\epsilon=0} X^a\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right), \qquad (4.36)$$
$$\text{for} \quad 1 \leq a \leq 2N \quad \text{and} \quad 0 \leq k \leq J.$$

We want to know whether this dynamics on a $2N(J+1)$ dimensional space is Hamiltonian if $X(\epsilon)$ is Hamiltonian on each $2N$ dimensional space ϵ =constant.

Darboux's theorem [Arnold, 1978] tells us that we may choose the coordinates $\{x^1, \ldots, x^{2N}\}$ on M to be canonical. Thus the Poisson bracket of any two coordinate functions,

$$\{x^a, x^b\} = J^{ab} \qquad 1 \leq a, b \leq 2N, \qquad (4.37)$$

is a constant independent of x. The dynamics then takes the form

$$X^a = \{x^a, H\} = J^{ab} \frac{\partial}{\partial x^b} H. \qquad (4.38)$$

The correct perturbation dynamics is then

$$\dot{x}_k^a = J^{ab} \frac{\partial}{\partial x^b} \left(\left.\frac{d^k}{d\epsilon^k}\right|_{\epsilon=0} H\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right)\right) \qquad (4.39)$$
$$\text{for } 1 \leq a \leq 2N \quad \text{and} \quad 0 \leq k \leq J$$

from the expression in 4.36 for the perturbation vector field.

4.4.1. The Jet Hamiltonian

This dynamics is Hamiltonian with the Hamiltonian function on JM given by

$$\bar{H}(x_0,\ldots,x_J) \equiv \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} H\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) \tag{4.40}$$

with respect to the Poisson bracket introduced in the next section. All of the various derivatives in 4.39 are contained in this expression and the Poisson bracket picks out the right one for each perturbation variable.

4.4.2. The Jet Poisson Bracket

To discover the Poisson structure, we calculate

$$\begin{aligned}\frac{\partial \bar{H}}{\partial x_k^a} &= \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} \frac{\partial}{\partial x_k^a} H\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) \\ &= \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} \left(\frac{\epsilon^k}{k!}\frac{\partial H}{\partial x^a}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right)\right) \\ &= \frac{J!}{k!(J-k)!}\left.\frac{d^{J-k}}{d\epsilon^{J-k}}\right|_{\epsilon=0} \frac{\partial H}{\partial x^a}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right).\end{aligned} \tag{4.41}$$

Thus

$$\dot{x}_k^a = \{x_k^a, \bar{H}\}_{JM} \tag{4.42}$$

gives the correct dynamics if the jet Poisson bracket is

$$\begin{gathered}\{x_k^a, x_m^b\} = J^{ab}\frac{k!m!}{J!}\delta_{k,J-m} \\ \text{for} \quad 0 \le k, m \le J \quad \text{and} \quad 1 \le a, b \le 2N.\end{gathered} \tag{4.43}$$

Notice that for $J = 1$ this gives $\{x_0^a, x_1^b\} = J^{ab}$, which was the bracket that we found in section 4.2 for first order perturbation theory.

4.5. Relation to the Iterated Tangent Bundle

We now need to show that this construction, defined in terms of coordinates, is really intrinsic. $\tilde{H}$ is clearly intrinsic, being the Jth derivative of H along any representative curve in P_1M of the point in JM (all such curves give the same answer by definition). That the structure of the Poisson bracket is intrinsic is not so obvious, but may be seen as follows.

4.5.1. Injecting Jets into the Iterated Tangent Bundle

Recall that we are letting I stand for the interval $[0,1]$. If we take the Jth derivative of the map $I \to M$, we get a map of the iterated tangent bundles:

$$T^J I \to T^J M. \tag{4.44}$$

($T^J M$ simply means $T(T(...(TM)...))$ where there are J T's. Each time we take the derivative of a map we get a map between the tangent bundles of the two manifolds.) We may think of this as a curve in $T^J M$, since a curve in M lifts to its tangent vector at each point in TM, this curve lifts to one in TTM, etc. The point $\epsilon = 0$ of this curve in $T^J M$ is then the image of $\partial^J/\partial\epsilon^J$. As we look at all curves in M, we don't get all points in the $2^J \cdot 2N$ dimensional $T^J M$, but rather only a $2N(J+1)$ dimensional submanifold identifiable with the jet space. This submanifold is made up of certain diagonals in the iterated tangent bundle which arise because the derivative of the derivative along a path is the same as the second derivative along a path. We give the details in the next section.

4.5.1.1. Coordinate Description of the Injection

If we look in coordinates, we see that this submanifold is given by certain diagonals in the iterated tangent bundle:

$$(x^a) = (x_0^a) \in M \tag{4.45}$$

$$\left(x^a, \left.\frac{dx^a}{d\epsilon}\right|_{\epsilon=0}\right) = (x_0^a, x_1^a) \in TM \tag{4.46}$$

$$\left(x^a, \left.\frac{dx^a}{d\epsilon}\right|_{\epsilon=0}, \left.\frac{dx^a}{d\epsilon}\right|_{\epsilon=0}, \left.\frac{d^2x^a}{d\epsilon^2}\right|_{\epsilon=0}\right) = (x_0^a, x_1^a, x_1^a, x_2^a) \in TTM \tag{4.47}$$

$$\left(x^a, \left.\frac{dx^a}{d\epsilon}\right|_{\epsilon=0}, \left.\frac{dx^a}{d\epsilon}\right|_{\epsilon=0}, \left.\frac{d^2x^a}{d\epsilon^2}\right|_{\epsilon=0}, \left.\frac{dx^a}{d\epsilon}\right|_{\epsilon=0}, \left.\frac{d^2x^a}{d\epsilon^2}\right|_{\epsilon=0}, \left.\frac{d^2x^a}{d\epsilon^2}\right|_{\epsilon=0}, \left.\frac{d^3x^a}{d\epsilon^3}\right|_{\epsilon=0}\right)$$
$$= (x_0^a, x_1^a, x_1^a, x_2^a, x_1^a, x_2^a, x_2^a, x_3^a) \in TTTM. \tag{4.48}$$

Each time we copy the last list and then copy it again adding one to each subscript. To count how many of each type of derivative we get, we may write this symbolically as

$$\left(1 + \frac{d}{d\epsilon}\right)^J. \tag{4.49}$$

From the binomial expansion, we get

$$\frac{J!}{(J-k)!k!} \tag{4.50}$$

copies of x_k^a in our list.

4.5.2. Symplectic Structure on the Iterated Tangent Bundle

But now recall that TM has a natural symplectic structure pulled back via ω from T^*M. We may use this to obtain a natural symplectic structure on TTM and then TT^*M, etc. We have just constructed a natural injection of $JM \to T^JM$, taking $\vec{\ }$ to an appropriate diagonal. We may pullback the symplectic structure on T^JM to get a natural one on JM.

4.5.2.1. Coordinate Description of Symplectic Structure

Let us introduce coordinates $\{y_0^a\}$ on M, $\{y_0^a, y_1^a\}$ on TM, $\{y_0^a, y_1^a, y_{10}^a, y_{11}^a\}$ on TTM. etc. Here we are using

$$y^a_{d_1,\ldots,d_J} \tag{4.51}$$

on T^JM where $d_i = 0, 1$ and leading zeroes are supressed. $y_{1,d,\ldots,d}$ are the coordinates in the fiber over the space described by $y_{d,\ldots,d}$. Let the symplectic structure on M be $\omega = \omega_{ab}\ dy_0^a \wedge dy_0^b$. T^*M pairs y_0 with y_1 and on TM the y_1 factor is twisted by ω. Thus the symplectic structure on TM is

$$\omega_{ab}\ dy_0^a \wedge dy_1^b. \tag{4.52}$$

T^*TM would pair y_0 with y_{10} and y_1 with y_{11}. On TTM y_{10} and y_{11} are twisted by TM's symplectic structure. Thus the structure for TTM is

$$\omega_{ab}dy_0^a \wedge dy_{11}^b - \omega_{ab}dy_1^a \wedge dy_{10}^b = \omega_{ab}(dy_0^a \wedge dy_{11}^b + dy_{10}^a \wedge dy_1^b). \tag{4.53}$$

If we think of the subscript as a binary number, then the prescription is to pair each y with the y whose digits have 1's and 0's switched. Thus it pairs y_m with

y_{2^J-1-m} such that the even binary number is the first in the wedge product. Thus the symplectic structure on T^JM is given by

$$\sum_{a,b}^{2N}\sum_{m=0}^{2^m-1}(-1)^m\frac{1}{2}\omega_{ab}\,dy^a_m\wedge dy^b_{2^J-1-m}. \tag{4.54}$$

How does the jet bundle map into T^JM in these coordinates? Each time there is a 1 in the binary expansion of m, it indicates that y^a_m coordinatizes another tangent to a tangent, i.e. another derivative in ϵ. Thus the injection of JM into T^JM is given by

$$y^a_m = x^a_{\text{sum of binary digits of m}}. \tag{4.55}$$

4.5.3. Pulled Back Symplectic Structure on the Jet Space

We may get a symplectic form on JM by pulling back the one on T^JM. This amounts to substituting the appropriate x for each y. Since 2^J-1-m is m with all 1's and 0's switched, the sum of the 1's in 2^J-1-m is J minus the number of 1's in m. Thus x^a_k is paired with $\omega_{ab}x^a_{J-k}$. We see that there are $J!/[k!(J-k)!]$ ways of choosing m with k 1's, and so $dx^a_k\wedge dx^b_{J-k}$ will get this coefficient. The pulled back symplectic form on JM is thus

$$\sum_{a,b=1}^{2N}\sum_{k=0}^{J}\frac{J!}{k!(J-k)!}\frac{1}{2}\omega_{ab}\,dx^a_k\wedge dx^b_{J-k}. \tag{4.56}$$

The corresponding Poisson bracket is exactly the one we obtained in the previous section. We have therefore shown that this bracket really is coordinate independent.

4.6. Relation to the Path Space Bracket

In this section we will show how JM's symplectic structure is related to P_1M's. The result is something like the J-th derivative of the path structure. This interpretation will be made explicit in the next section. We saw in section 4.3 that P_1M was essentially a direct integral of the spaces M_ϵ as ϵ goes from 0 to 1. The path space symplectic structure and Hamiltonian are integrals of the corresponding structures on M.

4.6.1. Weighted Path Bracket and Hamiltonian

Because the spaces do not interact for different values of ϵ, we may obtain equally viable structures by putting a weighting factor $\beta(\epsilon)$ into the integrals. Assuming that $\beta(\epsilon)$ doesn't vanish anywhere, we get the correct dynamics on P_1M with

$$\tilde{\omega}_{\beta p}(\tilde{V}_1, \tilde{V}_2) \equiv \int_0^1 \beta(\epsilon)\, \omega_{p(\epsilon)}\big(V_1(\epsilon, p(\epsilon)), V_2(\epsilon, p(\epsilon))\big)\, d\epsilon \tag{4.57}$$

$$\tilde{H}_\beta(p) \equiv \int_0^1 \beta(\epsilon)\, H\big(\epsilon, p(\epsilon)\big)\, d\epsilon. \tag{4.58}$$

When we project to one of the smaller spaces, we essentially take β to vanish on some domain. $P_\alpha M$ comes from taking $\beta(\epsilon) = 1$ for $0 \leq \epsilon \leq \alpha$ and $\beta(\epsilon) = 0$ for $\alpha < \epsilon \leq 1$. Taking β to be a delta function $\beta(\epsilon) = \delta(\epsilon)$ gives the original structure on $P_0M \sim M$.

In general, if we have a closed two-form on a manifold, it may not be symplectic due to degenerate directions (i.e. there exist tangent vectors such that the one-form that results from inserting them into the two-form vanishes). The set of degenerate

directions forms a subspace of the tangent space at each point of the manifold. Near points where the degenerate subspaces don't change in dimension, we may attempt to find a foliation by degenerate submanifolds (i.e. a smooth collection of disjoint submanifolds called leaves of the same dimension as and tangent to the degenerate subspaces, whose union is the whole manifold). Usually this is not possible even locally. Conditions under which it is possible locally are given by Frobenius' theorem (see [Spivak, 1979] p. 257). In the situation we are considering, the condition that the two-form be closed is sufficient to guarantee that the requirements of the Frobenius theorem are satisfied by the degenerate subspaces. We would like to consider the quotient of our manifold by the degenerate foliation. The quotient is a manifold whose points are the leaves of the foliation. It is always possible to form such a quotient locally and sometimes it is possible globally. The original degenerate two-form gives rise to a non-degenerate *symplectic* two-form on the quotient. The value of this two-form on two tangent vectors on the quotient is defined to be equal to the value of the original two-form applied to any two vectors on the original manifold that project to the two quotient vectors. The result of this is independent of the point we lift the vectors to because the original two-form is closed. It is independent of the vectors we choose at that point because the different choices differ by degenerate vectors on which the two-form vanishes.

These very general considerations apply to the path symplectic structure $\tilde{\omega}_\beta$ defined in terms of β. A degenerate vector on the path space at p is a vector field along p on which $\tilde{\partial}_\beta$ vanishes. Examining $\tilde{\omega}_\beta$'s defining integral and using the fact that ω is non-degenerate, we see that if β is a function (as opposed to a

distribution) then a degenerate V is non-zero only at those ϵ where β vanishes. If $\beta = 1$ for $0 \leq \epsilon \leq \alpha$ and $= 0$ for $\alpha < \epsilon \leq 1$, a degenerate vector is described by a $V(\epsilon)$ which is non-zero only for $\epsilon > \alpha$. The degenerate foliation has leaves that are some given $p(\epsilon)$ for $0 \leq \epsilon \leq \alpha$ and all possible extensions for $\alpha < \epsilon \leq 1$. The quotient by the degenerate foliation is then exactly $P_\alpha M$ and the quotient symplectic structure is $\tilde{\omega}_\beta$ viewed as acting on vector fields along paths defined for $0 \leq \epsilon \leq \alpha$. The case of distributional β's may be studied in a similar way.

4.6.2. Jet Bracket Arises from Derivative of Delta Function Weighting

We claim that taking β to be the Jth derivative of a delta function gives us the J-jet structure. Consider

$$\tilde{\omega}_{Jp}(\tilde{V}_1, \tilde{V}_2) \equiv \int_0^1 \left((-1)^J \frac{d^J}{d\epsilon^J} \delta(\epsilon) \right) \omega\big(V_1(\epsilon, p(\epsilon)), V_2(\epsilon, p(\epsilon))\big)\, d\epsilon \tag{4.59}$$

$$\tilde{H}_J(p) \equiv \int_0^1 \left((-1)^J \frac{d^J}{d\epsilon^J} \delta(\epsilon) \right) H\big(\epsilon, p(\epsilon)\big)\, d\epsilon. \tag{4.60}$$

This structure does respect the jet equivalence classes. We take symplectic coordinates on M so that $\omega = \frac{1}{2}\omega_{ab}\, dx^a \wedge dx^b$ and use the coordinates $\{x_0^a, \ldots, x_J^a\}$ as defined before on the J-jet space. Recall that a vector

$$\sum_{a=1}^{2N} \sum_{k=0}^{J} V_k^a \frac{\partial}{\partial x_k^a} \tag{4.61}$$

corresponds to a vector X along a curve with the same jet according to

$$V_k^a(x_0, \ldots, x_J) = \left. \frac{d^k}{d\epsilon^k} \right|_{\epsilon=0} X^a \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J \right). \tag{4.62}$$

We see that $\tilde{\omega}_J$ really depends only on the J-th jet of the path and the J-th jet of the vector field:

$$\begin{aligned}\tilde{\omega}_J(\tilde{X}_1, \tilde{X}_2) &= \frac{1}{2}\frac{d^J}{d\epsilon^J}\bigg|_{\epsilon=0}\left(\omega_{ab}\, X_1^a(\epsilon, x_0 + \ldots + \frac{\epsilon^J}{J!}x_J) X_2^b(\epsilon, x_0 + \ldots + \frac{\epsilon^J}{J!}x_J)\right) \\ &= \frac{1}{2}\sum_{k=o}^{J}\frac{J!}{k!(J-k)!}\omega_{ab}\left(\frac{d^k}{d\epsilon^k}\bigg|_{\epsilon=0} X_1^a(\epsilon, x_0 + \ldots + \frac{\epsilon^J}{J!}x_J)\right) \\ &\qquad \left(\frac{d^{J-k}}{d\epsilon^{J-k}}\bigg|_{\epsilon=0} X_2^b(\epsilon, x_0 + \ldots + \frac{\epsilon^J}{J!}x_J)\right)\end{aligned} \tag{4.63}$$

by the Leibniz rule for derivatives.

But these derivatives give the components of the jets of X:

$$\tilde{\omega}_J(\tilde{X}_1, \tilde{X}_2) = \frac{1}{2}\sum_{k=0}^{J}\frac{J!}{k!(J-k)!}\,\omega_{ab}\, V_{1,k}^a\, \tilde{v}_{2,J-k}^b. \tag{4.64}$$

So $\tilde{\omega}_J$ is really the pullback along $P_1 M \to JM$ of the form which we discovered before:

$$\sum_{a,b=1}^{2N}\sum_{k=0}^{J}\frac{J!}{k!(J-k)!}\frac{1}{2}\omega_{ab}\, dx_k^a \wedge dx_{J-k}^b. \tag{4.65}$$

4.6.3. Jet Hamiltonian from Derivative of Delta Function Weighting

Similarly,

$$\tilde{H}_J(p) = \frac{d^J}{d\epsilon^J}\bigg|_{\epsilon=0} H\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) = \tilde{H}(x_0, \ldots, x_J) \tag{4.66}$$

is the pullback of $\tilde{H}$ along $P_1 M \to JM$.

4.7. Jet Space as Derivative

In this section we will make more explicit the sense in which the symplectic structure on JM is a derivative of ω. Tulczyjew and Kijowski have shown how the natural structure on TM is a first derivative in [Kijowski and Tulczyjew, 1979].

4.7.1. The Sheet Quotient Spaces

We will need the spaces defined by

$$P_{0,\delta}M \equiv \Big\{\text{equivalence classes in } P_1M \text{ where } p_1 \sim p_2 \text{ iff } p_1(0) = p_2(0) \text{ and } p_1(\delta) = p_2(\delta)\Big\} \tag{4.67}$$

and in general by

$$P_{0,\delta,\ldots,J\delta}M \equiv \Big\{\text{equivalence classes in } P_1M \text{where } p_1 \sim p_2 \text{ iff } p_1(\epsilon) = p_2(\epsilon) \text{ for } \epsilon = 0,\delta,\ldots,J\delta\Big\}. \tag{4.68}$$

We require the curves to agree on sheets $\epsilon = 0, \delta, \ldots, J\delta$ spaced by δ. Let us call the coordinates on these sheets $\{z_0^a, z_1^a, \ldots, z_J^a\}$.

4.7.2. Sheet Symplectic Structure and Hamiltonian

We get the correct dynamics on these sheets if we take the symplectic structure and Hamiltonian to be

$$\omega_\delta \equiv \sum_{k=0}^{J} \frac{1}{2}\beta_k\ \omega_{ab}\ dz_k^a \wedge dz_k^b \qquad \beta_k \neq 0 \tag{4.69}$$

$$\check{H}_\delta(z_0, \ldots, z_J) \equiv \sum_{k=0}^{J} \beta_k\ H(k\delta, z_k) \tag{4.70}$$

These are discrete versions of the infinite dimensional structures mentioned in the last section. We want to map JM into this space and pull back ω_δ and H_δ. We choose the β_k as functions of δ so that the limit $\delta \to 0$ is both non-singular and non-trivial.

4.7.3. Map Between Sheet Space and Jet Space

With coordinates $\{x_0^a, \ldots, x_J^a\}$ on JM we can define the map

$$z_k^a = \sum_{m=0}^{J} k^m \frac{\delta^m}{m!} x_m^a. \tag{4.71}$$

This identifies $\{x_0, \ldots, x_J\}$ with the points where the curve $x_0 + \epsilon x_1 + \ldots + \frac{\epsilon^J}{J!} x_J$ intersects the sheets introduced above.

4.7.4. The Pulled Back Sheet Symplectic Structure and Hamiltonian

The pulled back ω_δ on JM is then

$$\omega_\delta = \frac{1}{2} \sum_{k,m,n=0}^{J} k^{m+n} \frac{\delta^{m+n}}{m!\, n!} \beta_k \, \omega_{ab} \, dx_m^a \wedge dx_n^b. \tag{4.72}$$

The pulled back Hamiltonian is

$$H_\delta(x_0, \ldots, x_J) = \sum_{k=0}^{J} \beta_k \, H\left(k\delta, x_0 + k\delta x_1 + \cdots + \frac{(k\delta)^J}{J!} x_J\right). \tag{4.73}$$

4.7.5. Sheet Structures Asymptote to Jet Structures for Small Spacing

Looking at the expression for H_δ, we see that there is no hope for x_J dependence as $\delta \to 0$ if $\beta_k \to \infty$ slower than δ^{-J} as $\delta \to 0$. We therefore assume that $\beta_k = b_k \delta^{-J}$ where b_k is independent of δ. δ then only appears in the expression for ω_δ as δ^{m+n-J}. Terms with $m+n > J$ will vanish when $\delta \to 0$. For ω_δ to be defined as $\delta \to 0$, we must choose b_0 so that the sum of terms with $i = m+n < J$ must have a vanishing coefficient. Thus the b_k must satisfy

$$\sum_{k=0}^{J} \sum_{m+n=i} k^{m+n} \frac{b_k}{m!\, n!} = 0 \qquad 0 \le i \le J-1. \tag{4.74}$$

This may be rewritten

$$= \sum_{k=0}^{J} \sum_{m=0}^{i} k^i \frac{b_k}{m!(i-m)!} \tag{4.75}$$

and the binomial theorem gives $\sum_{m=0}^{i} [m!(i-m)!]^{-1} = 2^i/i!$. Thus

$$\sum_{k=0}^{J} \frac{2^i\, k^i}{i!} b_k = 0 \qquad \text{for } 0 \le i \le J-1 \tag{4.76}$$

or

$$\sum_{k=0}^{J} k^i b_k = 0 \qquad \text{for } 0 \le i \le J-1. \tag{4.77}$$

This is J equations for $J+1$ unknowns. We may take $b_0 = 1$ and remember that an arbitrary multiplicative factor is allowed. We solve the equations by introducing a generating function

$$f(x) \equiv \sum_{k=0}^{J} b_k x^k. \tag{4.78}$$

The condition $b_0 = 1$ becomes $f(0) = 1$. Notice that

$$\left(x \frac{d}{dx} \right)^i f = \sum_{k=0}^{J} k^i b_k x^k. \tag{4.79}$$

So

$$\left(x\frac{d}{dx}\right)^i f\Big|_{x=1} = 0 \qquad \text{for } 0 \le i \le J-1. \tag{4.80}$$

This easily implies that

$$\frac{d^i}{dx^i} f\Big|_{x=1} = 0 \qquad \text{for } 0 \le i \le J-1. \tag{4.81}$$

Using these equations we may Taylor expand f about $x = 1$

$$f(x) = 0 + 0 + \cdots + \frac{C_J}{J!}(x-1)^J + \cdots. \tag{4.82}$$

Because f is a J-th order polynomial and $f(0) = 1$, we see that $C_J = J!(-1)^J$ and so

$$f(x) = (1-x)^J. \tag{4.83}$$

From the binomial expansion

$$b_k = (-1)^k \frac{J!}{k!(J-k)!} \quad \text{and} \quad \sum_{k=0}^{J} k^J = \frac{d^J}{dx^J}(1-x)^J\Big|_{x=1} = J!. \tag{4.84}$$

The only terms left in ω have $m + n = J$ and give us

$$\begin{aligned}\omega_\delta &= \sum_{k,m=0}^{J} \frac{1}{2} k^J \frac{1}{m!(J-m)!} b_k\, \omega_{ab}\, dx^a_m \wedge dx^b_{J-m} \\ &= \frac{1}{2}\sum_{m=0}^{J} \frac{J!}{m!(J-m)!}\, \omega_{ab}\, dx^a_m \wedge dx^b_{J-m}\end{aligned} \tag{4.85}$$

just as before.

Now

$$\begin{aligned}&H_\delta(x_0, \ldots, x_J) \\ &\quad = \sum_{k=0}^{J} \frac{1}{\delta^J}(-1)^k \frac{J!}{k!(J-k)!} H\left(k\delta, x_0 + k\delta x_1 + \cdots + \frac{(k\delta)^J}{J!} x_J\right).\end{aligned} \tag{4.86}$$

If we call S_δ the operator which shifts a function of ϵ by an amount δ, so $S_\delta f(\epsilon) = f(\epsilon + \delta)$, then we see that

$$H_\delta(x_0, \ldots, x_J) = \sum_{k=0}^{J} \frac{1}{\delta^J}(-1)^k \frac{J!}{k!(J-k)!} \left. S_\delta^k \right|_{\epsilon=0} H\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right)$$

$$= \frac{(1 - S_\delta)^J}{\delta^J} H\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right). \tag{4.87}$$

As δ goes to zero, the operator $(1 - S_\delta)/\delta$ becomes $d/d\epsilon$. In the limit we have

$$\tilde{H}(x_0, \ldots, x_J) = \left. \frac{d^J}{d\epsilon^J} \right|_{\epsilon=0} H\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right) \tag{4.88}$$

just as before. Our jet structure thus comes out of a limiting process almost uniquely. Similar kinds of arguments arise in the theory of finite differences used to approximate derivatives numerically. A good reference written from a theoretical perspective is [Stoer and Burlirsh, 1980].

4.8. Jets and Symmetry

In this section we study Hamiltonian group actions in the perturbation context we have been considering. We need to understand the perturbation analogs of group actions and the process of reduction. We first show that a Hamiltonian group action of G on M lifts to the path space P_1M and jet space JM. We calculate the momentum maps and show that they are equivariant. We then introduce the group PG of paths in G and the group JG of their J-jets. These too act on PM and JM and we find their momentum maps. Next we consider the case where M is a coadjoint orbit in g^* and show that PM is a coadjoint orbit in Pg^* and JM in Jg^*. We calculate the corresponding Lie symplectic structures and show that they give the path bracket and jet bracket that we discovered earlier. We then study the process of reduction and show that the reduced path space is the path space of the reduced space and the reduced jet space is the jet space of the reduced space.

4.8.1. ϵ-dependent Group Actions on M

Our starting point is an ϵ-dependent group action on the manifold:

$$\rho : I \times G \times M \to M. \tag{4.89}$$

Here I is the interval $[0,1]$, G is the group, and M is the manifold. If e is the identity of G then

$$\rho(\epsilon, e, m) = m \tag{4.90}$$

and

$$\rho(\epsilon, g_1, \rho(\epsilon, g_2, m)) = \rho(\epsilon, g_1 \cdot g_2, m). \tag{4.91}$$

For a compact group G acting on a compact manifold M, we can eliminate the ϵ-dependence of the group action by an ϵ-dependent coordinate change in M (this result is due to Palais). This is not true for non-compact G as the example of the real line demonstrates. In this case group actions are the flows of vector fields. Even if we restrict attention to arbitrarily small neighborhoods of $\epsilon = 0$, only the so-called "structurally stable" vector fields can be made ϵ-independent by an ϵ-dependent coordinate change.

4.8.1.1. Lift to G Action on the Path Space

This action lifts to the space of paths PM by defining

$$\tilde{\rho} : G \times PM \to PM \quad \text{by} \quad \tilde{\rho}(g,p)(\epsilon) \equiv \rho(\epsilon, g, p(\epsilon)). \tag{4.92}$$

4.8.1.2. Lift to G Action on the Jet Space

This respects the equivalence classes that define the J-jet space JM. We introduce coordinates x^a on M and x^a_i, $i = 0, 1, \ldots, J$ on JM as before. The components of ρ will be written ρ^a. The action of G on JM is then given by

$$\bar{\rho}^a_i(g,(x^0,x^1,\ldots,x^J)) \equiv \frac{d^i}{d\epsilon^i}\bigg|_{\epsilon=0} \rho^a\left(\epsilon, g, \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right)\right). \tag{4.93}$$

4.8.1.3. Maps from the Lie Algebra to Vector Fields on M, PM, and JM

The action of G on $I \times M$ gives us a map from the Lie algebra g of G to vector fields on $I \times M$ that leave ϵ invariant. If $v \in g$ is tangent to a curve $I \to G$, so

$$v = \left.\frac{d}{dt}\right|_{t=0} g(t), \tag{4.94}$$

then X_v on $I \times M$ is defined by

$$X_v(\epsilon, m) = \left.\frac{d}{dt}\right|_{t=0} \rho(\epsilon, g(t), m). \tag{4.95}$$

This induces a map from g to TPM defined by

$$\tilde{X}_v(p)(\epsilon) \equiv X_v(\epsilon, p(\epsilon)) \tag{4.96}$$

and to TJM defined in coordinates by

$$\begin{aligned}(\tilde{X}_v)^a_J &\equiv \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} X^a_v\left(\epsilon, \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right)\right) \\ &= \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} \left.\frac{d}{dt}\right|_{t=0} \rho^a\left(\epsilon, g(t), \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right)\right).\end{aligned} \tag{4.97}$$

4.8.1.4. The Momentum Map

In the case where M, ω is symplectic and our G action has an ϵ-dependent equivariant momentum map

$$J : I \times M \to g^* \tag{4.98}$$

then

$$i_{X_v}\omega = \langle dJ, v \rangle \tag{4.99}$$

on each ϵ =constant.

4.8.1.5. Momentum Map on PM

This gives us a momentum map for G's action on PM with the symplectic structure $\tilde{\omega}$ given by

$$\tilde{J} : PM \to g^* \quad \text{by} \quad p \mapsto \int_0^1 J(\epsilon, p(\epsilon))\, d\epsilon. \tag{4.100}$$

It's easy to see that this gives the correct action and is equivariant because J is.

4.8.1.6. Momentum Map on JM

Using coordinates $(x_0^a, \ldots, x_J^a)$ on JM as before where the x^a's are canonical, we see that we may define a momentum map from $JM \to g^*$ by

$$\tilde{J}(x_0, \ldots, x_J) = \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} J\left(\epsilon, \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right)\right). \tag{4.101}$$

4.8.1.7. Equivariance of JM's Momentum Map

The equivariance follows from that of J, as follows. Equivariance of J says

$$J(\epsilon, \rho(\epsilon, g, m)) = Ad_g^* \cdot J(\epsilon, m). \tag{4.102}$$

On JM we see

$$\begin{aligned} \tilde{J}(\bar{\rho}(g, (x_0, \ldots, x_J))) &= \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} J\left(\epsilon, \sum_i \frac{\epsilon^i}{i!} \bar{\rho}_i(g, (x_0, \ldots, x_J))\right) \\ &= \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} J\left(\epsilon, \sum_{i=0}^{J} \frac{\epsilon^i}{i!} \left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} \rho\left(\epsilon, g, \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right)\right)\right). \end{aligned} \tag{4.103}$$

The $d^J/d\epsilon^J\big|_{\epsilon=0}$ allows us to neglect terms in ϵ of higher than the Jth order. But to this order any function $f(\epsilon)$ is equal to its Jth order Taylor series

$$\sum_{i=0}^{J} \frac{\epsilon^i}{i!} \frac{d^i}{d\epsilon^i}\bigg|_{\epsilon=0} f(\epsilon) \sim f(\epsilon). \tag{4.104}$$

Thus

$$\begin{aligned} &= \frac{d^J}{d\epsilon^J}\bigg|_{\epsilon=0} J(\epsilon, \rho(\epsilon, g, \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right))) \\ &= \frac{d^J}{d\epsilon^J}\bigg|_{\epsilon=0} Ad_g^* \cdot J(\epsilon, \left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!} x_J\right)) \\ &= Ad_g^* \cdot \bar{J}(x_0, \ldots, x_J) \end{aligned} \tag{4.105}$$

as desired.

4.8.2. The Path Group: PG

When we do reduction, we'll want a much larger group to work with. In essence, we want a symmetry that can act on different level sets of ϵ independently. Thus we define

$$PG \equiv \left\{\text{all paths } \tilde{g} : I \to G, \;\; \tilde{g} : \epsilon \mapsto \tilde{g}(\epsilon)\right\}. \tag{4.106}$$

The product in this space is defined as

$$\tilde{g}_1 \cdot \tilde{g}_2(\epsilon) = \tilde{g}_1(\epsilon) \cdot \tilde{g}_2(\epsilon). \tag{4.107}$$

The identity in PG is the constant path at the identity $\tilde{e}(\epsilon) = e$ in G.

4.8.2.1. The Path Lie Algebra: Pg

Taking infinitesimal elements, we see that the Lie algebra is

$$Pg \equiv \left\{\text{all paths } \tilde{v} : I \to g, \quad \tilde{v} : \epsilon \mapsto \tilde{v}(\epsilon)\right\}. \tag{4.108}$$

The notation Pg would be ambiguous in that it could mean the path space of the Lie algebra or the Lie algebra of the path group, except that these two spaces are naturally isomorphic. The Lie bracket of two elements is defined pointwise

$$[\tilde{v}_1, \tilde{v}_2](\epsilon) = [\tilde{v}_1(\epsilon), \tilde{v}_2(\epsilon)]. \tag{4.109}$$

4.8.2.2. The Dual of the Lie Algebra of the Path Group

The dual of the Lie algebra is all (distributional) paths in g^*:

$$Pg^* = \left\{\tilde{\alpha} : I \to g^*, \quad \tilde{\alpha} : \epsilon \mapsto \tilde{\alpha}(\epsilon)\right\}. \tag{4.110}$$

The pairing is given by

$$\langle \tilde{v}, \tilde{\alpha} \rangle \equiv \int_0^1 \langle \tilde{v}(\epsilon), \tilde{\alpha}(\epsilon) \rangle d\epsilon. \tag{4.111}$$

4.8.2.3. The Action of the Path Group on the Path Space

PG acts on PM by

$$\tilde{R} : PG \times PM \to PM \quad \tilde{R}(\tilde{g}, p)(\epsilon) \equiv \rho(\epsilon, \tilde{g}(\epsilon), p(\epsilon)). \tag{4.112}$$

G is the subgroup of PG with $\tilde{g}(\epsilon) = g$ and this action of PG on PM extends the action of G. The momentum map for this action is $PM \to Pg^*$ by $p \mapsto J(\cdot, p(\cdot))$ where $\cdot$ is the parameter ϵ to be inserted in an element of $\tilde{g}^*$ to get $I \to g^*$.

4.8.3. The Jet Group: JG

Consider paths in G which begin at the identity and remain there to order ϵ^J, i.e. $\tilde{g}(0) = e$, $\partial^k/\partial\epsilon^k\big|_{\epsilon=0}\tilde{g}(\epsilon) = 0$ for $1 \leq k \leq J$. The product of two such paths is such a path as are inverses and the identity. Let us call this subgroup of PG, PG_{J-flat} (since a function whose derivative at $\epsilon = 0$ vanishes is flat there and one whose higher derivatives vanish is very flat there, i.e. J-flat). With the action $\tilde{R}$ on PM, we see that this subgroup leaves invariant the $J-jet$ equivalence classes, so we are naturally interested in

$$PG_J \equiv PG/PG_{J-flat}, \tag{4.113}$$

assuming that this is a group. It is easier to show PG_{J-flat} is a normal subgroup of PG by considering Lie algebras (recall that a normal subgroup $H \subset G$ satisfies $gHg^{-1} = H$ for every $g \in G$ and that this is a necessary and sufficient condition for the quotient G/H to be a group).

The Lie algebra of PG_{J-flat} is clearly

$$Pg_{J-flat} \equiv \{\tilde{v} \in Pg \mid \left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} \tilde{v}(\epsilon) = 0 \quad \text{for} \quad 0 \leq i \leq J\}. \tag{4.114}$$

To check for normality, we want to show that

$$[Pg, Pg_{J-flat}] \subset Pg_{J-flat} \tag{4.115}$$

(this is the Lie algebra analog of normality). But if

$$\tilde{u}(\epsilon) = \tilde{u}_0 + \epsilon\tilde{u}_1 + \cdots \in Pg \tag{4.116}$$

and

$$\tilde{v}(\epsilon) = \frac{\epsilon^{J+1}}{(J-1)!}\tilde{v}_{J+1} + \cdots \in Pg_{J-flat} \tag{4.117}$$

then

$$[\tilde{u}, \tilde{v}](\epsilon) = \frac{\epsilon^{J+1}}{(J+1)!}[\tilde{u}_0, \tilde{v}_{J+1}] + \cdots \in Pg_{J-flat} \tag{4.118}$$

as desired. Ordinarily, there are problems in relating results about infinite-dimensional Lie algebras to results about the corresponding infinite-dimensional groups (for example, the image of the exponential map of the diffeomorphism group does not contain any neighborhood of the identity). Path groups are especially well-behaved, however, and the argument here is valid in this context (for more discussion see Appendix A of [Freed and Uhlenbeck, 1984]). Thus we may introduce the group of J-jets of paths in G:

$$JG \equiv PG/PG_{J-flat}. \tag{4.119}$$

4.8.3.1. The Lie Algebra of the Jet Group: Jg

Its Lie algebra is

$$Jg \equiv Pg/Pg_{J-flat}. \tag{4.120}$$

As for Pg, the potentially ambiguous notation is not, by a simple theorem. We may put coordinates $u_0^c, u_1^c, \ldots, u_J^c$ on Jg by associating $(u_0, u_1, \ldots, u_J)$, $u_i \in g$ with the equivalence class of $u_0 + \epsilon u_1 + \cdots + (\epsilon^J/J!)u_J$. The bracket is then

$$\begin{aligned}
&[(u_0, u_1, \ldots, u_J), (v_0, v_1, \ldots, v_J)] \\
&\quad = [u_0 + \epsilon u_1 + \cdots + \frac{\epsilon^J}{J!}u_J, v_0 + \epsilon v_1 + \cdots + \frac{\epsilon^J}{J!}v_J] \\
&\quad - \sum_{k=0}^{J} \frac{\epsilon^k}{k!}\left(\sum_{i=0}^{k} \frac{k!}{i!(k-i)!}[u_i, v_{k-i}]\right).
\end{aligned} \tag{4.121}$$

So

$$[(u_0, u_1, \ldots, u_J), (v_0, v_1, \ldots, v_J)]_k = \sum_{i=0}^{k} \frac{k!}{i!(k-i)!}[u_i, v_{k-i}] \tag{4.122}$$

This is an explicit formula for the Lie bracket of the jet Lie algebra. We shall need it in determining the Lie-Poisson bracket.

4.8.3.3. Homomorphism from Jg to Vector Fields on JM

The group JG acts on JM and so we get a Lie algebra homomorphism from Jg to vector fields on JM (these will generate symmetries that hold up to order ϵ^J for the full dynamics but are exact symmetries on JM). Recall that X_v is the image of a map from $v \in g$ to vector fields on $I \times M$. When v depends on ϵ as above, we see that

$$\begin{aligned}
(\tilde{X}_{(v_0,v_1,\ldots,v_J)})_i &= \\
&= \left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} X_{(v_0+\epsilon v_1+\cdots+(\epsilon^J/J!)v_J)}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) \\
&= \left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} X_{v_0}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) + \\
&\quad + \left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} \epsilon X_{v_1}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) + \\
&\quad + \cdots + \left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} \frac{\epsilon^J}{J!} X_{v_J}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) \\
&= \sum_{k=0}^{i} \left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} \frac{\epsilon^k}{k!} X_{v_k}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right) \\
&= \sum_{k=0}^{i} \frac{i!}{k!(i-k)!}\frac{1}{k!}\left.\frac{d^{i-k}}{d\epsilon^{i-k}}\right|_{\epsilon=0} X_{v_k}\left(\epsilon, x_0 + \epsilon x_1 + \cdots + \frac{\epsilon^J}{J!}x_J\right).
\end{aligned} \tag{4.123}$$

So:

$$(\tilde{X}_{(v_0,v_1,\ldots,v_J)})_i = \sum_{k=0}^{i} \frac{i!}{k!(i-k)!k!}(\tilde{X}_{v_k})_{i-k}. \tag{4.124}$$

This gives the action of a jet Lie algebra element on the jet space of M in terms of the ϵ-dependent action of the group on M. If we have an approximate symmetry in the sense that our ϵ-dependent action preserves some structure to order J, then this J-jet action will be an exact symmetry.

4.8.3.3. The Dual of the Jet Lie Algebra: Jg^*

The dual of the Lie algebra, Jg^*, may be coordinatized by J elements of g^* with the pairing

$$\langle(\alpha_0, \alpha_1, \ldots, \alpha_J), (v_0, \ldots, v_J)\rangle = \sum_{i=0}^{J} \langle \alpha_i, v_i \rangle \quad \text{for} \quad \alpha_i \in g^*, \; v_i \in g. \tag{4.125}$$

4.8.3.4. The Jet Momentum Maps

By the definition given in section 2.7.2 of momentum maps on M we have in general that

$$X_v^a(\epsilon, m) = \omega^{ab} \frac{\partial}{\partial x^b} \langle J(\epsilon, m), v \rangle. \tag{4.126}$$

In this case

$$v = v_0 + \epsilon v_1 + \cdots + \frac{\epsilon^J}{J!} v_J \tag{4.127}$$

is ϵ-dependent, but nonetheless,

$$\langle J(\epsilon, m), v_0 + \cdots + \frac{\epsilon^J}{J!} v_J \rangle \tag{4.128}$$

is an ϵ-dependent Hamiltonian on M for the correct action. We know from section 4.4.2 that with

$$\{x_i^a, x_k^b\} = \omega^{ab} \frac{i!k!}{J!} \delta_{i, J-k} \tag{4.129}$$

we will get the correct dynamics from the Jth derivative of the Hamiltonian. So

$$\begin{aligned}\frac{d^J}{d\epsilon^J}\Big|_{\epsilon=0}\langle J(\epsilon,m),v_0+\cdots+\frac{\epsilon^J}{J!}v_J\rangle &= \\ &= \sum_{i=0}^{J}\frac{d^J}{d\epsilon^J}\Big|_{\epsilon=0}\frac{\epsilon^i}{i!}\langle J(\epsilon,m),v_i\rangle \\ &= \sum_{i=0}^{J}\frac{J!}{i!(J-i)!i!}\frac{d^{J-i}}{d\epsilon^{J-i}}\Big|_{\epsilon=0}\langle J(\epsilon,m),v_i\rangle\end{aligned} \tag{4.130}$$

generates the $(v_0,\ldots,v_J)$ dynamics on JM. Thus the momentum map is

$$\tilde{J}_i(x_0,\ldots,x_J)=\frac{J!}{i!(J-i)!i!}\frac{d^{J-i}}{d\epsilon^{J-i}}\Big|_{\epsilon=0}J\left(\epsilon,x_0+\epsilon x_1+\cdots+\frac{\epsilon^J}{J!}x_J\right). \tag{4.131}$$

This momentum map is again equivariant. This is important because this is the approximate constant of motion corresponding to an approximate symmetry. For example, we will see in chapter 5 that adiabatic invariants may be viewed in this way. If we can find an exact symmetry of the dynamics on JM, then this gives a constant of the motion up to order J in ϵ for the full system.

4.8.4. When M is a Coadjoint Orbit with the KKS Symplectic Structure

Thus we have discovered two groups, PG and JG, that contain G. Let us consider the special case in which M is a coadjoint orbit of G in g^* with the canonical Kostant-Kirillov-Souriou symplectic structure. (This is really no limitation since we may take G to be the group of symplectomorphisms of M, g is then Hamiltonian functions, g^* is distributions and we may identify M and its symplectic structure with the orbit of a δ-function and its KKS structure).

4.8.4.1. Coadjoint Action of PG

Because multiplication in PG is ϵ-wise, the adjoint action of PG on Pg is just

$$(\bar{Ad}_{\bar{g}} \cdot \bar{v})(\epsilon) = Ad_{\bar{g}(\epsilon)} \cdot \bar{v}(\epsilon) \tag{4.132}$$

and the coadjoint action is similar.

4.8.4.2. The KKS Symplectic Structure on Coadjoint Orbits in Pg^*

Consider any path

$$\bar{\alpha} : I \to M \subset g^*. \tag{4.133}$$

Since M is a coadjoint orbit of G, under the coadjoint action of PG on $\bar{\alpha}$, at each ϵ we will sweep out a copy of M. Since the different ϵ's are nearly independent, it is easy to see that the coadjoint orbit of PG through $\bar{\alpha}$ is the path space PM of M.

What is the Lie symplectic structure? Given an element $\bar{v}$ of g, we determine a curve of tangent vectors to M, i.e. an element $\bar{V}$ of $T_p PM$, by identifying a path p with an element $\bar{\alpha}$ of Pg^*. We let

$$\bar{\alpha}(\epsilon) = p(\epsilon) \tag{4.134}$$

and take

$$\bar{V}(p) = \bar{ad}^*_{\bar{v}} \bar{\alpha} \tag{4.135}$$

or

$$V(p(\epsilon)) = ad^*_{v(\epsilon)} \alpha(\epsilon). \tag{4.136}$$

Similarly associate $\tilde{u} \in \tilde{g}$ with $\tilde{U} \in T_p PM$. The Lie symplectic structure is then

$$\begin{aligned}\omega_p(\tilde{U},\tilde{V}) &= \omega(\tilde{ad}^*_{\tilde{u}}\tilde{\alpha}, \tilde{ad}^*_{\tilde{v}}\tilde{\alpha}) \\ &= \langle \tilde{\alpha}, [\tilde{u},\tilde{v}] \rangle \\ &= \int_0^1 \langle \alpha(\epsilon), [\tilde{u}(\epsilon), \tilde{v}(\epsilon)] \rangle d\epsilon \\ &= \int_0^1 \omega_{p(\epsilon)}(\tilde{U}(\epsilon), \tilde{V}(\epsilon)) d\epsilon\end{aligned} \tag{4.137}$$

since $\omega_{p(\epsilon)}$ on M was the Lie symplectic structure for G. Thus we obtain our previous path space symplectic structure via a coadjoint orbit of the group of paths in G.

4.8.5. Natural Projections and Injections of G, PG, and JG

We are interested in the groups G, PG, and JG. G is naturally a subgroup of PG given by constant paths:

$$\tilde{g}(\epsilon) = g. \tag{4.138}$$

We also have a projection $PG \to G$ which sends a path to its endpoint:

$$\tilde{g} \mapsto \tilde{g}(0), \tag{4.139}$$

which is also a homomorphism. PG also projects to JG by JG's definition as PG/PG_{J-flat}. G's image in PG gets sent to a subgroup of JG containing those points with zero jets. JG also projects onto G but is not naturally a subgroup of PG.

The picture summarizing these natural maps is

$$\begin{array}{ccccc} PG & \longrightarrow & JG & \rightleftarrows & G. \\ \uparrow & & & & \uparrow \end{array} \tag{4.140}$$

4.8.5.1. Projections and Injections of the Lie Algebras and Duals

Taking derivatives of these maps also gives us maps between the corresponding Lie algebras

$$\begin{array}{ccccc} Pg & \longrightarrow & Jg & \rightleftarrows & g. \\ \uparrow & & \uparrow & & \uparrow \end{array} \tag{4.141}$$

where $\rightarrow$ means a projection and $\leftarrow$ an injection. Elements of Pg are paths in g. They project to their J-jets in Jg and their value at $\epsilon = 0$ in g. Elements of g are sent to J-jets whose 0-jet is the element and whose higher order jets vanish. They are sent to constant paths of the given value in Pg.

Taking duals gives

$$\begin{array}{ccccc} Pg^* & \longrightarrow & Jg^* & \rightleftarrows & g^*, \\ \uparrow & & \uparrow & & \uparrow \end{array} \tag{4.142}$$

where $\rightarrow$ means a projection and $\leftarrow$ means an injection. Elements of Pg^* are distributional paths in g^*. They are sent to g^* by letting them act on an element of g by integrating their value on that element over ϵ:

$$\int_0^1 \langle \alpha(\epsilon), v \rangle d\epsilon. \tag{4.143}$$

Elements of Jg^* are in the dual to the J-jets of paths in g. We may coordinatize it by $J-1$ copies of g^* and give the pairing

$$\langle (\alpha_0, \ldots, \alpha_J), (v_0, \ldots, v_J) \rangle = \langle \alpha_0, v_0 \rangle + \cdots + \langle \alpha_J, v_J \rangle. \tag{4.144}$$

$(\alpha_0, \ldots, \alpha_J)$ is sent to the element

$$\alpha_0 \delta(\epsilon) - \alpha_1 \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \delta(\epsilon) + \cdots + \alpha_J (-1)^J \left.\frac{d^J}{d\epsilon^J}\right|_{\epsilon=0} \delta(\epsilon) \tag{4.145}$$

of Pg^* and to the element α_0 of g^*. g^* injects into Jg^* to send α_0 to $(\alpha_0, 0, \ldots, 0)$.

4.8.6. The Lie Poisson Bracket on g^*

The Lie-Poisson bracket on g^* is

$$\{f,g\}(\alpha) \equiv \left\langle \alpha, \left[\frac{\delta f}{\delta \alpha}, \frac{\delta g}{\delta \alpha}\right]\right\rangle \tag{4.146}$$

and M is a symplectic leaf for this bracket.

4.8.6.1. The Lie Poisson Bracket on Jg^*

The Lie-Poisson bracket on Jg^* may be defined with the help of our coordinates as

$$\begin{aligned}\{f,g\}(\alpha_0,\ldots,\alpha_J) &= \left\langle (\alpha_0,\ldots,\alpha_J), \left[\left(\frac{\delta f}{\delta \alpha_0},\ldots,\frac{\delta f}{\delta \alpha_J}\right), \left(\frac{\delta g}{\delta \alpha_0},\ldots,\frac{\delta g}{\delta \alpha_J}\right)\right]\right\rangle \\ &= \sum_{i=0}^{J}\left\langle \alpha_i, \sum_{k=0}^{i}\frac{i!}{k!(i-k)!}\left[\frac{\delta f}{\delta \alpha_k}, \frac{\delta g}{\delta \alpha_{i-k}}\right]\right\rangle \\ &= \sum_{i=0}^{J}\sum_{k=0}^{i}\frac{i!}{k!(i-k)!}\left\langle \alpha_i, \left[\frac{\delta f}{\delta \alpha_k}, \frac{\delta g}{\delta \alpha_{i-k}}\right]\right\rangle.\end{aligned} \tag{4.147}$$

4.8.6.2. Jg^* as J-jets of Paths in g^*

We would like to identify some coadjoint orbit in Jg^* as M's J-jet space JM. For this we would like to identify J-jets of paths in g^* with elements of Jg^*. What is a natural pairing of J-jets of paths in g with J-jets of paths in g^*? If we change the measure on our pairing of paths to

$$\langle \alpha, v\rangle_J \equiv \int_0^1 \langle \alpha(\epsilon), v(\epsilon)\rangle (-1)^J \frac{d^J}{d\epsilon^J}\delta(\epsilon)d\epsilon, \tag{4.148}$$

where α is a path in g^* and v a path in g, we see that the result only depends on the J-jet of α and the J-jet of v, i.e. it defines a pairing of the appropriate jet spaces. Now we would like to find the identification of J-jets of paths in g^*, say $\alpha_0 + \epsilon\alpha_1 + \cdots + \frac{\epsilon^J}{J!}\alpha_J$, with our previous coordinatization which gives duals of the components $(v_0, v_1, \ldots, v_J)$ of the jet of the path $v_0 + \epsilon v_1 + \cdots + \frac{\epsilon^J}{J!}v_J$. We see that

$$\begin{aligned}
&\left\langle \alpha_0 + \epsilon\alpha_1 + \cdots + \frac{\epsilon^J}{J!}\alpha_J, v_0 + \epsilon v_1 + \cdots + \frac{\epsilon^J}{J!}v_J \right\rangle_J = \\
&\quad = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \left\langle \alpha_0 + \epsilon\alpha_1 + \cdots + \frac{\epsilon^J}{J!}\alpha_J, v_0 + \epsilon v_1 + \cdots + \frac{\epsilon^J}{J!}v_J \right\rangle \\
&\quad = \sum_{i=0}^{J} \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \left\langle \frac{\epsilon^i}{i!}\alpha_i, \frac{\epsilon^{J-i}}{J-i!}v_{J-i} \right\rangle \\
&\quad = \sum_{i=0}^{J} \frac{J!}{i!(J-i)!}\langle \alpha_i, v_{J-i} \rangle.
\end{aligned} \tag{4.149}$$

Thus the J-jet of $\alpha_0 + \epsilon\alpha_1 + \cdots + \frac{\epsilon^J}{J!}\alpha_J$ is associated with the element

$$\left(\frac{J!}{J!\,0!}\alpha_J, \frac{J!}{(J-1)!\,1!}\alpha_{J-1}, \ldots, \frac{J!}{0!\,J!}\alpha_0 \right). \tag{4.150}$$

The J-jet space JM of M gets turned upside down when we put it into Jg^*.

4.8.6.3. Coadjoint Orbits in Jg^*

With this identification of J-jets of g^* with Jg^*, the coadjoint action of JG on J-jets of g^* is just the obvious one: pick a representative path in G with the right J-jet; let it act at each ϵ according to the coadjoint action of G on g^* and then take the J-jet of the resulting path.

By definition this is the way the adjoint action of G on g works. We would like

$$\langle \tilde{ad}_u v, \tilde{ad}_u{}^* \alpha \rangle_J = \langle v, \alpha \rangle \tag{4.151}$$

for all $a \in g$. But since

$$\langle ad_{u(\epsilon)}v(\epsilon), ad^*_{u(\epsilon)}\alpha(\epsilon)\rangle = \langle v(\epsilon)\ \alpha(\epsilon)\rangle \tag{4.152}$$

for all ϵ, it is clear by integrating over ϵ against $(-1)^J \frac{d^J}{d\epsilon^J}\delta(\epsilon)$ that it is true for the jets.

4.8.6.4. JM is a Coadjoint Orbit in Jg^* as a Manifold

Because the coadjoint orbit in Pg^* of a path which lies solely in M is the space of all paths in M, the coadjoint orbit in Jg^* of a J-jet of a path in M is the space of all J-jets of paths in M, i.e. the J-jet space JM.

4.8.6.5. The KKS Symplectic Structure is the Jet Symplectic Structure

So JM is a coadjoint orbit. What is the Lie symplectic structure? Again every J-jet of a path in g determines a J-jet of a path of tangent vectors to M, i.e. an element of TJM, by

$$\begin{aligned} &V_i(\alpha_0 + \epsilon\alpha_1 + \cdots + \frac{\epsilon^J}{J!}\alpha_J) \\ &\qquad = \frac{d^i}{d\epsilon^i}\Big|_{\epsilon=0} ad^*_{v_0+\epsilon v_1+\cdots+\frac{\epsilon^J}{J!}v_J}(\alpha_0 + \epsilon\alpha_1 + \cdots + \frac{\epsilon^J}{J!}\alpha_J). \end{aligned} \tag{4.153}$$

The Lie symplectic structure is then

$$\begin{aligned}
&\omega_{(\alpha_0,\ldots,\alpha_J)}((u_0,\ldots,u_J),(v_0,\ldots,v_J)) = \\
&= (\{\alpha_0,\ldots,\alpha_J),\{(u_0,\ldots,u_J),(v_0,\ldots,v_J)\})_J \\
&= \int_0^1 \left\langle \alpha_0 + \epsilon\alpha_1 + \cdots + \frac{\epsilon^J}{J!}\alpha_J, \right. \\
&\qquad \left. \left[u_0 + \epsilon u_1 + \cdots + \frac{\epsilon^J}{J!}u_J, v_0 + \epsilon v_1 + \cdots + \frac{\epsilon^J}{J!}v_J\right]\right\rangle (-1)^J \frac{d^J}{d\epsilon^J}\delta(\epsilon)d\epsilon \\
&= \int_0^1 \frac{d^1}{d\epsilon^1}\bigg|_{\epsilon=0} \omega_\alpha(U_0 + \epsilon U_1 + \cdots + \frac{\epsilon^J}{J!}U_J, V_0 + \epsilon V_1 + \cdots + \frac{\epsilon^J}{J!}V_J)d\epsilon,
\end{aligned} \tag{4.154}$$

but we have seen that this is exactly the J-jet bracket that we obtained before.

Jerry Marsden has pointed out that some of these symplectic constructions immediately generalize to corresponding constructions on Poisson manifolds. For example, to get a symplectic structure on TM from ω on M, we use ω to identify TM with T^*M and pull back the canonical structure. While a Poisson structure does not define an isomorphism, it does define a map from T^*M to TM. We may *push forward* the canonical Poisson structure on T^*M to obtain a non-canonical one on TM. Iterating this gives Poisson structures on each of the iterated tangent bundles. Presumably one can use these to define a Poisson structure on each of the jet spaces as well. It appears likely that the jet structure derived from a Lie-Poisson bracket on g^* is the Lie-Poisson bracket on Jg^*. Richard Montgomery has recently shown that the second order perturbation jet structure may be extended in this way to arbitrary Poisson manifolds and it appears that the same methods will work to all orders.

4.8.7. JG as a Semi-Direct Product

Notice that the set of elements of Jg of the form

$$(0.0.\ldots.0.u_i.u_{i+1}.\ldots.u_J) \tag{4.155}$$

form a nilpotent ideal, say Jg_i. We see that $ig \sim Jg/Jg_i$. Thus Jg is a semi-direct product of ig and Jg_i. Note in particular that $1G$ is TG and has the group structure of the semi-direct product of G with g with the adjoint action. Alan Weinstein has pointed out that this provides an "explanation" of why most of the coadjoint orbits of the Euclidean group (which is the tangent group of the rotation group) are the tangent bundle of the 2-sphere (most of the coadjoint orbits of the rotation group are 2-spheres). Note, however, that the symplectic structure on these orbits agrees with the jet symplectic structure only on the orbits whose tangent vectors are tangent to the spheres. The other orbits have extra "magnetic terms" in their symplectic forms. If the conjecture at the end of the last section is correct, then the general coadjoint orbits of JG are the symplectic leaves of the jet lift of the Lie Poisson structure on g^* to Jg^*. It is interesting that there is another natural jet type Lie algebra associated with an arbitrary Lie algebra g. Consider J-jets at the origin of *real valued functions* (as opposed to paths) defined on g^*. The Lie Poisson bracket gives this jet space a natural Lie algebra structure. Its dimension is much larger than that of Jg and the relation between the two Lie algebras is not clear.

4.8.8. Jet and Path Reduced Spaces are Reduced Jet and Path Spaces

Let us now return to the general setting of an ϵ dependent action of G on M, ω. If we have an ϵ-dependent invariant function $H(\epsilon)$, it is easy to see that if the orbit spaces for all ϵ are diffeomorphic (for example if the G action is ϵ independent or if G is compact and the variation is small enough) then PG leaves $\tilde{H}$ on PM invariant. Its orbit space is the space of paths in the orbit space of G on M. Similarly, the inverse image of a point in Pg^* under the momentum map $\tilde{J}$ is the space of paths that lie in the corresponding inverse image of J at each ϵ. Thus the reduced space of PG acting on PM is the path space of the reduced spaces at each ϵ.

Similarly JG acts on JM leaving

$$\left.\frac{d^i}{d\epsilon^i}\right|_{\epsilon=0} H(\epsilon, p(\epsilon)) \tag{4.156}$$

invariant. Again the reduced space for this action is the jet space of the reduced space for the action of G on M.

Chapter 5: Kruskal's Theory of Nearly Periodic Systems

"All nature and the graceful sky are symbolized in the art of Geometria."— Kepler, *Tertius Interveniens*

5.1. Introduction: Kruskal's Approach

In 1962 Martin Kruskal published "Asymptotic Theory of Hamiltonian and other Systems with all Solutions Nearly Periodic" [Kruskal, 1962]. In this paper he generalized and unified previous results due to himself and others showing that many specific physical systems had adiabatic invariants to all orders in a small parameter expressing the separation of slow and fast time scales. This work is the theoretical foundation for many concepts in plasma physics and elsewhere. It is therefore of interest to re-examine the underlying structure of the theory in the light of recent developments in geometric mechanics and dynamical systems. We have presented in earlier chapters a geometric formulation for ordinary perturbation theory and showed that it leads to deeper insights into the Hamiltonian nature of

perturbed systems. Here we will present a new version of Kruskal's result which leads to a completely unambiguous prescription for the perturbation analysis. We then show how the new procedure is expressed in coordinate-free language using the geometric structure of ordinary perturbation theory. The procedure is intimately tied with the process of reduction in the Hamiltonian case and so is connected with recent developments in finding Hamiltonian structures in plasma and other physical contexts.

Let us begin by setting up the problem in geometric language and then discussing Kruskal's method in this framework. We are interested in studying dynamical systems whose dynamical vector field depends on a small parameter ϵ in such a way that when $\epsilon = 0$ all orbits are periodic and the period is a smooth non-vanishing function on state space. Kruskal called the closed unperturbed orbits *loops*. These loops naturally give the state space the structure of a circle bundle. Locally we may express this bundle as a product: $S^1 \times \Re^n$. It is easy to see that we may choose coordinates θ and x^a, $a = 1, \ldots, n$ such that the unperturbed dynamical vector field is independent of θ and has no x component. In these coordinates the dynamics is given by the vector field

$$X = \psi(x)\frac{\partial}{\partial\theta} + \epsilon X_1 + \frac{\epsilon^2}{2}X_2 + \ldots, \tag{5.1}$$

where $\psi(x)$ describes how the frequency varies with x. As we turn on the perturbation by letting ϵ be nonzero, the dynamical vector field no longer points along loops and the orbits in state space drift from loop to loop along a helical path. If we are uninterested in the dynamics around the loops then the unperturbed system projects to trivial dynamics on the base of the bundle, i.e. the x coordinate

does not evolve. If the perturbed vector field did not depend on θ, then again by ignoring the fast θ motion we could project the dynamics to the base space leaving only the interesting slow dynamics. Unfortunately, typical perturbations will not be independent of θ and different points on a loop will evolve to different loops, precluding any well defined slow dynamics on the base. Kruskal attempts to find ϵ-dependent coordinates $\tilde{\theta}$ and $\tilde{x}^a$ which reduce to θ and x^a when ϵ vanishes, such that the dynamical vector field expressed in these coordinates is independent of $\tilde{\theta}$. In general this will not be possible for finite ϵ, but Kruskal was able to show that such $\tilde{\theta}$ and $\tilde{x}^a$ exist as asymptotic series to all orders in ϵ. His technique involves an intricate "bootstrap" argument which links two expansions together and obtains terms in one from lower order terms in the other.

The choice of coordinates is not unique because we can always apply an ϵ dependent diffeomorphism to the base and rotate the fibers (i.e. choose coordinates $\tilde{y}(\tilde{x})$ and $\tilde{\theta} + f(\tilde{x})$) without altering the desired properties. Kruskal called the ϵ dependent loop obtained by holding $\tilde{x}$ fixed and letting $\tilde{\theta}$ run from 0 to 2π, a *ring*. The set of rings determines a fibration of phase space for each ϵ and reduces to the original fibration when $\epsilon = 0$. The vector field

$$R \equiv \frac{\partial}{\partial \tilde{\theta}} \tag{5.2}$$

is tangent to the rings and was denoted the *roto-rate* by Kruskal. He showed that as an asymptotic series, R is uniquely defined to all orders in ϵ. In the paper he makes the interesting comment: "It does not appear obvious whether an explicit recursion formula to determine R in terms of f [the dynamical vector field] can be found. If so, the whole theory of this paper might be simplified and rendered less deep." In

this chapter we will exhibit such a formula and show its geometric significance.

5.2. The New Approach Expressed in Coordinates

Many perturbation methods, beginning with the Poincaré-Lindstedt method and leading up to the method of Lie transforms, are like Kruskal's approach in that they work by looking for an ϵ-dependent diffeomorphism of phase space which takes the perturbed system into a simpler one (equivalently one thinks of finding new coordinates in terms of which the dynamics looks simple). This has the advantage that closed loops are automatically taken to closed loops and if the system is Hamiltonian and the diffeomorphism is a canonical transformation, it preserves the Hamiltonian structure. In the present setup we will see that it is easier to explicitly require the rings to be closed as a constraint that helps determine terms in the expansion. In Kruskal's technique a similar constraint is required to ensure that the change of coordinates is periodic in θ. One advantage of the present approach over methods which change coordinates is the uniqueness of the desired expansion. This leads one to suspect (correctly as we will show in section 6.3) that the whole procedure has an intrinsic coordinate-free interpretation. Because the coordinate change is not uniquely specified in the standard approach, there will always be an arbitrary and unphysical choice to be made at some point. This becomes especially critical in infinite-dimensions where one wishes to apply these techniques to fields (we will see in chapter 13 that infinite-dimensional coordinate changes may involve many subtle phenomena).

We will now see that an asymptotic expansion for R can be uniquely determined on the basis of two constraints. The coordinate-free way of expressing the dynamical

vector field X's independence of $\bar{\theta}$ is to say that the Lie bracket of X and R vanishes:

$$[R, X] = 0. \tag{5.3}$$

The second requirement on R is that its integral curves all be closed. Thus

$$\text{if } \dot{y}(t) = R \text{ then } y(0) = y(2\pi). \tag{5.4}$$

If we assume an expansion for R and X as above then equation (5.3) taken order by order in ϵ gives us a hierarchy of equations for the terms in R:

$$\begin{aligned} [R_0, X_0] &= 0 \\ [R_1, X_0] &= -[R_0, X_1] \\ [R_2, X_0] &= -2[R_1, X_1] - [R_0, X_2] \\ &\cdots \end{aligned} \tag{5.5}$$

Notice that each right hand side is known from before as we determine successive terms in R. On the left hand side we always find the bracket of a term in R with $X_0 = v(x)\partial/\partial\theta$. The x components and the θ components of this bracket have a different structure and must be dealt with separately.

We use y^i, $1 \le i \le n+1$ to represent x^a, θ, $1 \le a \le n$ together. Recall that the coordinate expression for the Lie bracket of two vector fields A and B takes the form

$$[A, B]^i = \sum_{j=1}^{n+1} \left(A^j \frac{\partial}{\partial y^j} B^i - B^j \frac{\partial}{\partial y^j} A^i \right). \tag{5.6}$$

X_0 is special in that it has only a θ component, which depends only on x. Denoting the x components of a vector by an x superscript and the θ component by a θ superscript, we find

$$[R_i, X_0]^x = -v(x) \frac{\partial R_i^x}{\partial \theta}. \tag{5.7}$$

Since $\psi(x)$ doesn't vanish by assumption, and assuming we have already obtained all lower order terms, we may integrate this equation and find the solution up to an arbitrary n-component function $F(x)$:

$$R_i^x = -\frac{1}{\psi(x)} \int_0^\theta (\text{known terms}) d\theta + F(x) \tag{5.8}$$

The θ component of the Lie bracket is slightly more complicated:

$$[R_i, X_0]^\theta = -\psi(x) \frac{\partial R_i^\theta}{\partial \theta} + \frac{\partial \psi}{\partial \theta} R_i^\theta + \sum_{a=1}^{n} \frac{\partial \psi}{\partial x^a} R_i^a. \tag{5.9}$$

The second term vanishes because ψ has no θ dependence. If we are able to determine R_i^x first then we may obtain R_i^θ by integrating up to an arbitrary function $G(x)$:

$$R_i^\theta = -\frac{1}{\psi(x)} \int_0^\theta \left(-\sum_{a=1}^{n} \frac{\partial \psi}{\partial x^a} R_i^a + (\text{known terms}) \right) d\theta + G(x). \tag{5.10}$$

To determine $F(x)$ and $G(x)$ we employ the periodicity condition order by order. We must find $F(x)$ first without the help of R_i^θ so that we may use R_i^x in its determination as above. The flow of $\dot{y} = R$ satisfies the integral equation

$$y(t) = \int_0^t R(y(t'))dt'. \tag{5.11}$$

Let us expand $y(t)$ as an asymptotic series in ϵ and substitute this into this equation:

$$y_0(t) + \epsilon y_1(t) + \ldots = \int_0^t \left\{ R_0(y_0(t')) + \epsilon \left(\frac{\partial R_0}{\partial y_0} \cdot y_1 + R_1(y_0) \right) + \epsilon^2 \ldots \right\} dt'. \tag{5.12}$$

We again get a hierarchy of equations by collecting terms order by order in ϵ. For reference, let us work out the integrand to order ϵ^3. We first expand the terms in

R's asymptotic expansion in Taylor series to the needed order:

$$\begin{aligned} R(y) \sim\ & R_0(y_0) + \frac{\partial R_0}{\partial y_0}\cdot(\epsilon y_1 + \frac{\epsilon^2}{2}y_2 + \frac{\epsilon^3}{6}y_3) \\ & + \frac{1}{2}\frac{\partial^2 R_0}{\partial y_0^2}(\epsilon y_1 + \frac{\epsilon^2}{2}y_2)(\epsilon y_1 + \frac{\epsilon^2}{2}y_2) + \frac{1}{6}\frac{\partial^3 R_0}{\partial y_0^3}(\epsilon y_1)^3 \\ & + \epsilon(R_1(y_0) + \frac{\partial R_1}{\partial y_0}\cdot(\epsilon y_1 + \frac{\epsilon^2}{2}y_2) + \frac{1}{2}\frac{\partial^2 R_1}{\partial y_0^2}(\epsilon y_1)(\epsilon y_1)) \\ & + \frac{\epsilon^2}{2}(R_2(y_0) + \frac{\partial R_2}{\partial y_0}\cdot\epsilon y_1) \\ & + \frac{\epsilon^3}{6}R_3(y_0). \end{aligned} \tag{5.13}$$

Now let us collect terms to get the vector field to the various orders. To order ϵ^0 we have simply

$$R_0. \tag{5.14}$$

To order ϵ^1 we have

$$\frac{\partial R_0}{\partial y_0}\cdot y_1 + R_1. \tag{5.15}$$

To order ϵ^2 we have

$$\frac{1}{2}\frac{\partial R_0}{\partial y_0}\cdot y_2 + \frac{1}{2}\frac{\partial^2 R_0}{\partial y_0^2}\cdot y_1\cdot y_1 + \frac{\partial R_1}{\partial y_0}y_1 + \frac{1}{2}R_2. \tag{5.16}$$

Finally to order ϵ^3 we have

$$\begin{aligned} & \frac{1}{6}\frac{\partial R_0}{\partial y_0}\cdot y_3 + \frac{1}{2}\frac{\partial^2 R_0}{\partial y_0^2}\cdot y_1\cdot y_2 + \frac{1}{6}\frac{\partial^3 R_0}{\partial y_0^3}\cdot y_1^3 \\ & + \frac{1}{2}\frac{\partial R_1}{\partial y_0}\cdot y_2 + \frac{1}{2}\frac{\partial^2 R_1}{\partial y_0^2}\cdot y_1^2 + \frac{1}{2}\frac{\partial R_2}{\partial y_0}y_1 + \frac{1}{6}R_3. \end{aligned} \tag{5.17}$$

In the expression for $y_i(t)$, the term $y_i(t)$ appears only in the form $y_i\cdot(\partial R_0/\partial y_0)$ and all other terms are of lower order and therefore known. But $R_0 = (\partial/\partial\theta)$ is constant and so its derivative vanishes. Thus each y_i is a well defined integral over

known quantities:

$$\begin{aligned} \theta_0(t) &= \theta_0(t=0) + t \quad x_0(t) = x_0(t=0) \\ y_1 &= \int_0^t R_1(x_0,\theta)d\theta \\ y_2 &= \int_0^t 2\left(\frac{\partial R_1}{\partial y_0}\cdot y_1 + \frac{1}{2}R_2 + \frac{1}{2}\frac{\partial^2 R_0}{\partial y_0^2}\cdot y_1 \cdot y_1\right)d\theta \\ &\vdots \end{aligned} \tag{5.18}$$

We may now impose the constraint that the orbits be periodic:

$$\begin{aligned} y_1(2\pi) - y_1(0) = 0 &= \int_0^{2\pi} R_1\, d\theta \\ \int_0^{2\pi} R_2\, d\theta &= -\int_0^{2\pi} \frac{\partial R_1}{\partial y_0}\cdot y_1\, d\theta \\ &= -\int_0^{2\pi} \frac{\partial R_1}{\partial y}\cdot\left(\int_0^{\theta} R_1 d\theta'\right)d\theta \\ &\vdots \end{aligned} \tag{5.19}$$

The elementary asymptotic series for y is valid since we need it only for finite time (in fact, only time one). Each of these equations has an x component and a θ component. The asymptotic series for R is uniquely determined as follows: i) $R_0^\theta = 1$ and $R_0^x = 0$, ii) assuming R_j known for $j < i$, we obtain R_i^x up to the function $F(x)$ by means of equation (6.8), iii) we determine $F(x)$ by means of the x component of equation (6.14), iv) using the R_j's and R_i^x we determine R_i^θ up to the function $G(x)$ by means of equation (6.10), v) and finally we determine $G(x)$ by using the θ component of equation (6.14) giving us the entire R_i and allowing us to continue the iteration to R_{i+1}.

5.3. The Geometric Version of the New Approach

Let us now formulate the new procedure completely invariantly, in terms of the geometric picture introduced in chapter 4. Given a vector field $X(\epsilon)$ on $I \times M$ with no $\partial/\partial\epsilon$ component and $X(0)$ having all periodic orbits, we would like to find $R(\epsilon)$ on $I \times M$ with no $\partial/\partial\epsilon$ component such that $X(0)$ and $R(0)$ are tangent, $[X(\epsilon), R(\epsilon)] = 0$, and all of $R(\epsilon)$'s orbits are periodic of period 2π. We cannot do this for finite ϵ (due to homoclinic behavior, for example, which is discussed in chapter 14), but R is determined uniquely to all orders in ϵ.

Let $\psi : M \to \Re$ be the period function of $X(0)$ and assume that it does not vanish anywhere. Let $R(0) = X(0)/\psi$. We have the hierarchy of jets of paths at $\epsilon = 0$:

$$\infty M \to \ldots \to JM \to \ldots \to 2M \to 1M \to M. \tag{5.20}$$

$X(\epsilon)$ determines vector fields on each of these spaces. We are looking for R's on each space such that each projects to the one below. The bracket of two vector fields lifts to a vector field on JM which is the bracket on JM of the lifts to JM of the vector fields. Thus $[X(\epsilon), R(\epsilon)] = 0$ lifts to a condition on each JM. The flow of a vector field for time 1 is a diffeomorphism of $I \times M$ preserving ϵ. It therefore lifts to diffeomorphisms of each JM which project into one another. On JM the diffeomorphism is the time-1 flow of the lift of the vector field to JM. The condition $\dot{y} = R(y)$ implies $y(1) = y(0)$ lifts to similar conditions on the lifts to each JM.

We know from the coordinate calculations that to determine a given order of R, we first have to determine it modulo R_0 and then use this to get the full R. Thus we introduce the drift tangent bundles $DTJM$ which at each point of JM has fiber

$T_pJM/\tilde{R}_0(p)$, where $\tilde{R}_0$ is the lift to JM of $\epsilon^J R_0$. We get the natural projections

$$\begin{array}{ccc} TJM & \to & DTJM \\ \downarrow & & \downarrow \\ \vdots & & \vdots \\ T2M & \to & DT2M \\ \downarrow & & \downarrow \\ T1M & \to & DT1M \\ \downarrow & & \downarrow \\ TM & \to & DTM. \end{array} \tag{5.21}$$

The key relation is

$$[R_j, x_0] = [R_j, \psi R_0] = (R_j \cdot \psi)R_0 + \psi[R_j, R_0]. \tag{5.22}$$

The first term is along R_0 and so is killed by the D operation, the second may be solved for R_j by integrating along the flow of R_0 the quantity $1/\psi$ times the right hand side. The integral relation on D space then gives the constant of integration. Next, since $R_0 \cdot \psi = 0$, we see that

$$(R_j + fR_0) \cdot \psi = R_j \cdot \psi. \tag{5.23}$$

So the first term depends only on DR_j. So the R_0 component of R_j may be found by integrating: $\frac{1}{\psi}(-(DR_j \cdot \omega)R_0 + r.h.s.)$ and then the integral gives the constant. We will see several examples of this procedure in the next chapter.

5.4. Extensions and Limitations of Kruskal's Theory

In this section let us describe some of the limitations and extensions of Kruskal's and other secular perturbation theories. Most of the asymptotic theories of physics are based upon asymptotic rather than convergent expansions. An asymptotic expansion of a function of ϵ is a formal power series in ϵ such that the truncation after the Nth term approximates the function to order ϵ^N. Typically, as N gets larger, one must go to smaller ϵ to get a good approximation and there may be no ϵ for which all the terms in an asymptotic series are helpful (and in fact most such series are divergent). Most of the techniques one applies in perturbative analyses yield asymptotic series because the exact solution we are approximating does not have a convergent expansion. The notion of convergence rests on a complex analytic structure, while smooth non-analytic coordinate changes, which should be physically irrelevant, can destroy the convergence of an asymptotic expression. There are two limiting processes here, letting ϵ go to zero and letting the number of terms go to infinity. Asymptotic series have nice limits when they are done in that order; convergent series allow us to switch the order of the limiting processes. From a practical point of view, one never uses more than a finite number of terms of an expansion in any case and often there are asymptotic series which approximate an expression much more quickly than a corresponding convergent series.

5.4.1. Secular Perturbation Theory

If the coefficients of an asymptotic series are in addition a function of other parameters x, then one may inquire into the uniformity of the asymptotic approximation. We say that a series is asymptotic uniformly in x if the coefficient of ϵ^m in the error term of a truncation may be chosen to be independent of x for each m. How small we have to make ϵ to get a given approximation should be specifiable independently of x. Most of the asymptotic expansions in physics are nonuniform, otherwise known as singular or secular. If the nonuniformity occurs near a finite point, we may often insert a *boundary layer* expansion defined on an ϵ dependent region about the singular x which is uniform over that region, and patch it to the singular expansion using the method of matched asymptotic expansions (see [Nayfeh, 1973] p. 111). Such a technique is used for example in matching a thin boundary layer where viscosity is important in an almost inviscid fluid to the inviscid solution in the interior.

The most common occurrence of nonuniformity, however, takes place as x goes to infinity (again note that the order of the x limit and the ϵ limit may not be exchanged). A key example of this is where x represents the time in the solution of ordinary differential equations. We have seen that the most pedestrian perturbation technique is to simply expand the supposed solution of an O.D.E. in an asymptotic series, plug this into the equation, and solve for the terms order by order. We saw that this technique gave asymptotic solutions over times independent of ϵ to all orders in ϵ. For even the simplest problems with recurrent behavior, however, these expansions are non-uniform in time and in fact are no longer asymptotic

expansions at all when written in terms of a "slow" time: $\tau \equiv \epsilon t$. We have not made any mistakes in our analysis of the equation as is sometimes suggested; rather this nonuniform behavior is a property of the exact solution.

Consider for example the simple equation

$$\ddot{x} = -(1+\epsilon)^2 x. \tag{5.24}$$

This is just a harmonic oscillator of frequency $(1+\epsilon)$ and so has solutions like

$$x(t) = \sin(1+\epsilon)t. \tag{5.25}$$

If we expand this in an asymptotic expansion, we obtain

$$x(t) \sim \sin(t) + \epsilon t \cos(t) + \cdots. \tag{5.26}$$

This is the asymptotic expansion of the true solution, and yet we see that for times of order $1/\epsilon$, the second term does not go to zero as ϵ does. This is thus non-uniform in t and the non-uniformity is on a scale of $t = 1/\epsilon$. For this example we may make a change of independent variable to $s = (1+\epsilon)t$ and we see that we get a completely uniform (and even convergent) expansion in terms of s. The effect of this is to do our asymptotics along the lines s =constant in the (ϵ, t) plane instead of along t =constant.

5.4.1.1. Lie Transforms

The general technique of all secular perturbation theories is to introduce a new asymptotic process as in this example. For example, the method of Lie transforms makes an ϵ dependent canonical transformation of an underlying phase space, such that in terms of the new variables the standard perturbation method yields solutions that are asymptotic for longer than bounded times. This technique may be shown to work for nearly periodic dynamical systems for time $1/\epsilon$ (see for example [Cary, 1981] and [Nayfeh, 1973] p. 200). Even here it is not completely specified what Lie transform one should make (since there is always extra freedom in making changes of coordinates). There is no general theory describing other circumstances in which the method works and even when the method does work, there is no algorithmic procedure for carrying it out.

5.4.1.2. Two-Timing

The multiple time scale approach (two-timing) writes the solution in terms of two variables t and τ (see [Nayfeh, 1973] p. 228). The expression in terms of t is obtained by replacing τ by ϵt, but the limiting process is performed with both t and τ held fixed. We choose the representation so that the dependence on t is non-secular on times of order $1/\epsilon$ by writing the secular parts in terms of τ. The result is asymptotics good for times of order $1/\epsilon$. This procedure is not given algorithmically, it is not clear what systems this approach works for and it is not clear how to proceed order by order even on systems where it works (though in specific examples the technique is often quite useful). One may sometimes get expressions for longer

times by introducing more slow times: $t, \tau = 1/\epsilon, \tau' = 1/\epsilon^2, \ldots$. but again the theory seems to be restricted to special examples. It appears that when we express quantities as functions of more than one variable that we introduce an essential non-uniqueness which may limit the precision with which one can specify what one is doing.

5.4.2. $1/\epsilon$ Time of Validity for Kruskal's Technique

We have seen that Kruskal's method, while restricted to perturbations of an exactly periodic system, is perfectly general within this context and uniform to all orders in ϵ for times of order $1/\epsilon$. That the time of validity is $1/\epsilon$ is a significant point, clearly stated in Kruskal's original paper, but not often repeated when this paper is quoted. One often hears the phrase: "Kruskal showed that adiabatic invariants exist to all orders in ϵ." This then leads one to confuse the accuracy of the approximation with the *time* of validity.

Why is the theory good for even time $1/\epsilon$ (we have seen that the simplest examples make the basic perturbation technique fail on this scale)? The idea is to convert our system, by hook or by crook, to one of the form

$$\dot{x} = \epsilon X, \tag{5.27}$$

where X is an asymptotic vector field. In this case we may make the change of variables to $\tau = \epsilon t$ and obtain the equation in the usual form:

$$\frac{dx}{d\tau} = X. \tag{5.28}$$

We may apply standard perturbation theory to this to get an expression for the solution that is asymptotic for bounded τ. But going back to t, this is valid for t of time $1/\epsilon$. In Kruskal's technique, we get rid of the fast oscillations by successive coordinate changes and so the resulting dynamics has only the slow drifts which are of order ϵ and so susceptible to the above technique. There is no way to get longer times out of this technique in general.

One sometimes hears the plausible argument: "Since the adiabatic invariant's time derivative is zero to all orders in ϵ, it will be constant for exponentially long times." As an example to see that this reasoning is faulty, consider the expression

$$\dot{J} = e^{-1/\epsilon t}. \tag{5.29}$$

If we take derivatives of this with respect to ϵ, while holding t fixed, we see that J's time derivative is indeed zero to all orders in ϵ. Nonetheless, $\dot{J}$ is of order 1 on times of order $1/\epsilon$. Since when $\dot{J}$ is of order 1, J can change in times of order 1, we see that for this example J can undergo an order 1 change in times of order $1/\epsilon$.

To see that there are physical examples that fit into Kruskal's (and so everbody else's) framework and yet do not have preserved adiabatic invariants on times of order $1/\epsilon^2$ we need only consider parametric resonance. As in [Arnold, 1983], the simple harmonic oscillator with slowly varying frequency:

$$\ddot{x} = -\omega^2(1 + a\cos(\epsilon t))x, \tag{5.30}$$

may be shown to be unstable for arbitrarily small ϵ since there are resonances of arbitrarily high order. (An example is the child who pumps up a swing by

resonantly varying its length, and so its natural frequency.) It is easy to see that in this circumstance, the action will change significantly in times of order $1/\epsilon^2$.

5.4.3. Averaging with Multiple Frequencies

If, instead of a single fast degree of freedom, we have several fast degrees of freedom, the motion is close to being on tori that fill up the phase space. It appears at first that one could just average over the tori and obtain as adiabatic invariants the actions corresponding to the fundamental loops of the tori. This is fine if the unperturbed dynamics covers each torus ergodically and so the average perturbation over a torus really reflects the time average. If any of the frequencies are rationally related, however, then the orbit covers only a piece of the torus and we have no reason to expect that the average over the whole torus should have anything to do with the average over the orbit. Indeed one finds that one may have "trapping" at these so-called resonances, where the adiabatic invariant changes drastically in time $1/\epsilon$. Fortunately, for generic Hamiltonians, the measure of the trapped regions is small and goes to zero as $\sqrt{\epsilon}$. [Arnold, 1983] introduces the notion of an *almost adiabatic invariant* which is a quantity that is constant to first order in ϵ for times of order $1/\epsilon$ except for a set of phase space whose measure goes to zero with ϵ.

5.4.4. Averaging Over Ergodic Orbits

When the unperturbed system is ergodic on its energy surfaces, [Kubo et al., 1965], and later [Ott, 1979], showed that the volume contained inside an energy surface is preserved to order ϵ for time $1/\epsilon$. This is based on an averaging similar to the single frequency averaging case. This is of interest because as the dimension gets larger, the volume being preserved is equivalent to the entropy being preserved. This gives a mechanical justification for the adiabatic invariance of the entropy in thermodynamics. When we forget about the oscillatory directions, we have seen that the constancy of the adiabatic invariant in an oscillatory system forces it to give and take energy under a variation of its parameters in such a way that it simulates a potential, leading to the pseudoforces we have discussed earlier. The same circumstance in the thermodynamic case leads to thermodynamic forces. We shall see in the last chapter of this thesis that there is a symplectic structure and Lagrangian submanifold that describes these forces in a Hamiltonian way exactly analogous to the pseudo forces of the mechanical systems.

5.4.5. Non-uniqueness of Symmetry for Finite Perturbation

Even though Kruskal showed that in a nearly periodic system there is a unique circle action on the perturbed dynamics to all orders in ϵ, for *finite* perturbation there may be none or the circle action may not be unique. Here is give a simple example to demonstrate this non-uniqueness. Our system will be a cylinder with coordinates θ and y. The unperturbed system will just rotate in θ:

$$X_0 = \frac{\partial}{\partial \theta}. \tag{5.31}$$

A typical perturbation shifts the orbits on the cylinder so that almost all of them are helices. For example, we get orbits that look like the stripes on a barber pole for finite ϵ with

$$X = \frac{\partial}{\partial\theta} + \epsilon\frac{\partial}{\partial y}. \tag{5.32}$$

Since this vector field does not depend explicitly on θ, the unperturbed vector field generates a circle action which is a symmetry of this one. There are an infinite number of others, however.

Let R be a vector field on the cylinder, all of whose orbits are circles. By the relationship of the flow of a vector field to the Lie derivative along it, R will generate a circle symmetry of X if, and only if the Lie derivative of X by R vanishes. But for vector fields this is just the Lie bracket: $[R, X]$, which is antisymmetric– and so the Lie derivative of R along X must vanish as well. This says that if we know R at any place on an X orbit, for it to be a symmetry its value anywhere else on this orbit must be obtained by pushing it forward along the flow of X. The orbits of X are a circle of helices filling up phase space. If we choose any closed loop around the cylinder that intersects each helix exactly once (it is easy to see that there are an infinite number of ways of doing this), then we can make that loop one of the orbits of R. For R on the loop, just choose R tangent to the loop, and to get R everywhere else, we push this forward by X's flow.

Chapter 6: Ponderomotive Force and Gyromotion

"The purpose of computing is insight, not numbers."—Richard W. Hamming

6.1. Ponderomotive-like Forces

In this chapter we will apply the perturbation techniques of the last chapter to some example problems. The physical plasma phenomena that underlie our examples are the ponderomotive force and gyromotion. We wish to keep the physics of these asymptotic physical effects in the foreground and so we do not work with the most general situation in which these effects arise. Instead, we focus on the simplest situations which contain the relevant physical effects, and discuss the underlying asymptotic process in detail. This approach allows us to explicitly compare aspects of these problems which are obscure in more complex settings. The insights and intuitions gained are of course applicable to more general settings.

First we discuss a model problem that contains the essence of ponderomotive force. We consider a time-independent problem with two degrees of freedom. One degree of freedom behaves like a fast oscillator and the other evolves slowly in an

arbitrary potential. The example we will use is

$$\begin{aligned} \dot{x} &= v_x \\ \dot{y} &= v_y \\ \dot{v}_x &= -\frac{\partial V(x)}{\partial x} - \frac{1}{2}\frac{\partial W(x)}{\partial x}y^2 \\ \dot{v}_y &= -W(x)y \end{aligned} \tag{6.1}$$

with the initial conditions

$$\begin{aligned} x(0) &= x_0 \quad y(0) = y_0 \\ v_x(0) &= v_{x0} \quad v_y(0) = v_{y0}. \end{aligned} \tag{6.2}$$

This system is Hamiltonian with the canonical Poisson bracket in x, y, v_x, v_y and the Hamiltonian

$$H(x, y, v_x, v_y) = \frac{1}{2}v_x^2 + \frac{1}{2}v_y^2 + V(x) + \frac{1}{2}W(x)y^2. \tag{6.3}$$

A physical model with this dynamics is a ball rolling in a trough. The trough is a surface whose height over the (x, y) plane is given by $V(x) + W(x)y^2/2$. (v_x, v_y) represent the x and y velocities respectively. Along each line x =constant, the potential has a minimum at $y = 0$ and grows quadratically with strength $W(x)$ as $|y|$ increases. If the rolling particle starts at $y = 0$ with $v_y = 0$, then it remains on the line $y = 0$ for all time. In this case the system behaves like a one degree of freedom system with potential $V(x)$ and $W(x)$ plays no role. If there is any displacement from $y = 0$, then the particle continues to oscillate back and forth about the x-axis. The y^2 in the potential acts like a restoring force toward $y = 0$ and the strength $W(x)$ varies with x. Exactly this kind of model is behind mechanical systems with constraints. When we say we have a particle in the plane constrained

to the x-axis (like a bead on a wire), we really mean that there is a stiff restoring force transverse to the axis. It is sometimes implicitly assumed that if the restoring force is sufficiently strong, the motion of the particle is well approximated by the equations restricted to the constraint manifold (in this case the x-axis). To make this statement precise one must introduce asymptotics in the strength of the restoring force. We will see in fact that if $W(x)$ is not constant, then the transverse motions add a new pseudo-potential to the constrained motion.

We are interested, then, in the case where the transverse y motion is very fast compared to the x motion. We want to explicitly introduce asymptotic scaling with a parameter ϵ into our equations of motion to make this analysis precise. Usually one is somewhat sloppy and simply proceeds intuitively. Sometimes this gives a useful answer but the chances of this go down as the complexity of the problem goes up. It is also important to explicitly state the intended scaling as this represents the physical effect one is trying to study. Let us give an intuitive argument and then see what must be done to make it precise.

6.1.1. Intuitive Treatment

Because the y motion is supposed to be fast compared to the x motion, the y motion initially behaves as if x is frozen at x_0. The equations for the y motion are now those of a simple harmonic oscillator with frequency $\sqrt{W'(x_0)}$:

$$\dot{y} = v_y \qquad \dot{v}_y = -W'(x_0)y. \tag{6.4}$$

For simplicity, we take the y initial condition to be zero: $y(0) = 0$. In this case the solution is

$$y(t) = \frac{v_{y0}}{\sqrt{W(x_0)}} \sin(\sqrt{W(x_0)}t). \tag{6.5}$$

The v_x evolution equation

$$\dot{v}_x = -\frac{\partial V(x)}{\partial x} - \frac{1}{2}\frac{\partial W(x)}{\partial x}y^2 \tag{6.6}$$

depends on y. Because y traverses its periodic orbit many times before x or v_x evolve appreciably, y's effect on v_x's evolution will be almost the same as its average over y's orbit. Because the average of $\sin^2(\theta)$ over $0 \leq \theta \leq 2\pi$ is $1/2$, the average of $-\frac{\partial W(x)}{\partial x}y^2/2$ will be

$$-\frac{1}{2}\frac{\partial W(x)}{\partial x}\frac{v_{y0}^2}{2W(x_0)} \tag{6.7}$$

If we now let u be the value of v_y when $y = 0$, we see that the same argument applies to any time (not just the initial time) to give the v_x equation:

$$\dot{v}_x = -\frac{\partial V(x)}{\partial x} - \frac{1}{2}\frac{\partial W(x)}{\partial x}\frac{u^2}{2W(x)}. \tag{6.8}$$

This tells us how v_x varies if we know u (i.e. y's maximum velocity).

We may determine u from the constraint that the total energy must remain constant. At an arbitrary time, the average energy in the y motion is $u^2/2$, a result we obtain by holding x and v_x fixed and realizing that all of the oscillatory energy is kinetic when $y = 0$. The total energy (averaged over the y motion) is then

$$H = \frac{1}{2}v_x^2 + \frac{1}{2}u^2 + V(x). \tag{6.9}$$

The time derivative of H must vanish, so

$$\dot{H} = v_x\dot{v}_x + u\dot{u} + v_x\frac{\partial V}{\partial x} = 0. \tag{6.10}$$

We substitute in the expression for $\dot{v}_x$ to obtain

$$\begin{aligned}\dot{H} &= 0 \\ &= -v_x\frac{\partial V}{\partial x} - \frac{1}{2}v_x\frac{\partial W}{\partial x}\frac{u^2}{2W} + u\dot{u} + v_x\frac{\partial V}{\partial x} \\ &= u\dot{u} - \frac{1}{2}\dot{W}\frac{u^2}{2W}.\end{aligned} \tag{6.11}$$

This has an integrating factor of $1/(2\sqrt{W})$ by which we multiply to obtain:

$$\begin{aligned}0 &= \frac{u\dot{u}}{2\sqrt{W}} - \dot{W}\frac{u^2}{8W^{3/2}} \\ &= \frac{d}{dt}\left(\frac{u^2}{4\sqrt{W}}\right).\end{aligned} \tag{6.12}$$

We have found a constant of the motion, a multiple of 2 of which we shall call J (since it is equal to the action of the transverse oscillations):

$$J = \left(\frac{u^2}{2\sqrt{W(x)}}\right) = \left(\frac{v_{y0}^2}{2\sqrt{W(x_0)}}\right). \tag{6.13}$$

This may be solved for u:

$$u = \sqrt{2J\sqrt{W}} \tag{6.14}$$

and then substituted into the equation for v_x:

$$\begin{aligned}\dot{v}_x &= -\frac{\partial V(x)}{\partial x} - \frac{1}{2}\frac{\partial W(x)}{\partial x}\frac{u^2}{2W(x)} \\ &= -\frac{\partial V(x)}{\partial x} - \frac{1}{2}\frac{\partial W(x)}{\partial x}\frac{J}{\sqrt{W}} \\ &= -\frac{\partial V(x)}{\partial x} - \frac{\partial}{\partial x}(J\sqrt{W(x)}).\end{aligned} \tag{6.15}$$

The x motion behaves just like a particle in the potential

$$V(x) + J\sqrt{W(x)} = V(x) + \frac{v_{y0}^2\sqrt{W(x)}}{2\sqrt{W(x_0)}}. \tag{6.16}$$

The ordinary potential $V(x)$ is augmented by an extra pseudo-potential $J\sqrt{W(x)}$ which is non-zero when there is transverse motion and has an effect on the x motion when W varies with x.

6.1.2. Introducing the Asymptotic Scaling

This argument is intuitively appealing but has not been systematic. It is difficult to see how to get more accurate evolution equations. We therefore wish to introduce an explicit ϵ into our equations which captures the physical assumptions used in the intuitive discussion above as $\epsilon \to 0$ and which gives the correct equations for $\epsilon = 1$. y and v_y are supposed to be fast variables. If we view them as varying by order 1 on a time scale of ϵ (so the fast motion gets faster and faster as $\epsilon \to 0$), then x and v_x should vary by order 1 on times scales of order 1. In modelling constrained motion, one typically makes the constraining force stronger asymptotically. We therefore replace $W(x)$ with $W(x)/\epsilon^2$. If we keep the initial conditions $y(0) = y_0$ and $\dot{y}(0) = v_{y0}$, then as $\epsilon \to 0$ the energy of the transverse motion becomes infinite. Because the transverse energy is really of the same order as the energy of the constrained motion, we must scale the transverse displacement with ϵ as $y(0) = \epsilon y_0$. This leads to a system given by

$$\begin{aligned} \dot{x} &= v_x \\ \dot{y} &= v_y \\ \dot{v}_x &= -\frac{\partial V(x)}{\partial x} - \frac{1}{2\epsilon^2}\frac{\partial W(x)}{\partial x}y^2 \\ \dot{v}_y &= -\frac{1}{\epsilon^2}W(x)y \end{aligned} \tag{6.17}$$

with the initial conditions

$$\begin{aligned} x(0) &= x_0 \quad y(0) = \epsilon y_0 \\ v_x(0) &= v_{x0} \quad v_y(0) = v_{y0}. \end{aligned} \tag{6.18}$$

From the equations, it might appear that v_x can really evolve by order $1/\epsilon$ on times of order ϵ contrary to assumption. This doesn't happen, however, because

the maximal excursion in y scales as ϵ and so counteracts the $1/\epsilon^2$ in the equation for v_x. This is because the initial y has this scaling and because the total energy is of order 1. That the transverse potential energy $Wy^2/2\epsilon^2$ is of order 1 means that y's maximal excursion is of order ϵ. This argument shows that the time derivative of v_y is of order $1/\epsilon$. This system is Hamiltonian with the canonical Poisson bracket in x, y, v_x, v_y and the Hamiltonian is

$$H(x, y, v_x, v_y) = \frac{1}{2}v_x^2 + \frac{1}{2}v_y^2 + V(x) + \frac{1}{2\epsilon^2}W(x)y^2. \tag{6.19}$$

This scaling has the fast motion getting very fast as $\epsilon \to 0$ as is physically reasonable. The interesting pseudo-potential has its effects in time 1 (which is presumably the time scale used by the observer). Unfortunately, this scaling is not a perturbation around any well understood system for $\epsilon = 0$. As ϵ approaches zero, the equations of motion become singular. It is also not apparent how the operations we performed in our intuitive approach may be expressed systematically with this scaling (e.g., we held the x variables fixed while studying y–why could we do this?)

One way to resolve these difficulties is to redo our study on a stretched time scale (i.e. imagine the clocks in your sensory apparatus getting faster as $\epsilon \to 0$). We introduce $\tilde{t} \equiv t/\epsilon$. If we interpret dot to mean derivative with respect to $\tilde{t}$ (while keeping the interpretation of the velocities v_x and v_y as derivatives with respect to t), then the equations of motion become

$$\begin{aligned}
\dot{x} &= \epsilon v_x \\
\dot{y} &= \epsilon v_y \\
\dot{v}_x &= -\epsilon \frac{\partial V(x)}{\partial x} - \frac{1}{2\epsilon}\frac{\partial W(x)}{\partial x}y^2 \\
\dot{v}_y &= -\frac{1}{\epsilon}W(x)y
\end{aligned} \tag{6.20}$$

with the initial conditions

$$\begin{aligned} x(0) &= x_0 \quad y(0) = \epsilon y_0 \\ v_x(0) &= v_{x0} \quad v_y(0) = v_{y0}. \end{aligned} \tag{6.21}$$

To get rid of the apparently singular terms, we may introduce a rescaled y defined by $Y \equiv y/\epsilon$. The initial conditions are now ϵ independent and the equations are non-singular:

$$\begin{aligned} \dot{x} &= \epsilon v_x \\ \dot{Y} &= v_y \\ \dot{v}_x &= -\epsilon \frac{\partial V(x)}{\partial x} - \frac{\epsilon}{2} \frac{\partial W(x)}{\partial x} Y^2 \\ \dot{v}_y &= -W(x) Y \end{aligned} \tag{6.22}$$

with the initial conditions

$$\begin{aligned} x(0) &= x_0 \quad Y(0) = y_0 \\ v_x(0) &= v_{x0} \quad v_y(0) = v_{y0}. \end{aligned} \tag{6.23}$$

These equations are perfectly set up for the non-Hamiltonian versions of the methods given in the last chapter. They limit as $\epsilon \to 0$ on a system with only periodic orbits:

$$\dot{x} = 0, \quad \dot{Y} = v_y, \quad \dot{v}_x = 0, \quad \dot{v}_y = -W(x) Y. \tag{6.24}$$

We will study these equations momentarily. Unfortunately, they are not Hamiltonian with respect to the canonical Poisson bracket. They are Hamiltonian with respect to the bracket

$$\{f, g\} = \epsilon \frac{\partial f}{\partial x} \frac{\partial g}{\partial v_x} - \epsilon \frac{\partial f}{\partial v_x} \frac{\partial g}{\partial x} + \frac{\partial f}{\partial Y} \frac{\partial g}{\partial v_y} - \frac{\partial f}{\partial v_y} \frac{\partial g}{\partial Y} \tag{6.25}$$

with the Hamiltonian

$$H(x, Y, v_x, v_y) = \frac{1}{2} v_x^2 + \frac{1}{2} v_y^2 + V(x) + \frac{1}{2} W(x) Y^2. \tag{6.26}$$

This bracket becomes degenerate (and functions of x and v_x become Casimirs) when $\epsilon = 0$.

One way to preserve the Hamiltonian nature with respect to the canonical bracket is to scale the potentials $V(x)$ and $W(x)$ to vary more and more slowly, the strength of the restoring potential to grow and the contribution of v_x to the kinetic energy to vanish as $\epsilon \to 0$. This leads to the Hamiltonian

$$H(x, y, v_x, v_y) = \frac{1}{2}\epsilon v_x^2 + \frac{1}{2}v_y^2 + V(\epsilon x) + \frac{1}{2}W(\epsilon x)y^2. \tag{6.27}$$

The equations of motion are

$$\begin{aligned} \dot{x} &= \epsilon v_x \\ \dot{y} &= v_y \\ \dot{v}_x &= -\epsilon \frac{\partial V}{\partial x}(\epsilon x) - \frac{\epsilon}{2}\frac{\partial W}{\partial x}(\epsilon x)y^2 \\ \dot{v}_y &= -W(\epsilon x)y \end{aligned} \tag{6.28}$$

with the initial conditions

$$\begin{aligned} x(0) = x_0 \quad y(0) = y_0 \\ v_x(0) = v_{x0} \quad v_y(0) = v_{y0}. \end{aligned} \tag{6.29}$$

Again this has nice limiting behavior as $\epsilon \to 0$. Physically, the picture is that we are stretching the x coordinate, so that the trough becomes flatter and flatter in the x direction. If we hold v_x fixed, then it takes a longer and longer time to make W vary substantially.

6.1.3. Using the Kruskal-like Perturbation Method

Let us now obtain the adiabatically invariant action to first order in ϵ using the method outlined in chapter 5. The dynamical vector field is:

$$\begin{aligned} X &= X_0 + \epsilon X_1 \\ &= \left(v_y \frac{\partial}{\partial Y} - W(x) Y \frac{\partial}{\partial v_y} \right) \\ &+ \epsilon \left(v_x \frac{\partial}{\partial x} - \left(\frac{\partial V(x)}{\partial x} + \frac{1}{2} \frac{\partial W(x)}{\partial x} Y^2 \right) \frac{\partial}{\partial v_x} \right). \end{aligned} \tag{6.30}$$

The solution curves of the unperturbed piece X_0 are all closed:

$$\begin{aligned} x(t) &= x_0 \qquad v_x(t) = 0 \\ Y(t) &= A\cos(\sqrt{W(x)}t) \\ v_y(t) &= -A\sqrt{W(x)}\sin(\sqrt{W(x)}t). \end{aligned} \tag{6.31}$$

6.1.3.1. The Coordinates x, v_x, A, θ

Let us define the angle on these orbits as θ and introduce an amplitude A:

$$Y = A\cos\theta \qquad v_y = -A\sqrt{W(x)}\sin\theta. \tag{6.32}$$

Writing these relations in the other direction gives

$$\tan\theta = -\frac{v_y}{Y\sqrt{W(x)}} \qquad A^2 = Y^2 + \frac{1}{W(x)} v_y^2. \tag{6.33}$$

Comparing these expressions with the earlier discussion we expect the zero order action to be given by

$$J_0 = \frac{(A\sqrt{W(x)})^2}{2\sqrt{W(x)}} = \frac{1}{2} A^2 \sqrt{W(x)}. \tag{6.34}$$

We may take the time derivatives of these expressions to obtain the dynamical vector field in the new coordinates.

$$\begin{aligned}(1+\tan^2\theta)\dot{\theta} &= -\frac{\dot{v}_y}{Y\sqrt{W(x)}} + \frac{v_y\dot{Y}}{Y^2\sqrt{W(x)}} + \frac{v_y}{2YW(x)^{3/2}}\frac{\partial W}{\partial x}\dot{x} \\ &= \sqrt{W(x)} + \tan^2\theta\sqrt{W(x)} - \epsilon v_x\frac{\tan\theta}{2W(x)}\frac{\partial W}{\partial x}.\end{aligned} \tag{6.35}$$

And so

$$\dot{\theta} = \sqrt{W(x)} - \epsilon v_x \sin\theta\cos\theta\frac{1}{2W(x)}\frac{\partial W}{\partial x}. \tag{6.36}$$

Similarly for A:

$$\begin{aligned}2A\dot{A} &= 2Y\dot{Y} + 2\frac{v_y}{W(x)}\dot{v}_y - \frac{v_y^2}{W(x)^2}\frac{\partial W}{\partial x}\dot{x} \\ &= 2Yv_y - 2v_yY - \epsilon v_x\frac{v_y^2}{W(x)^2}\frac{\partial W}{\partial x} \\ &= -\frac{\epsilon v_x A^2}{W(x)}\sin^2\theta\frac{\partial W}{\partial x}\end{aligned} \tag{6.37}$$

leading to

$$\dot{A} = -\frac{\epsilon v_x A}{2W(x)}\sin^2\theta\frac{\partial W}{\partial x}. \tag{6.38}$$

Thus the dynamical vector field in these coordinates is

$$\begin{aligned}X = &\sqrt{W(x)}\frac{\partial}{\partial\theta} + \epsilon v_x\frac{\partial}{\partial x} - \epsilon\left(\frac{\partial V(x)}{\partial x} + \frac{1}{2}\frac{\partial W(x)}{\partial x}A^2\cos^2\theta\right)\frac{\partial}{\partial v_x} \\ &+ \epsilon v_x\sin\theta\cos\theta\frac{1}{W(x)}\frac{\partial W}{\partial x}\frac{\partial}{\partial\theta} - \epsilon\frac{v_xA}{2W(x)}\sin^2\theta\frac{\partial W}{\partial x}\frac{\partial}{\partial A}.\end{aligned} \tag{6.39}$$

6.1.3.2. Result of the Method of Averaging

The method of averaging says that the projection to x, v_x, A space of the integral curves of X agrees with the evolution on that space of the averaged vector field:

$$\bar{X} \equiv \epsilon \left(v_x \frac{\partial}{\partial x} - \left(\frac{\partial V}{\partial x} + \frac{A^2}{4} \frac{\partial W(x)}{\partial x} \right) \frac{\partial}{\partial v_x} - \frac{v_x A}{4W(x)} \frac{\partial W}{\partial x} \frac{\partial}{\partial A} \right) \tag{6.40}$$

to first order in ϵ for a time $1/\epsilon$. From this we again (though this time it is rigorous) see that for the averaged dynamics

$$\begin{aligned} \dot{J}_0 &= \frac{d}{dt}\left(\frac{1}{2}A^2\sqrt{W(x)}\right) \\ &= A\dot{A}\sqrt{W(x)} + \frac{A^2}{4\sqrt{W(x)}}\frac{\partial W}{\partial x}\dot{x} \\ &= -\frac{v_x A^2}{4\sqrt{W(x)}}\frac{\partial W}{\partial x} + \frac{v_x A^2}{4\sqrt{W(x)}}\frac{\partial W}{\partial x} \\ &= 0. \end{aligned} \tag{6.41}$$

So J is exactly a constant of the motion for the averaged dynamics. This implies that J_0 is conserved to within ϵ for time $1/\epsilon$, i.e. it is an adiabatic invariant. If we consider a level set of J_0 then the averaged vector field restricted to that level set has the form

$$\bar{X}_J = \epsilon \left(v_x \frac{\partial}{\partial x} - \frac{\partial}{\partial x}\left(V(x) + J\sqrt{W(x)}\right)\frac{\partial}{\partial v_x} \right) \tag{6.42}$$

and so is Hamiltonian with an extra pseudo-potential $J\sqrt{W(x)}$.

6.1.3.3. Calculation of the Roto-rate Vector Field

Let us now find the roto-rate symmetry vector field to first order in ϵ. We know that

$$R_0 = \frac{\partial}{\partial\theta} \tag{6.43}$$

from the general theory. We wish to impose the requirement that

$$[R_1, X_0] = -[R_0, X_1]. \tag{6.44}$$

Let us solve this for R_1 component by component. Since

$$X_0 = \sqrt{W(x)}\frac{\partial}{\partial\theta} \tag{6.45}$$

we find that

$$\begin{aligned}
[R_1, X_0]^x &= -\sqrt{W(x)}\frac{\partial R_1^x}{\partial\theta} \\
&= -[R_0, X_1]^x \\
&= -[\frac{\partial}{\partial\theta}, X_1]^x \\
&= -\frac{\partial}{\partial\theta}X_1^x \\
&= 0.
\end{aligned} \tag{6.46}$$

We conclude that

$$R_1^x = f_1^x(x, v_x, A) \tag{6.47}$$

is a constant. The next step in the method is to impose the constraint that the orbits R close after time 2π:

$$\int_0^{2\pi} R_1 d\theta = 0. \tag{6.48}$$

We conclude that the constant vanishes and that

$$R_1^x = 0. \tag{6.49}$$

Similarly, we obtain the v_x component:

$$\begin{aligned}[R_1, X_0]^{v_x} &= -\sqrt{W(x)}\frac{\partial R_1^{v_x}}{\partial \theta} \\ &= -[\frac{\partial}{\partial \theta}, X_1]^{v_x} \\ &= -\frac{\partial}{\partial \theta} X_1^{v_x} \\ &= -\frac{\partial W(x)}{\partial x} A^2 \cos\theta \sin\theta.\end{aligned} \tag{6.50}$$

So

$$\begin{aligned}R_1^{v_x} &= \frac{1}{\sqrt{W(x)}} \int_0^\theta \frac{\partial W(x)}{\partial x} A^2 \cos\tilde{\theta}\sin\tilde{\theta} d\tilde{\theta} + f_1^{v_x}(x, v_x, A) \\ &= -\frac{1}{4\sqrt{W(x)}} \frac{\partial W(x)}{\partial x} A^2 \cos 2\theta + \tilde{f}_1^{v_x}(x, v_x, A).\end{aligned} \tag{6.51}$$

Again the integral condition forces the constant of integration to vanish and we obtain

$$R_1^{v_x} = -\frac{1}{2}\frac{\partial}{\partial x}(\sqrt{W(x)}) A^2 \cos 2\theta. \tag{6.52}$$

We proceed to find R_1^A:

$$\begin{aligned}[R_1, X_0]^A &= -\sqrt{W(x)}\frac{\partial R_1^A}{\partial \theta} \\ &= -[\frac{\partial}{\partial \theta}, X_1]^A \\ &= -\frac{\partial}{\partial \theta} X_1^A \\ &= \frac{v_x A}{W(x)} \sin\theta\cos\theta \frac{\partial W}{\partial x}.\end{aligned} \tag{6.53}$$

Thus we see that

$$\begin{aligned}R_1^A &= -\frac{1}{2\sqrt{W(x)}} \int_0^\theta \frac{v_x A}{W(x)} \frac{\partial W}{\partial x} \sin 2\tilde{\theta} d\tilde{\theta} + f_1^A(x, V_x, A) \\ &= \frac{v_x A}{4W(x)^{3/2}} \frac{\partial W}{\partial x} \cos 2\theta + \tilde{f}_1^A(x, v_x, A).\end{aligned} \tag{6.54}$$

Imposing the integral condition gives

$$R_1^A = \frac{v_x A}{4W(x)^{3/2}} \frac{\partial W}{\partial x} \cos 2\theta. \tag{6.55}$$

Finally we obtain R_1^θ:

$$\begin{aligned}
[R_1, X_0]^\theta &= -\sqrt{W(x)}\frac{\partial R^\theta}{\partial \theta} + R_1^x \frac{\partial}{\partial x}\sqrt{W(x)} \\
&= -\sqrt{W(x)}\frac{\partial R^\theta}{\partial \theta} \\
&= -[\frac{\partial}{\partial \theta}, X_1]^\theta \\
&= -\frac{\partial}{\partial \theta} X_1^\theta \\
&= \frac{1}{2} v_x \cos 2\theta \frac{1}{W(x)} \frac{\partial W}{\partial x}.
\end{aligned} \tag{6.56}$$

Thus we find that

$$\begin{aligned}
R^\theta &= \frac{-v_x}{2W(x)^{3/2}} \frac{\partial W}{\partial x} \int_0^\theta \cos 2\tilde{\theta} d\tilde{\theta} + f_1^\theta(x, v_x, A) \\
&= \frac{-v_x}{4W(x)^{3/2}} \frac{\partial W}{\partial x} \sin 2\theta
\end{aligned} \tag{6.57}$$

where we have already imposed the integral condition.

Thus to first order in ϵ, the roto-rate vector field is

$$\begin{aligned}
R &= R_0 + \epsilon R_1 \\
&= \frac{\partial}{\partial \theta} \\
&- \epsilon \frac{1}{2} \frac{\partial}{\partial x}(\sqrt{W(x)}) A^2 \cos 2\theta \frac{\partial}{\partial v_x} \\
&+ \epsilon \frac{v_x A}{4W(x)^{3/2}} \frac{\partial W}{\partial x} \cos 2\theta \frac{\partial}{\partial A} \\
&- \epsilon \frac{v_x}{4W(x)^{3/2}} \frac{\partial W}{\partial x} \sin 2\theta \frac{\partial}{\partial \theta}.
\end{aligned} \tag{6.58}$$

6.1.3.4. The Hamiltonian Structure

In the original coordinates, the non-trivial Poisson bracket relations are

$$\{x, v_x\} = \epsilon \qquad \{Y, v_y\} = 1. \tag{6.59}$$

Let us determine the Poisson bracket in terms of the variables x, v_x, A and θ. Since neither A or θ depend on v_x, we have that

$$\{A, x\} = \{\theta, x\} = 0. \tag{6.60}$$

We may find the other relations most simply by using the derivation property of the bracket:

$$\begin{aligned} 2A\{A, v_x\} &= \{A^2, v_x\} \\ &= \{Y^2 + \frac{1}{W(x)} v_y^2, v_x\} \\ &= -\frac{v_y^2}{W(x)^2} \frac{\partial W(x)}{\partial x} \{x, v_x\} \\ &= -\frac{\epsilon v_y^2}{W(x)^2} \frac{\partial W(x)}{\partial x} \\ &= -\frac{\epsilon A^2}{W(x)} \sin^2\theta \frac{\partial W}{\partial x}. \end{aligned} \tag{6.61}$$

Thus we obtain the relation

$$\{A, v_x\} = -\frac{\epsilon A}{2W(x)} \sin^2\theta \frac{\partial W}{\partial x}. \tag{6.62}$$

Similarly

$$\begin{aligned} (1 + \tan^2\theta)\{\theta, v_x\} &= \{\tan\theta, v_x\} \\ &= \{-\frac{v_y}{Y\sqrt{W(x)}}, v_x\} \\ &= -\frac{\epsilon\sqrt{W(x)}\sin\theta}{2A\cos\theta} \frac{1}{W(x)^{3/2}} \frac{\partial W}{\partial x}, \end{aligned} \tag{6.63}$$

so

$$\{\theta, v_x\} = -\frac{\epsilon}{2W(x)}\sin\theta\cos\theta\frac{\partial W}{\partial x}. \tag{6.64}$$

Lastly

$$\begin{aligned} 2A\{A,\theta\}(1+\tan^2\theta) &= \{A^2, \tan\theta\} \\ &= \{Y^2 + \frac{1}{W(x)}v_y^2, -\frac{v_y}{Y\sqrt{W(x)}}\} \\ &= -2Y\{Y, v_y\}\frac{1}{Y\sqrt{W(x)}} + \frac{2v_y}{W(x)}\{v_y, Y\}\frac{1}{Y^2}\frac{v_y}{\sqrt{W(x)}} \\ &= -\frac{2}{\sqrt{W(x)}} - \frac{2v_y^2}{Y^2W(x)^{3/2}} \\ &= -\frac{2}{\sqrt{W(x)}} - \frac{2}{\sqrt{W(x)}}\tan^2\theta, \end{aligned} \tag{6.65}$$

so

$$\{A, \theta\} = -\frac{1}{A\sqrt{W(x)}}. \tag{6.66}$$

In the new coordinates the Hamiltonian is

$$\begin{aligned} H &= \frac{1}{2}v_x^2 + \frac{1}{2}v_y^2 + V(x) = \frac{1}{2}W(x)Y^2 \\ &= \frac{1}{2}v_x^2 + \frac{1}{2}A^2W(x)\sin^2\theta + V(x) + \frac{1}{2}W(x)A^2\cos^2\theta \\ &= \frac{1}{2}A^2W(x) + \frac{1}{2}v_x^2 + V(x). \end{aligned} \tag{6.67}$$

One easily checks that this Hamiltonian generates the dynamical vector field in these coordinates using the above Poisson bracket relations.

6.1.4. Obtaining the Action to First Order

Let us now use the Poisson bracket relations to find the function $J_0 + \epsilon J_1 + \ldots$ which generates the roto-rate vector field $R_0 + \epsilon R_1 + \ldots$. From the defining expression

$$R_0 + \epsilon R_1 + \ldots = \{\cdot, J_0 + \epsilon J_1 + \ldots\}_0 + \{\cdot, J_0 + \epsilon J_1 + \ldots\}_1 \tag{6.68}$$

for the action we find a hierarchy of equations

$$\begin{aligned} R_0 &= \{\cdot, J_0\}_0 \\ R_1 &= \{\cdot, J_1\}_0 + \{\cdot, J_0\}_1 \\ &\vdots \end{aligned} \tag{6.69}$$

A priori each J_i is only determined up to Casimirs for $\{,\}_0$ until we get to the next stage. In this example, we may always take these Casimirs to vanish (is this a general phenomenon?).

The zero-order vector field generated by J is

$$\{\cdot, J_0\}_0 = -\frac{\partial J_0}{\partial \theta}\frac{1}{A\sqrt{W(x)}}\frac{\partial}{\partial A} + \frac{\partial J_0}{\partial A}\frac{1}{A\sqrt{W(x)}}\frac{\partial}{\partial \theta}. \tag{6.70}$$

Comparing this with the desired

$$R_0 = \frac{\partial}{\partial \theta}, \tag{6.71}$$

we obtain the relations

$$\frac{\partial J_0}{\partial \theta} = 0 \qquad \frac{\partial J_0}{\partial A} = A\sqrt{W(x)}. \tag{6.72}$$

Up to possible Casimirs, we obtain the expected zero order action:

$$J_0 = \frac{1}{2}A^2\sqrt{W(x)}. \tag{6.73}$$

The first-order vector field generated by J has two pieces. The first is

$$\begin{aligned}\{\cdot, J_0\}_1 &= \{\cdot, A\}A\sqrt{W'(x)} + \{\cdot, x\}\frac{A^2}{4\sqrt{W'(x)}}\frac{\partial W'(x)}{\partial x} \\ &= \frac{A}{2W'(x)}\sin^2\theta\frac{\partial W'}{\partial x}A\sqrt{W'(x)}\frac{\partial}{\partial v_x} - \frac{A^2}{4\sqrt{W'(x)}}\frac{\partial W'}{\partial x}\frac{\partial}{\partial v_x} \\ &= -\frac{A^2}{4\sqrt{W'(x)}}\frac{\partial W'}{\partial x}(\sin^2\theta - \cos^2\theta)\frac{\partial}{\partial v_x} \\ &= -\frac{A^2}{4\sqrt{W'(x)}}\frac{\partial W'}{\partial x}\cos 2\theta\frac{\partial}{\partial v_x}.\end{aligned} \tag{6.74}$$

The second term is $\{\cdot, J_1\}_0$ and has the same form as in the zero order calculation.

Together the two terms give

$$\begin{aligned}\{\cdot, J_1\}_0 + \{\cdot, J_0\}_1 = &-\frac{\partial J_1}{\partial\theta}\frac{1}{A\sqrt{W'(x)}}\frac{\partial}{\partial A} + \frac{\partial J_1}{\partial A}\frac{1}{A\sqrt{W'(x)}}\frac{\partial}{\partial\theta} \\ &- \frac{A^2}{4\sqrt{W'(x)}}\frac{\partial W'}{\partial x}\cos 2\theta\frac{\partial}{\partial v_x}.\end{aligned} \tag{6.75}$$

We must choose J_1 so that this vector field agrees with R. Comparing the $\partial/\partial v_x$ terms, we see that they already agree (this just says that the undetermined Casimirs in the zero order step were actually zero). For the θ and A terms to agree we must have:

$$-\frac{\partial J_1}{\partial\theta}\frac{1}{A\sqrt{W'(x)}} = \frac{1}{4}\frac{v_x A}{W'(x)^{3/2}}\frac{\partial W'}{\partial x}\cos 2\theta \tag{6.76}$$

and

$$\frac{\partial J_1}{\partial A}\frac{1}{A\sqrt{W'(x)}} = -\frac{v_x}{4W'(x)^{3/2}}\frac{\partial W'}{\partial x}\sin 2\theta. \tag{6.77}$$

These are satisfied by

$$J_1 = -\frac{v_x A^2}{8W'(x)}\frac{\partial W'}{\partial x}\sin 2\theta. \tag{6.78}$$

We have thus shown that the function

$$J_0 + \epsilon J_1 = \frac{1}{2}A^2\sqrt{W'(x)} - \epsilon\frac{v_x A^2}{8W'(x)}\frac{\partial W'}{\partial x}\sin 2\theta. \tag{6.79}$$

varies only by order ϵ^2 over times of order $1/\epsilon$

6.2. Some Comments on Perturbation Calculations

Let us use the opportunity of this example to make explicit some very important points about physical perturbation theory which are implicitly contained in the rest of the thesis. Since the time of Galileo, the scientific method has thrived on theorists making all assumptions explicit and basing the acceptance or rejection of a theory on the testable validity of the assumptions and conclusions of a theory. For science to advance, it is important to be precise about the distinction between physical assumptions and rigorous mathematical deduction. In the asymptotic physical theories discussed in this thesis there are two distinct phases of analysis. The first is the "putting the ϵ's in". This process requires physical intuition as to which aspects of a problem are physically important and should be emphasized in the scaling with ϵ. One must make the assumed scaling precise at the beginning of the problem if one is to have hope for a self-consistent theory. Too often workers eliminate terms haphazardly as a calculation is progressing and the resulting theory does not have its assumptions made explicit. Without this one cannot build anything else on top of the theory and the result is the fragmentation of a field.

Once the ϵ's have been inserted, the rest of a derivation should be rigorous mathematics. One must state precisely what problem is to be solved and in particular what time scale of validity is desired and obtained. Let us try to clearly state some facts about coordinates. When we are given a problem in terms of physical variables including ϵ, each physical state is associated with a well-defined point in the state space manifold with these coordinates. The dynamics is a precisely defined vector field on this manifold. When we start with a definite physical state, its time

evolution goes through other definite physical states, regardless of any coordinate system. By changing coordinates, one cannot change the evolution of the system. We have seen in chapter 2 that perturbation theory is independent of coordinates as well. This means that if we change our coordinate system, do nonsingular perturbation theory in the new coordinates and change back to the original system, we always get the same result as working with the original coordinates all along. If the asymptotic expansion of a solution is secular in time (i.e. the coefficients in the expansion blow up as $t \to \infty$, usually on a time scale of order $1/\epsilon$), then this is a physical fact which cannot be changed by going to another coordinate system. (One can introduce fake coodinates which shrink with time making the system apparently non-secular, but rewriting this solution in the original coordinates shows that it really is secular. If the orbits are bounded, then Oseledec's theorem (see section 10.3) shows that one cannot do even this.)

If one is happy with time scales of order 1 then one may use simple non-secular perturbation theory (as discussed in chapter 2) in any coordinate system one desires (in particular there is no advantage or need to do anything fancy like Lie transforms or Kruskal's method). If one wishes to do secular perturbation theory, then again there is a *physical* operation involved which is independent of coordinates. One must eliminate the variable that gives rise to the secularity, say by using the method of averaging or by reduction by an approximate symmetry. This is a physical operation which changes the identification of physical states with points in state space. One must lose the information that causes the secularity. Merely changing coordinates does not lose any information and cannot eliminate secularity by itself. (Changing

coordinates can make the physical operation of reduction or averaging much easier to carry out in practice, but one must not lose sight of the fact that it is mere convenience and has no fundamental signifigance.)

We have seen only two methods for eliminating secularity expressed independently of the coordinate system: the method of averaging and Kruskal's method. Both of these rest in an absolutely essential way on the unperturbed system being made up of purely closed orbits. If we wish to use these methods, we must scale our system to be periodic at $\epsilon = 0$. Sometimes one sees the argument made that while the unperturbed system isn't periodic, we are only interested in the case where orbits almost come back to their starting point, validating averaging. If this is the case, it should be put into the scaling with ϵ! The whole point of the scaling is to make explicit what variation is small compared to what other variation; in a self-consistent theory all such assumptions are put in at the beginning.

One other point in connection with these polemics is that one must he careful not to expend great effort producing a long time theory for a system of equations which are themselves only an approximation for a short time. One place where this issue must be clarified is in plasma physics. One often begins with the Vlasov equation, assumes wave amplitudes are small, and then works with the linearized Vlasov equation. This is just the first order of non-secular perturbation theory, where the disturbance amplitude is scaled with ϵ, and so is valid only for bounded times as $\epsilon \to 0$ as we have seen. One then often proceeds with an analysis of the linearized system using secular perturbation theory (say in analysing ponderomotive effects) which purportedly will give asymptotic answers uniformly on long time

scales. If the long time scale is longer than the bounded time scale for which the linear theory is correct, then as far as the actual physical behavior is concerned the result is not correct. This is not to say that such studies are wrong, only that there is a further assumption behind them (that the scaling of the process studied using secular theory is such that the secular time scale is bounded as far as the original linearization is concerned) which must be made explicit. Higher order *non-secular* theory gives the two-wave, three-wave, etc. interaction equations. Again these are often studied using fancy secular techniques. Again the scaling which makes this a worthwhile endeavor must be made explicit. This will also give restrictions on the domain of validity of the theory.

6.3. 2D Gyromotion via the New Kruskal-like Method

In this section we will apply the perturbation technique developed in chapter 5 to the problem of two-dimensional gyromotion. We will carry out all calculations in explicit detail, so as to provide a model for other calculations done with this method. This work was done in collaboration with Richard Montgomery. The problem we wish to consider is the motion of a charged particle in the (x, y) plane, in the presence of a purely perpendicular magnetic field:

$$B(x, y)\hat{z}. \tag{6.80}$$

The Lorentz force law:

$$m\frac{d\mathbf{v}}{dt} = q(\mathbf{v} \times \mathbf{B}) \tag{6.81}$$

implies that when we set $q = m = c = 1$, the exact non-relativistic equations of motion arise from the dynamical vector field:

$$X = v_x\frac{\partial}{\partial x} + v_y\frac{\partial}{\partial y} + Bv_y\frac{\partial}{\partial v_x} - Bv_x\frac{\partial}{\partial v_y}. \tag{6.82}$$

To do a perturbation analysis, we must introduce the scaling factor ϵ. There are a variety of ways of doing this, but we choose ϵ so that the unperturbed orbits have their velocity vectors rotating at the gyro-frequency and the particle position remains stationary. This scaling is equivalent to making the charge to mass ratio infinite (though we rescale time so that the unperturbed system has well defined dynamics). This scaling is the standard one used in plasma physics. More discussion may be found in [Northrop, 1963] and in the papers [Littlejohn, 1979] and [Littlejohn, 1981] which are also good references for the rest of this section.

This dynamical vector field has only an unperturbed part, X_0, and a first-order perturbation, X_1:

$$X = X_0 + \epsilon X_1 = B\left(v_y \frac{\partial}{\partial v_x} - v_x \frac{\partial}{\partial v_y}\right) + \epsilon\left(v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y}\right) \tag{6.83}$$

If B doesn't vanish, the unperturbed system X_0 has all of its orbits periodic. In this case, we are in the appropriate situation for the application of the method of chapter 5. The goal of the perturbation analysis is to find the (roto-rate) symmetry vector field R, order by order in ϵ:

$$R = R_0 + \epsilon R_1 + \frac{\epsilon^2}{2} R_2 + \ldots.$$

The technique used here requires no special knowledge and makes no arbitrary choices. It is therefore suitable for a computer implementation which could work symbolically order by order to as high an accuracy as desired.

Recall that the condition for R to be a symmetry of X is that their Lie bracket vanish:

$$[R, X] = 0. \tag{6.84}$$

We write this equation order by order. This example has only two terms in X, and

so each equation contains two brackets:

$$
\begin{aligned}
&[R_0, X_0] = 0 \\
&[R_1, X_0] = -[R_0, X_1] \\
&[R_2, X_0] = -2[R_1, X_1] \\
&[R_3, X_0] = -3[R_2, X_1] \\
&\vdots \\
&[R_j, X_0] = -j[R_{j-1}, X_1] \\
&\vdots
\end{aligned}
\tag{6.85}
$$

We saw in chapter 5 that it is convenient to introduce the coordinate θ describing the phase on the unperturbed orbits. This makes it simpler to do the integrals, but is in no way required for any fundamental reason. We introduce polar coordinates in the (v_x, v_y) plane at each point:

$$
\theta \equiv \tan^{-1}\left(\frac{-v_x}{v_y}\right) \qquad v \equiv \sqrt{v_x^2 + v_y^2}. \tag{6.86}
$$

The inverse relations are

$$
v_x = v\cos\theta \qquad v_y = -v\sin\theta. \tag{6.87}
$$

Expressed in these coordinates, the unperturbed dynamical vector field is

$$
X_0 = B\frac{\partial}{\partial\theta}. \tag{6.88}
$$

The first order perturbation of the dynamics is

$$
X_1 = v\cos\theta\frac{\partial}{\partial x} - v\sin\theta\frac{\partial}{\partial y}.
$$

The frequency of the unperturbed orbits is B. As discussed in chapter 5, the zero order symmetry R_0 is parallel to these orbits, but normalized so that all orbits have period 2π:

$$R_0 \equiv \frac{\partial}{\partial\theta} = \frac{1}{B}X_0. \tag{6.89}$$

The four components of the Lie bracket of R_1 with X_0 are given by

$$\begin{aligned}
[R_j, X_0]^x &= -B\frac{\partial R_j^x}{\partial\theta} = -j[R_{j-1}, X_1]^x \\
[R_j, X_0]^y &= -B\frac{\partial R_j^y}{\partial\theta} = -j[R_{j-1}, X_1]^y \\
[R_j, X_0]^v &= -B\frac{\partial R_j^v}{\partial\theta} = -j[R_{j-1}, X_1]^v \\
[R_j, X_0]^\theta &= -B\frac{\partial R_j^\theta}{\partial\theta} + \frac{\partial B}{\partial x}R_j^x + \frac{\partial B}{\partial y}R_j^y = -j[R_{j-1}, X_1]^\theta.
\end{aligned} \tag{6.90}$$

These are the expressions that we use to explicitly perform the integral to determine each component of R_j:

$$\begin{aligned}
R_j^x &= \frac{j}{B}\int_0^\theta [R_{j-1}, X_1]^x d\tilde{\theta} + f_j^x(x, y, v) \\
R_j^y &= \frac{j}{B}\int_0^\theta [R_{j-1}, X_1]^y d\tilde{\theta} + f_j^y(x, y, v) \\
R_j^v &= \frac{j}{B}\int_0^\theta [R_{j-1}, X_1]^v d\tilde{\theta} + f_j^v(x, y, v) \\
R_j^\theta &= \frac{1}{B}\int_0^\theta \Big(j[R_{j-1}, X_1]^\theta + \frac{\partial B}{\partial x}R_j^x \\
&\quad + \frac{\partial B}{\partial y}R_j^y\Big)d\tilde{\theta} + f_j^\theta(x, y, v).
\end{aligned} \tag{6.91}$$

The right hand sides of the Lie bracket equations are all of the form

$$[R_{j-1}, X_1] = [R_{j-1}, v\cos\theta\frac{\partial}{\partial x} - v\sin\theta\frac{\partial}{\partial y}]. \tag{6.92}$$

Let us work out the four components of this in general, so as to make later work a

matter of simple substitution:

$$
\begin{aligned}
[R_{j-1}, X_1]^x &= R^x_{j-1}\cos\theta - R^\theta_{j-1} v \sin\theta - v\cos\theta \frac{\partial}{\partial x}(R^x_{j-1}) + v\sin\theta \frac{\partial}{\partial y}(R^x_{j-1}) \\
[R_{j-1}, X_1]^y &= -R^x_{j-1}\sin\theta - R^\theta_{j-1} v\cos\theta - v\cos\theta \frac{\partial}{\partial x}(R^y_{j-1}) + v\sin\theta \frac{\partial}{\partial y}(R^y_{j-1}) \\
[R_{j-1}, X_1]^v &= -v\cos\theta \frac{\partial}{\partial x}(R^v_{j-1}) + v\sin\theta \frac{\partial}{\partial y}(R^v_{j-1}) \\
[R_{j-1}, X_1]^\theta &= -v\cos\theta \frac{\partial}{\partial x}(R^\theta_{j-1}) + v\sin\theta \frac{\partial}{\partial y}(R^\theta_{j-1}).
\end{aligned}
\tag{6.93}
$$

6.3.1. The Four Components of R_1

From the general theory, we know that R_0 is given by

$$
R_0 = \frac{\partial}{\partial \theta}. \tag{6.94}
$$

Let us now proceed to find the four components of R_1. The constants of integration: $f_1^x, f_1^y, f_1^v, f_1^\theta$, will be determined by the condition that the orbits of R close to first order. We saw in chapter 5 that this implied that

$$
\int_0^{2\pi} R_1 d\theta = 0. \tag{6.95}
$$

We will apply this condition to each component of R_1 after evaluating it. The components of the Lie bracket $[R_0, X_1]$ are

$$
\begin{aligned}
[R_0, X_1]^x &= -v\sin\theta \\
[R_0, X_1]^y &= -v\cos\theta \\
[R_0, X_1]^v &= 0 \\
[R_0, X_1]^\theta &= 0.
\end{aligned}
\tag{6.96}
$$

Substituting these into the general expression gives

$$\begin{aligned} R_1^x &= \frac{-1}{B}\int_0^\theta v\,\sin\tilde{\theta}\;d\tilde{\theta} + f_1^x(x,y,v) \\ &= \frac{1}{B}v\,\cos\theta + \tilde{f}_1^x(x,y,v). \end{aligned} \tag{6.97}$$

Since $\cos\theta$ has zero average over the interval $[0,2\pi]$, the integral condition

$$\int_0^{2\pi} R_1^x d\theta = 0 \tag{6.98}$$

implies that the constant of integration is

$$\tilde{f}_1^x = 0. \tag{6.99}$$

The x component of R_1 is therefore

$$R_1^x = \frac{1}{B}v\,\cos\theta. \tag{6.100}$$

Similarly,

$$\begin{aligned} R_1^y &= -\frac{1}{B}\int_0^\theta v\,\cos\tilde{\theta}\;d\tilde{\theta} + f_1^y(x,y,v) \\ &= -\frac{1}{B}v\,\sin\theta + f_1^y(x,y,v). \end{aligned} \tag{6.101}$$

The integral condition

$$\int_0^{2\pi} R_1^y d\theta = 0 \tag{6.102}$$

implies that the constant of integration is

$$f_1^y = 0. \tag{6.103}$$

The y component of R_1 is therefore

$$R_1^y = -\frac{1}{B}v\,\sin\theta. \tag{6.104}$$

Similarly

$$\begin{aligned} R_1^v &= \frac{1}{B}\int_0^\theta 0\, d\tilde{\theta} + f_1^v(x,y,v) \\ &= f_1^v(x,y,v). \end{aligned} \tag{6.105}$$

The integral condition

$$\int_0^{2\pi} R_1^v d\theta = 0 \tag{6.106}$$

implies that the constant of integration is

$$f_1^v(x,y,v) = 0. \tag{6.107}$$

The v component of R_1 is therefore

$$R_1^v = 0. \tag{6.108}$$

The R_1^θ equation uses these results:

$$\begin{aligned} R_1^\theta &= \frac{1}{B}\int_0^\theta (0 + B_x R_1^x + B_y R_1^y) d\tilde{\theta} + f_1^\theta(x,y,v) \\ &= \frac{1}{B}\int_0^\theta \left(\frac{B_x}{B} v\cos\tilde{\theta} - \frac{B_y}{B} v\sin\tilde{\theta}\right) d\tilde{\theta} + f_1^\theta(x,y,v) \\ &= \frac{1}{B}\left(\frac{B_x}{B} v\sin\theta + \frac{B_y}{B} v\cos\theta\right) + \tilde{f}_1^\theta(x,y,v). \end{aligned} \tag{6.109}$$

The integral condition

$$\int_0^{2\pi} R_1^\theta d\theta = 0 \tag{6.110}$$

implies that the constant of integration is

$$\tilde{f}_1^\theta(x,y,v) = 0. \tag{6.111}$$

The θ component of R_1 is therefore

$$R_1^\theta = \frac{B_x}{B^2} v\sin\theta + \frac{B_y}{B^2} v\cos\theta. \tag{6.112}$$

These expressions may be simplified by introducing a variable representing the reciprocal of the magnetic field:

$$C \equiv \frac{1}{B}.$$

We see that

$$C_x = -\frac{B_x}{B^2} \quad \text{and} \quad C_y = -\frac{B_y}{B^2}. \tag{6.113}$$

We may then write the entire expression for R_1 as

$$R_1 = Cv\cos\theta\frac{\partial}{\partial x} - Cv\sin\theta\frac{\partial}{\partial y} + (-C_x v\sin\theta - C_y v\cos\theta)\frac{\partial}{\partial\theta}. \tag{6.114}$$

6.3.2. The Four Components of R_2

We will now proceed to obtain R_2. The algebraic manipulations are more tedious than those for R_1, but conceptually the calculation proceeds identically.

We begin by calculating the four components of $[R_1, X_1]$. In these expressions it is convenient to use the double angle trigonometric relations (purely for ease of notation):

$$\sin^2\theta - \cos^2\theta = -\cos 2\theta \tag{6.115}$$

and

$$\sin\theta\cos\theta = \frac{1}{2}\sin 2\theta. \tag{6.116}$$

The x component is

$$\begin{aligned}[R_1, X_1]^x &= -(-C_x v\sin\theta - C_y v\cos\theta)v\sin\theta - v\cos\theta C_x v\cos\theta + v\sin\theta C_y v\cos\theta \\ &= v^2 C_x(\sin^2\theta - \cos^2\theta) + 2v^2 C_y\sin\theta\cos\theta \\ &= -v^2 C_x\cos 2\theta + v^2 C_y\sin 2\theta.\end{aligned} \tag{6.117}$$

The y component is

$$\begin{aligned}[R_1, X_1]^y =&(C_x v \sin\theta + C_y v \cos\theta)v\cos\theta + v\cos\theta\, C_x v \sin\theta - v\sin\theta C_y v \sin\theta \\ =&2C_x v^2 \sin\theta\cos\theta + C_y v^2(\cos^2\theta - \sin^2\theta) \\ =&C_x v^2 \sin 2\theta + C_y v^2 \cos 2\theta.\end{aligned} \tag{6.118}$$

The v component is

$$[R_1, X_1]^v = 0. \tag{6.119}$$

The θ component is

$$\begin{aligned}[R_1, X_1]^\theta =& -v\cos\theta(-C_{xx}v\sin\theta - C_{yx}v\cos\theta) + v\sin\theta(-C_{xy}v\sin\theta - C_{yy}v\cos\theta) \\ =&\frac{v^2}{2}C_{xx}\sin 2\theta + v^2 C_{xy}\cos 2\theta - \frac{v^2}{2}C_{yy}\sin 2\theta.\end{aligned} \tag{6.120}$$

As we saw in chapter 5, the condition that the integral curves of R close to second order is

$$\int_0^{2\pi} R_2 d\theta = -\int_0^{2\pi} \frac{\partial R_1}{\partial y_0}\cdot\left(\int_0^\theta R_1 d\tilde{\theta}\right) d\theta. \tag{6.121}$$

This will eventually determine the constants of integration. To prepare for that we will evaluate the right hand side using the value of R_1 obtained above. First find the components of the

$$\int_0^\theta R_1 d\tilde{\theta}$$

integral:

$$\int_0^\theta R_1^x d\tilde{\theta} = \int_0^\theta Cv\cos\tilde{\theta}\; d\tilde{\theta} = Cv\sin\tilde{\theta} \tag{6.122}$$

$$\int_0^\theta R_1^y d\tilde{\theta} = \int_0^\theta -Cv\sin\tilde{\theta}\; d\tilde{\theta} = Cv\cos\tilde{\theta} \tag{6.123}$$

$$\int_0^\theta R_1^v d\tilde{\theta} = 0 \tag{6.124}$$

$$\begin{aligned}\int_0^{\theta} R_1^{\theta} d\tilde{\theta} &= \int_0^{\theta} (-C_x v \sin\tilde{\theta} - C_y v \cos\tilde{\theta}) d\tilde{\theta} \\ &= C_x v \cos\tilde{\theta} - C_y v \sin\tilde{\theta}\end{aligned} \tag{6.125}$$

We now substitute in these results to obtain the general expression:

$$\begin{aligned}\int_0^{2\pi} R_2^{x,y,v,\theta} d\theta = -\int_0^{2\pi} \Big(&\frac{\partial R_1^{x,y,v,\theta}}{\partial x} Cv \sin\theta + \frac{\partial R_1^{x,y,v,\theta}}{\partial y} Cv \cos\theta \\ &+ \frac{\partial R_1^{x,y,v,\theta}}{\partial \theta} (C_x v \cos\theta - C_y v \sin\theta)\Big) d\theta.\end{aligned} \tag{6.126}$$

We now explicitly calculate the four components of this expression.

The x component yields

$$\begin{aligned}\int_0^{2\pi} R_2^x d\theta &= -\int_0^{2\pi} (C_x v \cos\theta \; Cv \sin\theta + C_y v \cos\theta \; Cv \cos\theta \\ &\qquad - Cv \sin\theta (C_x v \cos\theta - C_y v \sin\theta)) d\theta \\ &= -\int_0^{2\pi} (CC_y v^2 \cos^2\theta + CC_y v^2 \sin^2\theta) d\theta \\ &= -2\pi CC_y v^2.\end{aligned} \tag{6.127}$$

The y component yields

$$\begin{aligned}\int_0^{2\pi} R_2^y d\theta &= -\int_0^{2\pi} (-C_x v \sin\theta \; Cv \sin\theta - C_y v \sin\theta \; Cv \cos\theta \\ &\qquad - Cv \cos\theta (C_x v \cos\theta - C_y v \sin\theta)) d\theta \\ &= \int_0^{2\pi} CC_x v^2 (\sin^2\theta + \cos^2\theta) d\theta \\ &= 2\pi CC_x v^2.\end{aligned} \tag{6.128}$$

The v component yields

$$\int_0^{2\pi} R_2^v d\theta = 0. \tag{6.129}$$

The θ component yields

$$\begin{aligned}\int_0^{2\pi} R_2^\theta d\theta = &- \int_0^{2\pi} [(-C_{xx}v\sin\theta - C_{yx}v\cos\theta)Cv\sin\theta + \\ &(-C_{xy}v\sin\theta - C_{yy}v\cos\theta)Cv\cos\theta \\ &+ (-C_x v\cos\theta + C_y v\sin\theta)(C_x v\cos\theta - C_y v\sin\theta)]d\theta \\ = &- \int_0^{2\pi} [-CC_{xx}v^2\sin^2\theta - CC_{yy}v^2\cos^2\theta \\ &- C_xC_xv^2\cos^2\theta - C_yC_yv^2\sin^2\theta]d\theta \\ = &\pi v^2(CC_{xx} + CC_{yy} + C_x^2 + C_y^2).\end{aligned} \tag{6.130}$$

We may finally solve for the four components of R_2. The x component is

$$\begin{aligned}R_2^x = &\frac{2}{B}\int_0^\theta [R_1, X_1]^x d\tilde{\theta} + f_2^x(x,y,v) \\ = &\frac{2}{B}\int_0^\theta (-v^2C_x\cos 2\tilde{\theta} + v^2C_y\sin 2\tilde{\theta})d\tilde{\theta} + f_2^x(x,y,v) \\ = &-v^2CC_x\sin 2\theta - v^2CC_y\cos 2\theta + \tilde{f}_2^x.\end{aligned} \tag{6.131}$$

We next determine the constant of integration f_2^x. We have seen that

$$\begin{aligned}-2\pi CC_y v^2 &= \int_0^{2\pi} R_2^x d\theta \\ &= \int_0^{2\pi} \tilde{f}_2^x d\theta \\ &= 2\pi\tilde{f}_2^x.\end{aligned} \tag{6.132}$$

Thus

$$\tilde{f}_2^x = -CC_y v^2. \tag{6.133}$$

Finally we obtain

$$R_2^x = -v^2CC_x\sin 2\theta - v^2CC_y\cos 2\theta - CC_y v^2. \tag{6.134}$$

The y component of R_2 is

$$\begin{aligned} R_2^y &= \frac{2}{B}\int_0^\theta [R_1, X_1]^y d\tilde{\theta} + f_2^y(x, y, v) \\ &= \frac{2}{B}\int_0^\theta (C_x v^2 \sin 2\theta + C_y v^2 \cos 2\theta) d\tilde{\theta} + f_2^y(x, y, v) \\ &= -v^2 C C_x \cos 2\theta + v^2 C C_y \sin 2\theta + f_2^y(x, y, v). \end{aligned} \tag{6.135}$$

The constant of integration comes from

$$\begin{aligned} 2\pi C C_x v^2 &= \int_0^{2\pi} R_2^y d\theta \\ &= \int_0^{2\pi} f_2^y d\theta \\ &= 2\pi f_2^y. \end{aligned} \tag{6.136}$$

Thus

$$f_2^y = C C_x v^2. \tag{6.137}$$

Finally we obtain

$$R_2^y = -v^2 C C_x \cos 2\theta + v^2 C C_y \sin 2\theta + C C_x v^2. \tag{6.138}$$

The v component of R_2 is

$$\begin{aligned} R_2^v &= \frac{2}{B}\int_0^\theta [R_1, X_1]^v d\tilde{\theta} + f_2^v \\ &= f_2^v. \end{aligned} \tag{6.139}$$

The constant of integration comes from

$$\int_0^{2\pi} R_2^v d\theta = 0. \tag{6.140}$$

Thus

$$f_2^v = 0.$$

Finally we obtain

$$R_2^v = 0. \tag{6.141}$$

The θ component of R_2 requires the above results and is given by

$$\begin{aligned}
R_2^\theta &= \frac{1}{B}\int_0^\theta (2[R_1, X_1]^\theta + B_x R_2^x + B_y R_2^y)d\tilde{\theta} + f_2^\theta(x,y,v) \\
&= \frac{1}{B}\int_0^\theta (v^2 C_{xx}\sin 2\theta + 2v^2 C_{xy}\cos 2\theta - v^2 C_{yy}\sin 2\theta \\
&\qquad + B_x(-v^2 C C_x \sin 2\theta - v^2 C C_y \cos 2\theta - C C_y v^2) \\
&\qquad + B_y(-v^2 C C_x \cos 2\theta + v^2 C C_y \sin 2\theta + C C_x v^2))d\tilde{\theta} + f_2^\theta \\
&= \frac{1}{B}(-\frac{1}{2}v^2 C_{xx}\cos 2\theta + v^2 C_{xy}\sin 2\theta + \frac{1}{2}v^2 C_{yy}\cos 2\theta \\
&\qquad + \frac{1}{2}v^2 B_x C C_x \cos 2\theta - \frac{1}{2}v^2 B_x C C_y \sin 2\theta - B_x C C_y v^2\theta \\
&\qquad - \frac{1}{2}v^2 B_y C C_x \sin 2\theta - \frac{1}{2}v^2 B_y C C_y \cos 2\theta + B_y C C_x v^2\theta) + \tilde{f}_2^\theta(x,y,v) \\
&= -\frac{1}{2}v^2 C C_{xx}\cos 2\theta + v^2 C C_{xy}\sin 2\theta + \frac{1}{2}v^2 C C_{yy}\cos 2\theta \\
&\qquad - \frac{1}{2}v^2 C_x^2 \cos 2\theta + \frac{1}{2}v^2 C_x C_y \sin 2\theta + C_x C_y v^2\theta \\
&\qquad + \frac{1}{2}v^2 C_y C_x \sin 2\theta + \frac{1}{2}v^2 C_y^2 \cos 2\theta - C_y C_x v^2\theta + \tilde{f}_2^\theta(x,y,v) \\
&= -\frac{v^2}{2}[(C C_{xx} - C C_{yy} + C_x^2 - C_y^2)\cos 2\theta - (2C C_{xy} + 2C_x C_y)\sin 2\theta] + \tilde{f}_2^\theta.
\end{aligned} \tag{6.142}$$

To determine the constant of integration, we use

$$\begin{aligned}
\pi v^2(C C_{xx} + C C_{yy} + C_x^2 + C_y^2) &= \int_0^{2\pi} R_2^\theta d\theta \\
&= 2\pi f_2^\theta.
\end{aligned} \tag{6.143}$$

Thus

$$f_2^\theta = \frac{v^2}{2}(C C_{xx} + C C_{yy} + C_x^2 + C_y^2).$$

Finally we obtain

$$R_2^6 = -\frac{v^2}{2}(CC_{xx} - CC_{yy} + C_x^2 - C_y^2)\cos 2\theta + v^2(CC_{xy} + C_xC_y)\sin 2\theta + \frac{v^2}{2}(CC_{xx} + CC_{yy} + C_x^2 + C_y^2). \tag{6.144}$$

6.3.3. Summary of the Calculation

We have thus succeeded in finding the roto-rate vector field up to second order:

$$\begin{aligned} R \sim R_0 + \epsilon R_1 + \frac{\epsilon^2}{2}R_2 \\ = \frac{\partial}{\partial\theta} + \\ \epsilon Cv\cos\theta\frac{\partial}{\partial x} \\ -\epsilon Cv\sin\theta\frac{\partial}{\partial y} \\ +\epsilon(-C_x v\sin\theta - C_y v\cos\theta)\frac{\partial}{\partial\theta} \\ \frac{\epsilon^2}{2}(-v^2CC_x\sin 2\theta - v^2CC_y\cos 2\theta - CC_y v^2)\frac{\partial}{\partial x} \\ \frac{\epsilon^2}{2}(-v^2CC_x\cos 2\theta + v^2CC_y\sin 2\theta + CC_x v^2)\frac{\partial}{\partial y} \\ \frac{\epsilon^2}{2}\Big(-\frac{v^2}{2}(CC_{xx} - CC_{yy} + C_x^2 - C_y^2)\cos 2\theta + v^2(CC_{xy} + C_xC_y)\sin 2\theta \\ +\frac{v^2}{2}(CC_{xx} + CC_{yy} + C_x^2 + C_y^2)\Big)\frac{\partial}{\partial\theta}. \end{aligned} \tag{6.145}$$

Let us summarize what we have done. We began with the exact equations of motion for a two-dimensional particle in a magnetic field. We introduced the scaling parameter ϵ to empbasize the physically important dynamics in such a way that the limiting system for $\epsilon = 0$ has only periodic orbits. We then used the procedure presented in chapter 5 to find the roto-rate vector field order by order. This required no special coordinate system, though we did introduce θ to make the integrals easier

to write. There may exist coordinate systems in which the calculation is simpler and if one is clever enough to find them certainly one should certainly use them. The virtue of our method is that it requires no cleverness and by slugging away doing more integrals like the above we could continue order by order. This explicitness is very important when we want to delegate this labor to machines. We showed in chapter 5 (and Kruskal showed it before) that this roto-rate vector field is the unique vector field whose orbits are all closed loops of period 2π to all orders in ϵ and which commutes to all orders in ϵ with the dynamical vector field X. If we wish to introduce guiding center coordinates $\tilde{X}, \tilde{Y}, \tilde{J}, \tilde{\theta}$, there is a lot of freedom in the choice. What is not free, if the dynamics is to be independent of $\tilde{\theta}$, is that $\tilde{X}$ $\tilde{Y}$ and $\tilde{J}$ must be constant on the orbits of R. We have calculated these orbits in the course of the calculation.

In a Hamiltonian context, the roto-rate vector field is generated by the adiabatically conserved action, which we may then determine order by order (in any coordinates). We turn to this issue in the next section.

6.4. The Hamiltonian Structure of Gyromotion

In the previous section we studied gyromotion without using any Hamiltonian structures. Here we would like to introduce such structures and so make the connection with adiabatic invariance. There are a variety of ways to Hamiltonianize the equations of motion for a particle in a magnetic field. The standard approach introduces canonical momenta which depend on the (unphysical) vector potential $\boldsymbol{A}$. In the paper [Littlejohn, 1979] an approach is developed based on a non-canonical Poisson bracket which itself depends on the magnetic field. The points in phase space represent the true particle position and velocity and the Hamiltonian is just the kinetic energy $v^2/2$. This approach allows a particularly nice formulation of the perturbative scaling used in the last section. The subtlety is that the Poisson bracket itself depends on ϵ and becomes degenerate when $\epsilon = 0$. This makes for a much more interesting analysis and is one motivation for extending the Hamiltonian perturbation results of chapter 4 to singular Poisson systems.

6.4.1. The Poisson Bracket

Explicitly the bracket for the two dimensional system is

$$\begin{aligned}\{f,g\} &= \{f,g\}_0 + \epsilon\{f,g\}_1 \\ &= B\left(\frac{\partial f}{\partial v_x}\frac{\partial g}{\partial v_y} - \frac{\partial f}{\partial v_y}\frac{\partial g}{\partial v_x}\right) + \\ &\quad \epsilon\left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial v_x} - \frac{\partial f}{\partial v_x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial v_y} - \frac{\partial f}{\partial v_y}\frac{\partial g}{\partial y}\right).\end{aligned} \tag{6.146}$$

With the Hamiltonian

$$H = \frac{1}{2}v_x^2 + \frac{1}{2}v_y^2 \tag{6.147}$$

this gives the scaled equations of motion used in the last section:

$$\{\cdot, H\} = \{\cdot, H\}_0 + \epsilon\{\cdot, H\}_1 = X_0 + \epsilon X_1. \tag{6.148}$$

When $\epsilon \to 0$, the bracket reduces to $\{,\}_0$. This bracket is singular and any function of x and y alone is a Casimir.

6.4.2. The Symplectic Structure

It is interesting to look at the ϵ-dependent symplectic structure ω which corresponds to our bracket. Since the bracket is non-singular when ϵ isn't zero, we may invert it to give a well defined symplectic form. Since the bracket becomes singular as ϵ approaches zero, the symplectic form must become infinite in this limit. It is easiest to introduce the matrix J^{ij} representing the components of the contravariant tensor that defines the bracket:

$$\{f, g\} = \sum_{ij} \frac{\partial f}{\partial z^i} J^{ij} \frac{\partial g}{\partial z^j}. \tag{6.149}$$

In the coordinates x, v_x, y, v_y, the matrix J has the form

$$J = \begin{pmatrix} 0 & \epsilon & 0 & 0 \\ -\epsilon & 0 & 0 & B \\ 0 & 0 & 0 & \epsilon \\ 0 & -B & -\epsilon & 0 \end{pmatrix}. \tag{6.150}$$

If we introduce a matrix representing ω by

$$\omega = \sum_{ij} dz^i \omega_{ij} dz^j, \tag{6.151}$$

then it will be the inverse of the matrix corresponding to J. We easily see (say by Gauss' method) that

$$\begin{pmatrix} 0 & \epsilon & 0 & 0 \\ -\epsilon & 0 & 0 & B \\ 0 & 0 & 0 & \epsilon \\ 0 & -B & -\epsilon & 0 \end{pmatrix} \begin{pmatrix} 0 & -1/\epsilon & B/\epsilon^2 & 0 \\ 1/\epsilon & 0 & 0 & 0 \\ -B/\epsilon^2 & 0 & 0 & -1/\epsilon \\ 0 & 0 & 1/\epsilon & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6.152}$$

This shows that the symplectic structure is

$$\omega = \frac{1}{\epsilon^2}(B dx \wedge dy + \epsilon dx \wedge dv_x + \epsilon dy \wedge dv_y). \tag{6.153}$$

This is indeed singular as $\epsilon \to 0$, but if we rescale by multiplying by ϵ^2, we get a well defined form for all ϵ. Now, however, when $\epsilon \to 0$ the form becomes $dx \wedge dy$ which is degenerate. This vanishes when we insert any vector which is a linear combination of v_x and v_y. The correct dynamics is obtained by the usual Hamiltonian prescription for any $\epsilon \neq 0$:

$$\begin{aligned} i_X\omega &= -Bv_y\frac{1}{\epsilon}dx + Bv_x\frac{1}{\epsilon}dy + v_x dv_x - \frac{B}{\epsilon}v_x dy \\ &\quad + \frac{B}{\epsilon}v_y dx + \frac{1}{\epsilon}v_y dy \\ &= v_x dv_x + v_y dv_y \\ &= dH. \end{aligned} \tag{6.154}$$

6.4.3. Uniqueness of the Generator for a Vector Field

The Poisson structure we have given becomes degenerate when $\epsilon = 0$. We must therefore study its properties carefully. First, it is not at all obvious that the association of vector fields with Hamiltonians is unique with this bracket (since to zero order, say, there are Casimirs which may be added to any Hamiltonian without changing the dynamics). We will show that if we have two Hamiltonian vector fields with respect to this bracket, which are asymptotically equal to all orders in ϵ, then their Hamiltonians are also asymptotically equal to all orders in ϵ. Let us subtract the two vector fields in question to obtain a vector field which vanishes to all orders in ϵ. We want to show that only a Hamiltonian which vanishes to all orders can produce such a vector field.

The Poisson bracket has the form

$$\{,\} = \{,\}_0 + \epsilon\{,\}_1. \tag{6.155}$$

The zero order piece $\{,\}_0$ is degenerate and the Casimirs are exactly the functions of x and y alone. The first order piece $\{,\}_1$ is canonical and so non-degenerate. Let us assume that H generates the zero vector field to all orders in ϵ. Expanding H in an asymptotic series

$$H \sim H_0 + \epsilon H_1 + \frac{1}{2}\epsilon^2 H_2 + \ldots \tag{6.156}$$

and working out the generated vector field order by order gives us a hierarchy of

equations:

$$\begin{aligned}
\{\cdot, H_0\}_0 &= 0 \\
\{\cdot, H_1\}_0 &= -\{\cdot, H_0\}_1 \\
\{\cdot, \frac{1}{2}H_2\}_0 &= -\{\cdot, H_1\}_1 \\
&\vdots \\
\{\cdot, \frac{1}{k}H_k\}_0 &= -\{\cdot, H_{k-1}\}_1 \\
&\vdots
\end{aligned} \tag{6.157}$$

The only thing the zero order bracket can produce is derivatives with respect to p_x and p_y. These derivatives must be equal to the right hand sides. If any of the H_k depended on anything but x and y, then the right hand side of its equation would have a derivative which could not be matched on the left. So each H_k depends only on x and y. But these are then Casimirs for the left hand sides which therefore vanish. But the right hand brackets are non-degenerate and so they only vanish when the corresponding H_k is constant. Adding a meaningless constant to the energy, we see that H must vanish order by order.

6.4.4. Comparison With Robert Littlejohn's Results

Let us now show that our results agree with the results obtained by Robert Littlejohn . The two references of interest here are [Littlejohn, 1979] and [Littlejohn, 1981]. To avoid confusion, we will use the notation of these papers and refer to formulas within them. To help other workers make explicit comparisons, we will describe manipulations in detail. Readers without these papers available may want to skip this section.

Let us begin with [Littlejohn, 1979]. In this paper Littlejohn introduces a very clever method based on Darboux's theorem to manipulate the two-dimensional guiding center problem into a form suitable for the application of Lie transforms (the Lie transform is clever too, because the bracket is ϵ-dependent). His Poisson bracket is $1/\epsilon$ times the one we have used and his variable θ is rotated by $\pi/2$ from ours (ours is the clockwise angle of the velocity vector from the $\hat{x}$ axis, his is the clockwise angle of the gyro-radius vector from the $\hat{x}$ axis in a uniform field). Using the Darboux algorithm introduced in the paper, he defines a set of variables (X, Y, θ, J), where θ is the original θ, J is a function whose Poisson bracket with θ is $-1/\epsilon$ to all orders in ϵ and X and Y are coordinates that Poisson commute with both J and θ to all orders in ϵ. These variables are the natural ones for a uniform magnetic field, but other than that have no dependence on the Hamiltonian. These functions are obtained as asymptotic series in ϵ. Using his formulas 4.8, 4.21, 4.20, and 4.29 and letting C denote $1/B$, we find that to first order in ϵ they are

$$\begin{aligned} X &= x - \epsilon v C \cos\theta + \dots \\ Y &= y + \epsilon v C \sin\theta + \dots \\ J &= \frac{v^2}{2} C - \frac{\epsilon v^3}{6}(C\cos C_x - C\sin C_y) + \dots \end{aligned} \tag{6.158}$$

He then performs the Lie transform to find new variables $\bar{X}, \bar{Y}, \bar{J}, \bar{\theta}$ which have the same Poisson bracket relations and such that the Hamiltonian is independent of $\bar{\theta}$ to all orders in ϵ. This then implies that $\bar{J}$ is the adiabatic invariant to all orders in ϵ. We are interested only in this adiabatic invariant (since it is the only quantity that is uniquely determined). By the Poisson bracket relations, the vector field which $\bar{J}$ generates must be simultaneously tangent to the level sets of $\bar{X}, \bar{Y}$, and $\bar{J}$.

The closed orbits of this vector field are parameterized by $\bar{\theta}$ and the Hamiltonian is constant on them (implying that the Hamiltonian vector field commutes with the field generated by $\bar{J}$). These conditions are exactly the ones imposed on our roto-rate vector field R and it is the unique vector field which satisfies them. Thus the adiabatic invariant $\bar{J}$ must generate the roto-rate vector field. We will show that the adiabatic invariant introduced by Littlejohn does indeed generate the roto-rate vector field we calculated (at least to first order).

The expression 5.27 in the paper gives $\bar{J}$. To first order in ϵ, it is

$$\bar{J} = J + \epsilon \frac{(2BJ)^{3/2}}{3B^3}(\hat{a} \cdot \nabla B). \tag{6.159}$$

Using the expressions for these quantities in terms of θ, x, y, and v, we find:

$$\begin{aligned} \bar{J} &= \frac{v^2}{2B} - \frac{\epsilon v^3}{6}(-\cos\theta \frac{B_x}{B^3} + \sin\theta \frac{B_y}{B^3}) \\ &+ \frac{\epsilon v^3}{3B^3}(-\cos\theta B_x - \sin\theta B_y) \\ &= \frac{v^2}{2B} + \frac{\epsilon v^3}{2B^3}(\cos\theta B_x - \sin\theta B_y). \end{aligned} \tag{6.160}$$

Let us check that this agrees with the much more general results presented in [Littlejohn, 1981]. Formula 82e of that paper gives the expression for the adiabatic invariant:

$$\bar{M} = [\frac{w^2}{2B} + \frac{\epsilon}{B^2}[wG_6 + \frac{w^2}{4}(G_8 - 2G_2) + \frac{w^3}{2}F_6]. \tag{6.161}$$

To compare with our much simplified situations we utilize his expressions 17g, 17i, 17c, 16g, 9b, 8, and 11. These show that $w = v$, $G_6 = 0$, $G_8 = 0$, $G_2 = 0$, and $F_6 = \hat{a} \cdot \nabla B/B$. Substituting these in we obtain to first order in ϵ that

$$\bar{M} = \frac{v^2}{2B} + \frac{\epsilon v^3}{2B^3}(\cos\theta B_x - \sin\theta B_y). \tag{6.162}$$

This agrees with the first paper.

Let us now determine the vector field which $\bar{J}$ (or $\bar{M}$) generates. For this we need the Poisson bracket relations:

$$\begin{aligned} x, v &= -\sin\theta \qquad y, v = -\cos\theta \\ x, \theta &= -\frac{1}{v}\cos\theta \qquad y, \theta = \frac{1}{v}\sin\theta \\ \theta, v &= \frac{B}{\epsilon v}. \end{aligned} \tag{6.163}$$

We then see the dynamical vector field generated by $\bar{J}$ to lowest two orders is

$$\begin{aligned} \bar{J} = & \frac{1}{\epsilon}\frac{\partial}{\partial\theta} + \\ & \frac{B}{v}\Big(\frac{3v^2}{2B^3}(\cos\theta B_x - \sin\theta B_y)\frac{\partial}{\partial\theta} - \frac{v^3}{2B^3}(-\sin\theta B_x - \cos\theta B_y)\frac{\partial}{\partial v}\Big) \\ & - \sin\theta\frac{v}{B}\frac{\partial}{\partial x} - \cos\theta\frac{v}{B}\frac{\partial}{\partial y} + \\ & \sin\theta\frac{v^2}{2}\frac{\partial}{\partial x}\Big(\frac{1}{B}\Big)\frac{\partial}{\partial v} + \cos\theta\frac{v^2}{2}\frac{\partial}{\partial y}\Big(\frac{1}{B}\Big)\frac{\partial}{\partial v} \\ & + \frac{v^2}{2}\frac{1}{v}\cos\theta\frac{\partial}{\partial x}\Big(\frac{1}{B}\Big)\frac{\partial}{\partial\theta} - \frac{1}{v}\sin\theta\frac{v^2}{2}\frac{\partial}{\partial y}\Big(\frac{1}{B}\Big)\frac{\partial}{\partial\theta} \\ = & \frac{1}{\epsilon}\frac{\partial}{\partial\theta} \\ & - \frac{v}{B}\sin\theta\frac{\partial}{\partial x} - \frac{v}{B}\cos\theta\frac{\partial}{\partial y} \\ & \Big(\frac{v^2}{2B^2}\sin\theta B_x + \frac{v^2}{2B^2}\cos\theta B_y \\ & - \frac{v^2}{2B^2}B_x\sin\theta - \frac{v^2}{2B^2}B_y\cos\theta\Big)\frac{\partial}{\partial v} \\ & \Big(\frac{3v}{2B^2}\cos\theta B_x - \frac{3v}{2B^2}\sin\theta B_y \\ & - \frac{v}{2B^2}\cos\theta B_x + \frac{v}{2B^2}\sin\theta B_y\Big)\frac{\partial}{\partial\theta} \\ = & \frac{1}{\epsilon}\frac{\partial}{\partial\theta} - \frac{v}{B}\sin\theta\frac{\partial}{\partial x} - \frac{v}{B}\cos\theta\frac{\partial}{\partial y} + \\ & \Big(\frac{v}{B^2}\cos\theta B_x - \frac{v}{B^2}\sin\theta B_y\Big)\frac{\partial}{\partial\theta}. \end{aligned} \tag{6.164}$$

If we make the conversion (due to the different definitions of θ):

$$\begin{aligned} -\sin\theta &\rightarrow \cos\theta \\ -\cos\theta &\rightarrow -\sin\theta \end{aligned} \tag{6.165}$$

and multiply by ϵ, we find that this agrees to order ϵ with the roto-rate vector field R that we obtained in the last section.

6.4.5. Prospects for a Fully Hamiltonian Theory

We may essentially apply the argument of the last section in reverse to obtain the adiabatically invariant action from the roto-rate vector field. Because the generating function is unique, we can solve for it order by order using the expression for the Poisson bracket and the roto-rate vector field. It is interesting that because the zero order piece of the Poisson bracket is degenerate, we can determine a given *order of* J *from the corresponding order of* R only up to Casimirs. These become determined by the next order term in R (only a piece of this higher order term is actually needed, so the full calculation need not be carried out). When one inserts R into the symplectic form ω, the ϵ^{-2} term and the ϵ^{-1} terms vanish, though this is by no means obvious a *priori*. Can this be shown in general? [Kruskal, 1962] proves that if the dynamical vector field is Hamiltonian with respect to an ϵ-independent symplectic structure, then the corresponding roto-rate vector field R is as well, leading to an adiabatic invariant. We have seen that the same is true in this example, even though the Poisson bracket was ϵ-dependent and became degenerate at $\epsilon = 0$. Is it always guaranteed that the symmetry vector field is Hamiltonian in such degenerate cases?

One would like to implement the perturbation analysis we have given directly in terms of the Hamiltonian structures involved instead of going through the essentially non-Hamiltonian analysis. Such a theory would look like the one developed in chapter 4, but requires some extensions. In particular, I expect that:

1. We must do everything with Poisson brackets that have a singular structure and Casimir functions.

2. We must deal with approximate group actions, i.e. we have an action on the J-jet space which does not arise from a true action. Thus, for example, our "circle action" is generated by a vector field given as an asymptotic series in ϵ. The orbits of the ϵ^J truncation of this vectorfield are only closed to order ϵ^{J+1}. We may still reduce asymptotically in this setting.

3. The reduction map, which for $\epsilon = 0$ takes the form $M \to N$, was previously defined from $JM \to JN$. Here we would like to work with the J-jet of diffeomorphisms from M to N. Thus the reduction map is given as an asymptotic series, not the phase space and the reduced space. In the example at hand this will give us the guiding center coordinates as asymptotic series in the original coordinates.

PART II:

WAVES

"Before Maxwell, people conceived of physical reality–insofar as it is supposed to represent events in nature–as material points, whose changes consist exclusively of motions, which are subject to total differential equations. After Maxwell they conceived of physical reality as represented by continuous fields, not mechanically explicable, which are subject to partial differential equations."—Albert Einstein (1931) [Hirsch, 1984]

Chapter 7: Asymptotic Wave Theory

Let us now turn to the next large class of systems: waves. These systems are particularly interesting from a foundational point of view, because the symplectic structure of classical mechanics arises from a natural symplectic structure occuring in asymptotic wave theory as applied to quantum mechanics. All of the elegant and physically important Hamiltonian notions of mechanics may be seen as examples of the new conceptual structures that can arise from asymptotics, as we have discussed throughout this thesis. The concepts of momentum, position, energy, action, rays, and Lagrangian submanifold do not make sense for a given wave or wavepacket. As we separate the wavelength from the scale length, however, they take on asymptotic meaning and give us classical mechanics as we know it.

7.1. Wave Asymptotics and Approximate Symmetry

A crucial element of this simplification proceedure is the idea of approximate symmetry. Part of the asymptotic stretching we will perform on our system will make it approximately symmetric under translations. In any given region of space, an eikonal wave looks like a plane-wave. Consider the translation group acting

on a plane-wave. Translations in a direction contained in the wave-front leave the wave invariant. Translations transverse to the wave-front bring the wave back to its initial state at integral multiples of some distance. The space of distinct plane-waves obtainable from the initial one by translation is identifiable with the quotient of the group of translations by the subgroup which leaves the wave invariant. This quotient is a circle, and we call the angle on the circle the *phase* of the translated wave relative to the reference wave.

For an asymptotically eikonal wave, we have an approximate symmetry given by translation. On the large scale this is not a real translation because we must translate by different amounts at different points in space. Asymptotically, however, it becomes closer and closer to a true translation. On the small scale a translation that slowly varies cannot make neighboring wave crests exactly line up. Asymptotically, however, they will match up more and more closely. We may therefore think of an asymptotic circle action on our space of eikonal waves (using the philosophy of chapters 4 and 5, we may define an asymptotic symmetry to be an exact symmetry on a jet space). It is in this sense that we may assign a unique asymptotic phase to our eikonal wave. Our eikonal methods to eliminate the wavelength scale details and to obtain scale-length sized dynamics and concepts may be seen as asymptotic reduction by this circle action.

The key technical tool which allows us to carry out this asymptotic simplification is the method of stationary phase. On a large scale only average behavior is important. Functions that obey the eikonal separation have zero average. The systematic effects that mount up near places where the wavevector vanishes are ac-

counted for by the method of stationary phase and so allow us to obtain large-scale behavior. These eikonal structures are intimately connected with the essentially mechanical material of the earlier chapters. We shall see in the last chapter that there is also an intimate connection with the statistical averaging that leads to thermodynamics. We will develop the needed wave structures to make this connection clear. This is the reason we discuss contact structures and Legendre submanifolds in places where symplectic structures and Lagrangian submanifolds alone could have sufficed for waves.

7.1.1. Eikonal Waves

A beautiful description of eikonal wave theory for the example of light waves in the geometric optics limit is given in [Born and Wolf, 1970]. Let us begin by defining what we mean by an eikonal wave. Intuitively we mean a physical field in space which is locally like a plane-wave. The local plane-wave is described by a *one-form*, called the *wavevector* k, which is the differential of the wave phase. As discussed above, these notions are not precise for a real wave. To make them precise, we introduce an asymptotic class of waves parameterized by ϵ. When ϵ is 1 this should give the real wave. As ϵ goes to 0, the class should emphasize the tendency that we feel to be physically important for the behavior of the real wave. For the systems we have in mind, this tendency is the separation of wavelength and scale length. We therefore introduce ϵ in such a way as to separate them infinitely as it goes to zero.

There are two natural ways to do this (though any combination of them is also possible). We may let the wavelength go to zero while holding the scale length fixed or we may let the scale length go to infinity as we hold the wavelength fixed as in figure (7.1). I prefer the second tactic because the wavelength often determines the type of wave and its fundamental physical behavior (eg. in plasmas, how close the wave is to the Debye length plays a fundamental role, similarly for electron wavelength compared to the atomic spacing of a crystal in solid state physics). The scale length is often something quite variable (i.e. determined by engineering rather than physics). We may change the size of a tokamak, silicon chip, or auditorium. When we feel eikonal methods provide a valid description, we are saying that we have made things large enough so that the waves propagate in an eikonal fashion. Of course we may let the wavelength go to zero without changing the wave's evolutionary behavior by altering the physical constants in the equation. One common example of this is to say that letting $\hbar$ go to zero in the Schrödinger equation is the way to take the classical limit (in reality we make the potential the particle moves in very slowly varying). We shall see, however, that regardless of how we do the asymptotics, it is often useful to introduce coordinates on the slowly varying scale, and many of the asymptotic concepts will make sense only in these coordinates.

From the geometric viewpoint we have been emphasizing, it is important to regard the wave fields as living on manifolds. There are several reasons that this generality is important. General relativity says that spacetime is really a curved manifold, and so studies of quantum mechanics or electromagnetic radiation on a cosmic scale must utilize a geometric formulation. We have seen in our discussion of

mechanics that even when the manifold structure of spacetime is not important, the natural spaces of interest are products (for more than one component), quotients (by symmetries and ignored variables), and submanifolds (for constrained systems and invariant subsystems) of regions of spacetime and these often have a non-trivial manifold structure. (One class of rich examples arises in the study of linkages. e.g., a system of 3 rigid links in three-dimensions joined in a line by universal joints and constrained to have the two ends fixed in space, has a state space that is naturally a 3-sphere and rotations about the line joining the ends define the Hopf fibration.) Corresponding operations on wave systems lead to the study of waves on non-trivial manifolds (for example, the Y_{lm} spherical harmonics that arise in systems with rotational symmetry are the normal modes of a wave operator on a sphere). As we have been emphasizing, formulations of a theory that make explicit which structures are essential for which phenomena (e.g., what aspects of a coordinate system are used in an essential way in a derivation) give insight into the underlying physics. Lastly, coordinate-free expressions may be evaluated in any coordinates. This allows us to work in the system most convenient for the problem at hand. This is important for systems based on a complicated geometry (e.g., modern plasma fusion devices).

How are we to incorporate the asymptotic scaling into the geometric structure of the manifold in an invariant way? The manifold structure represents slow-scale behavior. We therefore want to "blow up" the manifold asymptotically. For example, most invariant P.D.E.'s use a Riemannian structure on the underlying state space (e.g., the Laplace-Beltrami operator, which is the invariant Laplacian). The

metric should depend on ϵ in such a way that the distance between two points grows indefinitely as $\epsilon \to 0$. The curvature of the manifold then goes to zero with ϵ. Asymptotic fast-scale (ϵ-independent) objects may be defined near a point of the slow manifold as if they were defined on a linear space (which invariantly is the tangent space at the point of interest). One way to see this without a metric is to use the Whitney embedding theorem to embed the manifold smoothly in $\Re^{2N+1}$ with the point of interest sent to the origin. In $\Re^{2N+1}$ we do the scaling by sending $x \in \Re^{2N+1}$ to x/ϵ. If we look in a fixed neighborhood of the origin as $\epsilon \to 0$, our manifold approaches a linear space identifiable with the tangent space. If we change the embedding by a map f that leaves the origin invariant, the asymptotic linear space is changed by the Jacobian of f at the point. This shows that the asymptotic linear approximation space transforms under coordinate changes just like the tangent space. One advantage of using the other type of scaling (where fast-scale objects are scaled smaller and smaller) is that the geometric structure is more immediately recognizable. Two points that are fixed on the fast-scale (for example two crests of an eikonal wave) asymptotically approach one another in the physical manifold as $\epsilon \to 0$. The equivalence class of all points that approach a given point with a given first order rate may be identified as a tangent vector.

7.1.1.1. Sources with Time Scales Generate Eikonal Waves

Eikonal waves appear in nearly every discipline of physics and are one of the most useful analytic tools. Why do they arise in practice? One reason is simply that the systems we wish to study have a fast natural time scale and we vary this

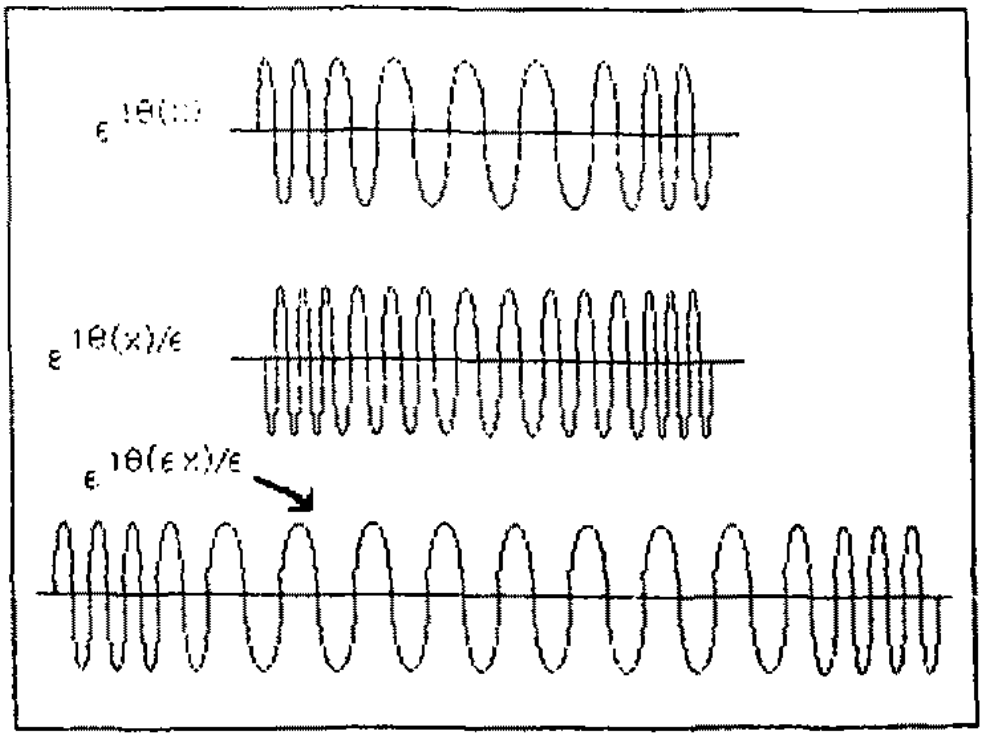

Figure 7.1: Two ways of making a wave asymptotically eikonal.

slowly. The waves that are generated are thus eikonal. For example, a tuning fork has a natural period of oscillation and the dissipative processes that make the amplitude decay and the period change are much slower than this. The tuning fork's generated sound wave is then eikonal. Musical instruments are usually considered to be playing with a definite pitch. The amplitude and timbre (the harmonic mix which determines the characteristic sound of an instrument) vary slowly during the note and distinguish different instruments. Similar examples abound in plasma wave generation processes, solid state waves, atomic light emmission, etc.

7.1.1.2. Dispersive Media Create Eikonal Waves

The other common reason for the prevalence of eikonal waves is the fact that in linear dispersive media, any bounded initial wavepacket will eventually turn into an eikonal wave, and the eikonality will get better and better with time. Imagine throwing a stone into a large lake. This will create a bounded disturbance with some spectrum of wavevectors. For surface gravity waves on water, the long wavelengths have a greater group velocity than the short ones. As our disturbance evolves, the long wavelengths will congregate at the outer edge of the spreading wave. After a long enough time, the different wavelengths present in the intial disturbance will have sorted themselves out radially. As time goes on, the fastest waves will go further than the slower ones and new wave crests will be formed in between. As we wait asymptotically long times, we get asymptotically long stretches of wavetrain close to each wavelength.

7.1.1.3. Whitham's Generalization to Nonlinearity

Whitham has shown that many of the notions of linear eikonal theory carry over to nonlinear waves as well [Whitham, 1974]. The essential change is that, for nonlinear systems, the basic periodic solutions are not necessarily sinusoidal. When we make eikonal waves, they will not be slowly varying sine waves; instead, they will be locally like the nonlinear periodic wavetrains.... with parameters that slowly vary. The nonlinearity also leads to amplitude dependent dispersion relations, which lead to some interesting effects. We will develop parts of this theory to encompass Whitham's ideas in later sections, but let us here focus on linear waves for clarity.

7.1.1.4. Sinusoidal Waves

In this case we shall be interested in slowly varying sinusoids. In a typical situation, we are given a wave of the form

$$A(x)e^{i\theta(x)} \tag{7.1}$$

and are told that the amplitude $A(x)$ and the wavevector $k \equiv d\theta(x)$ are slowly varying functions of x compared to the wavelength $\lambda = 1/|k|$. Using the two schemes discussed above we may write down an asymptotic family as either

$$A(x)e^{i\theta(x)/\epsilon} \tag{7.2}$$

or

$$A(\epsilon x)e^{i\theta(\epsilon x)/\epsilon}. \tag{7.3}$$

Both of these give the original wave when $\epsilon = 1$ and the first shrinks the wavelength, while the second stretches the scale length. We will obtain expressions and concepts relevant to $A(y)$ and $k(y)$ which are slow scale variables (note the absence of ϵ), if we take $y = \epsilon x$. Eikonal waves of this form are described by k as a function of y, an overall phase (since k doesn't set the zero of phase), and the amplitude as a function of y. If we don't care about the phase (and on the large scale we shouldn't, since it changes by order 1 on scales of order ϵ) then we can represent the asymptotic features of our wave as a distribution on y, k space, with support on the surface $k = k(y)$.

7.1.2. The Local Fourier Transform

This representation will result from taking a local Fourier transform. This is a notion one would often like to use in physical situations, but that is usually either ill defined or has very arbitrary components to it. When we introduce the asymptotics as above, however, it becomes precisely defined. The idea is to take an ordinary Fourier transform, but to restrict the domain of integration to the neighborhood of the point y we are interested in in such a way that the domain shrinks to zero on the large scale, but grows to infinity on the small-scale, asymptotically. We can implement this with a *window function* $W_\epsilon(y)$ that asymptotes to a delta function of y, but a constant function of x. A convenient choice is to use a family of Gaussians:

$$W_\epsilon(y) = e^{-y^2/\epsilon}. \tag{7.4}$$

Given an arbitrary asymptotic family, we would like to explicitly obtain its local Fourier transform. Let us define the local Fourier transform of an eikonal family $\psi_\epsilon(x)$ to be

$$\hat{\psi}_\epsilon(y,k) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\epsilon x^2 - ikx}\psi_\epsilon\left(x+\frac{y}{\epsilon}\right)dx. \tag{7.5}$$

This definition is related to ideas presented in [Guillemin and Sternberg, 1977] on page 394 and in [Weinstein, 1978]. If we ignore the asymptotic aspect of this definition which makes the Gaussian convenient but arbitrary, this definition is related to the so-called Iagolnitzer transform (see [Iagolnitzer, 1975]). This expression has many of the nice properties we desire of a local Fourier transform asymptotically. As one can easily see from the manipulations below, the only properties of the window $\exp(-\epsilon x^2)$ that are actually needed are that it is 1 when x is zero, and that it

grows with a scale greater than 1 but less than $1/\epsilon$. Here the width of the Gaussian is of order $1/\sqrt{\epsilon}$ as far as x is concerned.

A specific nice property of this definition is given in the following lemma:

Lemma 7.1. *If an eikonal wave is of the form:*

$$\psi_\epsilon(y) = A(\epsilon y)e^{i\theta(\epsilon y)/\epsilon}, \tag{7.6}$$

then the modulus of its local Fourier transform is asymptotic to:

$$|\hat{\psi}(y,k)| \sim A(y)\delta(k-\theta'(y)). \tag{7.7}$$

Proof. We will show that $\hat{\psi}(y,k)$ itself is asymptotic to $Ae^{i\theta/\epsilon}\delta(k-\theta')$ from which the lemma follows. Choose any smooth test function $f(k)$. Then, letting $\hat{f}(x)$ represent the ordinary Fourier transform of $f(k)$,

$$\begin{aligned}&\int_{-\infty}^{\infty}\hat{\psi}_\epsilon(y,k)f(k)dk = \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\epsilon x^2}\hat{f}(x)A(\epsilon x+y)e^{i\theta(\epsilon x+y)/\epsilon}dx.\end{aligned} \tag{7.8}$$

Now change variables to $X = \epsilon x$:

$$= \frac{1}{2\pi\epsilon}\int_{-\infty}^{\infty}e^{-X^2/\epsilon}\hat{f}(\frac{X}{\epsilon})A(X+y)e^{i\theta(X+y)/\epsilon}dX. \tag{7.9}$$

Since the last three factors are bounded in magnitude at each X as ϵ goes to zero, the first factor allows us to replace the integral by one over an arbitrarily small interval about zero, asymptotically. In fact, we get a contribution to the integral only when X is of order $\sqrt{\epsilon}$. We may thus expand A and θ in Taylor series and keep only the highest order asymptotic contribution, when X is of this order. We obtain

$$= \frac{1}{2\pi\epsilon}A(y)e^{i\theta(y)/\epsilon}\int_{-\infty}^{\infty}e^{-X^2/\epsilon}\hat{f}(\frac{X}{\epsilon})e^{iX\cdot\theta'(y)/\epsilon}dX. \tag{7.10}$$

Let us now go back to the coordinate $x = X/\epsilon$:

$$\frac{1}{2\pi\epsilon} A(y)e^{i\theta(y)/\epsilon} \int_{-\infty}^{\infty} e^{-\epsilon x^2} \hat{f}(x)e^{ix\,\theta'(y)}dx. \tag{7.11}$$

Now let $\epsilon \to 0$ in the integral to obtain the desired result:

$$\int_{-\infty}^{\infty} \tilde{\psi}_\epsilon(y,k)f(k)dk \sim A(y)e^{i\theta(y)/\epsilon} f(\theta'(y)). \tag{7.12}$$

Since f was arbitrary, $\hat{\psi}_\epsilon(y,k)$ is weakly asymptotic to (i.e. agrees when integrated against test functions)

$$A(y)e^{i\theta(y)/\epsilon}\delta(k-\theta'(y)).$$

Q.E.D.

7.1.3. Stationary Phase, Laplace's Method, and Steepest Descents

The method of *stationary phase* is the central tool in doing wave asymptotics. It and steepest descents will be the central theoretical tools of chapter 16. There are many different cases and situations where one might apply this method, but we will restrict ourselves to the simplest cases. The basic idea of the method is that the integral of a short wavelength wave against a slowly varying function will vanish asymptotically. In fact, such integrals vanish to all orders in the asymptotic parameter ϵ, as is quite easy to show. Consider, for instance, the integral

$$\int_{-\infty}^{\infty} f(y)\cos(y/\epsilon)dy, \tag{7.13}$$

or equivalently in terms of $x = y/\epsilon$,

$$\int_{-\infty}^{\infty} f(\epsilon x)\cos(x)\epsilon dx, \tag{7.14}$$

where $f(x)$ is assumed to be smooth and to die off at infinity. If we integrate by parts, we obtain

$$\epsilon f(\epsilon x)\sin(x)|_{-\infty}^{\infty} - \epsilon \int_{-\infty}^{\infty} f'(\epsilon x)\sin(x)\epsilon dx. \tag{7.15}$$

The first term vanishes and the second is ϵ times an integral of the type we are considering. Repeating this procedure puts as many ϵ's out in front as we desire, showing that the integral vanishes to all orders in ϵ.

If we have a slowly varying frequency in the cosine, such as $\cos(g(x)/\epsilon)$, and if dg never vanishes, we may change coordinates using g to get an integral with $\cos(x)$ and a slowly varying Jacobian of the transformation. This is the situation above, and we may again conclude that the integral vanishes to all orders. Thus if an integral of this type is to have a non-zero asymptotic value, it must arise from the regions where $dg = 0$. In fact, it is easy to see, by chopping our integral into one on an interval around $dg = 0$ and one on its complement, that the width of the non-uniform region for the above argument is of order $\sqrt{\epsilon}$. This shows that if g's second derivative is non-zero, the value of the integral will be of order $\sqrt{\epsilon}$ and the only terms in the Taylor series of g that can contribute to leading order are those of the second order. The integral then becomes one over a Gaussian, which may be evaluated by elementary methods. That this gives the highest order asymptotics may also be seen by using the Morse lemma to make a slow change of coordinate to a quadratic (see [Guillemin and Sternberg, 1977] p. 16).

The same idea may be used in integrals over an arbitrary number n of dimensions, and assuming a single stationary point at $y = 0$ leads to the formula (see

[Guillemin and Sternberg, 1977] p. 6)

$$\int f(y)e^{ig(y)/\epsilon}dy \sim \frac{f(0)(2\pi\epsilon)^{n/2}}{\sqrt{|\det g_{yy}(0)|}}\, e^{ig(0)/\epsilon+(i\pi/4)\mathrm{Sign}\, g_{yy}(0)} + O(\epsilon^{1+n/2}). \tag{7.16}$$

Here we assume that the Hessian g_{yy} is non-degenerate and "Sign" denotes the signature which is the number of positive eigenvalues minus the number of negative ones. The term in the exponential with this factor gives rise to the "extra phase shifts in going through caustics" and is responsible for the notorious correction factors in the Bohr-Sommerfeld expression for energy eigenvalues in the old quantum theory (eg. the extra $\hbar\omega/2$ for the harmonic oscillator).

If we have a real exponent instead of an imaginary one, then we may use Laplace's method. Maxima of the exponent tend to completely dominate under exponentiation, asymptotically. In this situation, the Taylor expansion of the exponent near the maxima (and possibly the endpoints) completely determine the asymptotic expansion of the integral (though one typically has to consider terms of order $2j$ in the Taylor expansions to get terms of order j in the expansion of the integral). To highest order, if ϕ has a maximum at the point $a < c < b$ and $\phi''(c) < 0$ then asymptotically

$$\int_a^b f(t)e^{\phi(t)/\epsilon}dt \sim \sqrt{\frac{2\pi\epsilon}{-\phi''(c)}}f(c)e^{\phi(c)/\epsilon}. \tag{7.17}$$

We will use this result when we look at statistical mechanics in a way that is analogous to the use of stationary phase in wave mechanics.

The method of steepest descents generalizes stationary phase and Laplace's method to saddle points occuring anywhere in the complex plane. The idea is to

deform the contour of integration in the neighborhood of the saddle point so that the phase of the integrand has constant imaginary part and a maximum in the real part. We can then use Laplace's method asymptotically.

Lastly, the stationary phase situation with no stationary points or maxima, may often be converted to a situation where we can use the *Riemann-Lebesgue lemma*. This says that

$$\int_a^b f(t)e^{it/\epsilon}dt \to 0 \quad \text{as} \quad \epsilon \to 0, \tag{7.18}$$

if

$$\int_a^b |f(t)|\, dt \tag{7.19}$$

exists. We don't have to assume any differentiablity for integration by parts in this situation.

7.1.3.1. Heisenberg's Uncertainty Principle

Heisenberg's uncertainty principle puts limitations on how tightly one can localize a function and its Fourier transform at the same time. In quantum mechanics this puts fundamental limitations on how accurately one may measure the position and the momentum of a particle at the same time. As we discuss in section 11.2, one can think of quantum state as corresponding to a region in phase space of volume h^n (where h is Planck's constant). The usual proofs of the uncertainty principle rely on inequalities that are hard to remember. Let us demonstrate it here using functional derivatives.

We use the Dirac notation where $\langle\phi|\psi\rangle$ represents the L^2 pairing of ϕ and ψ. We work with a normalized function ψ defined on the real line:

$$\langle\psi|\psi\rangle = 1. \tag{7.20}$$

By a transformation of the form

$$\psi(x) \to e^{ik_0 x}\psi(x - x_0) \tag{7.21}$$

we can put the mean values of x and k to zero without affecting the dispersions. We therefore assume that

$$\langle\psi|x|\psi\rangle = 0 = \langle\psi|k|\psi\rangle. \tag{7.22}$$

The dispersions on which we want to put bounds then take the form

$$(\Delta x)^2 \equiv \langle\psi|x^2|\psi\rangle \qquad (\Delta k)^2 \equiv \langle\psi|k^2|\psi\rangle. \tag{7.23}$$

One easily evaluates the functional derivatives

$$\begin{aligned} \frac{\delta\langle\psi|\psi\rangle}{\delta\psi}(x) &= \psi^*(x) \\ \frac{\delta(\Delta x)^2}{\delta\psi}(x) &= x^2\psi^*(x) \\ \frac{\delta(\Delta k)^2}{\delta\psi}(k) &= k^2\psi^*(k). \end{aligned} \tag{7.24}$$

Fourier transforming this last expression gives

$$\frac{\delta(\Delta k)^2}{\delta\psi}(x) = -\frac{d^2}{dx^2}\psi^*(x). \tag{7.25}$$

We want to show that the product of the dispersions reaches some minimum value on normalized functions. We do this by showing that the Gaussian has the minimum

uncertainty of any wave and then evaluating its uncertainty. We formulate the minimum uncertainty normalized wavepacket problem as a variational expression with Lagrange multiplier α. With the function F defined as

$$F = (\Delta x)^2(\Delta k)^2 - \alpha\langle\psi|\psi\rangle, \tag{7.26}$$

a minimal uncertainty wavepacket will be a point where F has a vanishing functional derivative with respect to ψ. So

$$0 = \frac{\delta F}{\delta\psi}(x) = x^2\psi^*(x)(\Delta k)^2 + (\Delta x)^2\left(-\frac{d^2}{dx^2}\psi^*(x)\right) - \alpha\psi^*(x). \tag{7.27}$$

Since $(\Delta k)^2$ and (Δx^2) are just real numbers, this yields the differential equation

$$\frac{d^2}{dx^2}\psi^* = \left(\frac{\alpha}{(\Delta x)^2} - \frac{(\Delta k)^2}{(\Delta x)^2}x^2\right)\psi^*. \tag{7.28}$$

We may easily solve this, and imposing the normalization condition determines α; imposing zero means for x and k and choosing a phase factor so that ψ is real determines the constants of integration and gives a relation between Δx and Δk. We are left with the solution (which is easily checked by plugging into the equation):

$$\psi(x) = [2\pi(\Delta x)^2]^{-1/4}e^{-x^2/4(\Delta x)^2}. \tag{7.29}$$

To see that this extremal is really a minimum, we calculate the second functional derivative:

$$\frac{\delta^2 F}{\delta\psi^2}(x,y) = x^2\psi^*(x)\left(-\frac{d^2}{dy^2}\psi^*(y)\right) + y^2\psi^*(y)\left(-\frac{d^2}{dx^2}\psi^*(x)\right). \tag{7.30}$$

When we plug in the Gaussian, both terms are positive, showing that the Gaussian is indeed the minimum uncertainty packet. We need only find the uncertainties for this packet, which entails doing some Gaussian integrals. We see that

$$\int_{-\infty}^{\infty} x^2\psi^2(x)dx = (2\pi(\Delta x)^2)^{-1/2}\int_{-\infty}^{\infty} x^2 e^{-x^2/2(\Delta x)^2}dx = (\Delta x)^2 \tag{7.31}$$

and

$$\begin{aligned}(\Delta k)^2 &= \int_{-\infty}^{\infty} k^2 \hat{\psi}(k)^2 dk \\ &= \int_{-\infty}^{\infty} k^2 \left(\frac{2(\Delta x)^2}{\pi}\right)^{1/2} e^{-2(\Delta x)^2 k^2} dk \\ &= \frac{1}{4(\Delta x)^2}.\end{aligned} \tag{7.32}$$

Thus we find the relation

$$\Delta k \Delta x > \frac{1}{2} \tag{7.33}$$

which is Heisenberg's uncertainty relation (with $\hbar$ set to 1).

7.1.3.2. Asymptotic Waves with a Definite k and y

Heisenberg's uncertainty principle tells us that the product of the *absolute* uncertainty in x space and k space (i.e. the dispersion in x and k without regard for the size of x or k) is bounded from below. The minimum uncertainty wave packet centered at a given x_0 and k_0 is given by a complex Gaussian:

$$[2\pi(\Delta x)^2]^{-1/4} \exp\left\{-\frac{(x-x_0)^2}{4(\Delta x)^2} + ik_0 x\right\}. \tag{7.34}$$

The dispersion in x is given by Δx and in k it is $1/2\Delta x$. The relative dispersions are $\Delta x/x$ and $\Delta k/k$. If we let either x or k get asymptotically large, we may make both of these relative dispersions go to zero. In the stretched cordinates: (x, K) or (y, k) where $\epsilon x = y$ and $\epsilon k = K$, the local Fourier transform can be an asymptotic δ-function in both directions. Any eikonal wave may be asymptotically decomposed into these δ states.

For example, in quantum mechanics $\hbar$ is often used as the asymptotic parameter. The momentum operator $\hat{p}$, defined as

$$\hat{p} \equiv \frac{\hbar}{i}\frac{\partial}{\partial x} \tag{7.35}$$

is exactly a stretched version of the wave vector k as $\hbar$ asymptotes to zero. We may therefore create wavepackets that have a definite x and p as $\hbar$ goes to zero. Heisenberg's uncertainty principle for x and p takes the form:

$$\Delta x \cdot \Delta p \geq \frac{1}{2}\hbar. \tag{7.36}$$

We see then that as $\hbar \to 0$ we may make both Δx and Δp vanish.

These asymptotic states are intimately related to the theory of coherent states and have been connected with the Lie Poisson structures associated with the Heisenberg group. Let us quickly sketch some of the results in [Yaffe, 1982], but reformulate them in terms of momentum maps. The idea is to consider an asymptotic family of "quantum mechanics's" labeled by the parameter $\hbar$. For each value of $\hbar$ we have a Hilbert space $H_\hbar$ of L^2 wavefunctions on $\Re^n$ and each operator (like $\hat{p}$ above) is defined for each $\hbar$ on $H_\hbar$. We choose a special state $|0\rangle_\hbar$ in each $H_\hbar$ which will asymptotically represent the state with zero position and zero momentum. This is chosen to be one of our special states with vanishing position and momentum dispersions as $\hbar \to 0$, such as

$$\langle x|0\rangle_\hbar \equiv (\pi\hbar)^{-1/4} e^{-(x^2/2\hbar)}. \tag{7.37}$$

Quantum mechanics has a natural Hamiltonian structure. We may consider $1/\hbar$ times the imaginary part of the Hermitian inner product $\langle|\rangle$ as a symplectic

structure on the Hilbert space (being a linear space, we may lift this structure to each tangent space). The Schrödinger evolution defined by the Hermitian Hamiltonian operator $\hat{H}$ is Hamiltonian with respect to this symplectic structure and a Hamiltonian *function* given by the expectation value of $\hat{H}$:

$$H(\psi) \equiv \langle \psi | \hat{H} | \psi \rangle. \tag{7.38}$$

There is a natural $2n+1$ dimensional Lie group that is intimately connected with the asymptotics of quantum mechanics, called the Heisenberg group. We may consider group elements to lie in $\Re^n \times \Re^{n*} \times \Re$ with coordinates (q, p, α) (sometimes the α factor is taken to lie in a circle). The multiplication law is

$$(q, p, \alpha) \cdot (q', p', \alpha') = (q + q', p + p', \alpha + \alpha' - \langle q, p' \rangle). \tag{7.39}$$

Here we use $\langle , \rangle$ to mean the pairing of $\Re^n$ with $\Re^{n*}$. This is just the translation group on $\Re^{2n}$ with the extra α factor twisted into the multiplication. The Heisenberg group naturally arises through an irreducible representation on the Hilbert spaces $H_\hbar$ defined by the mapping to unitary operators:

$$(q, p, \alpha) \mapsto e^{i\alpha/\hbar} e^{i\langle p, \hat{x} \rangle/\hbar} e^{-i\langle q, \hat{p} \rangle/\hbar}. \tag{7.40}$$

$(q, 0, 0)$ translates wavefunctions by q *in position space*, $(0, p, 0)$ *translates the* Fourier transform of wavefunctions by p in momentum space, and $(0, 0, \alpha)$ changes the phase.

The elements of the orbit of the special state $|0>_\hbar$ under this group action are called coherent states and are labeled by the (q, p, α) which acts to produce

them. The irreducibility of the group action may be used to immediately prove many interesting relations such as the decomposition of the identity:

$$\hat{1} = c_\hbar \int |q,p,\alpha\rangle\langle q,p,\alpha|, \tag{7.41}$$

where the integral is over an invariant measure on the Heisenberg group. This group action is Hamiltonian and we may ask for its momentum map. Each element u in the Lie algebra g of the Heisenberg group has an associated Hermitian operator $\hat{u}_\hbar$ defined on $H_\hbar$ which generates the action of the one-parameter subgroup tangent to u. As for the Hamiltonian operator, the corresponding Hamiltonian function simply associates to each state $|\psi\rangle$ the expectation value $\langle\psi|\hat{u}_\hbar|\psi\rangle$. The momentum map J then sends ψ to that element of the dual of the Lie algebra g^* which satisfies

$$\langle J(\psi), u\rangle = \langle\psi|\hat{u}_\hbar|\psi\rangle \tag{7.42}$$

for each u in g. We may easily see that this is equivariant since if β is an element the Heisenberg group, then

$$\begin{aligned}\langle J(\hat{\beta}\psi), u\rangle &= \langle\hat{\beta}\psi|\hat{u}_\hbar|\hat{\beta}\psi\rangle \\ &= \langle\psi|\hat{\beta}^{-1}\hat{u}_\hbar\hat{\beta}|\psi\rangle \\ &= \langle\psi|\widehat{Ad_\beta \cdot u}|\psi\rangle \\ &= \langle J(\psi), Ad_\beta \cdot u\rangle \\ &= \langle Ad^*_\beta \cdot J(\psi), u\rangle.\end{aligned} \tag{7.43}$$

The dual of the Lie algegra g^* is $2n+1$ dimensional. The coadjoint orbits of the Heisenberg group in g^* consist of $2n$ dimensional planes labeled by a parameter when that parameter is non-zero and an entire plane of individual points when the

parameter ·.ishes. On each of these spaces the KKS symplectic structure is a multiple of the canonical structure on $\Re^n \times \Re^{n*}$. Because the coherent states are an orbit of the Heisenberg group, their image under the momentum map J is exactly one coadjoint orbit. It turns out that coherent states that get sent to the same element in g^* are not distinguishable by means of operators that have a nice classical limit as $\hbar \to 0$ (the eliminated degree of freedom is the phase). Asymptotically, the $2n$-dimensional coadjoint orbit is the natural arena for dynamics. Associated with each operator with a nice classical limit is a function on this coadjoint orbit whose value on an element is the expectation value of the operator in any of the coherent states corresponding to that element (they all give the same value). This real-valued function on the coadjoint orbit is called the symbol of the operator. The symbol of the product of two such operators is simply the product of their symbols as $\hbar \to 0$. The symbol of the limit of $i/\hbar$ times the commutator of two operators is the Lie-Poisson bracket of the symbols of the operators. It would be interesting to extend these definitions to the entire dual of the Lie algebra (say by giving a family of $|0>$'s, one for each coadjoint orbit).

7.1.4. Eikonal Waves and Lagrangian Submanifolds

"The correspondence is an illustration of what I might call the "symplectic creed": EVERYTHING IS A LAGRANGIAN SUBMANIFOLD."--Alan Weinstein on p. 5 of [Weinstein, 1981]

We have seen that the result of our aysmptotic local Fourier transform on an eikonal wave family is a distribution on (y, k) space whose support is on the set

$k = d\theta$. We see that k is naturally a one-form based at the point y, it being equal to the differential of the phase function there. Geometrically we are to think of (y, k) space as the cotangent bundle of y space. We have seen that this has a natural symplectic structure that in coordinates is minus the differential of the canonical one-form: $k_a\ dy^a$. That the surface defined by our wave's singular support is the graph of the differential of a function is locally equivalent to the symplectic property *of being Lagrangian (if the submanifold projects diffeomorphically to y space).*

We may see this connection geometrically as follows. If we think of an arbitrary one-form α on y space as a mapping from y space to its cotangent bundle, then the pullback of the canonical one-form back to y space yields the form α. (In coordinates: $\alpha = \alpha_i dy^i$ and the canonical one form is $k_i dy^i$. The mapping defined by α takes the point with coordinates y^i to the point with coordinates (y^i, α_i). The canonical one-form on the image is $\alpha_i dy^i$ which pulls back to α as desired.) The canonical one-form restricted to the graph of the differential of a function has zero exterior derivative, since its pullback to y space does (since $d \circ d = 0$ and exterior differentiation and pullback commute). Thus the symplectic form restricts to zero on the graph. This graph is of the same dimension as the base and so is Lagrangian. If a submanifold is Lagrangian and projects diffeomorphically to the base, then by Poincaré's lemma the canonical one-form is locally the differential of a function. But then so is the one-form whose graph the manifold is.

Thus eikonal waves are asymptotically associated with Lagrangian submanifolds that don't "bend over" in the cotangent space and so don't have a singular projection. When we allow an eikonal wave to evolve in time, the dynamics may

bend the corresponding Lagrangian submanifold over. At such times, the originally eikonal wave has ceased to be eikonal. The image of the points with a singular projection forms the caustic of the wave (so named because such places have a high intensity and so tend to get hot in optical fields). The straightforward asymptotics of traditional WKB theory breaks down at these points and, a priori, one might not expect the corresponding Lagrangian submanifold to have anything more to do with the wave. Maslov introduced the concept and name of Lagrangian submanifolds in [Maslov, 1965] while generalizing earlier one-dimensional work of Keller in [Keller, 1958]. Maslov was able to show that there is a more general asymptotic class of asymptotic families than eikonal waves which is associated with arbitrary Lagrangian submanifolds. The asymptotic dynamics of an eikonal wave does not leave this larger class and in fact the bent over Lagrangian manifold continues to represent the wave. The basic idea is to treat our wave as being on a higher dimensional space where the correponding Lagrangian submanifold is not singular, but still projects onto the singular one over the space we are really interested in. Projection of one space onto another corresponds to integrating the wave over the fibers of the projection. Since the evolution equations are linear, one may introduce dynamics on the large space which projects to the correct dynamics on the space of interest. On the large space, everything is eikonal and so we get representatives of the caustic wave fields as integrals of eikonal waves on a higher dimensional space ([Guillemin and Sternberg, 1977] p. 428).

7.2. WKB Theory and Asymptotic Equations

Let us quickly sketch the theory of eikonal waves for linear P.D.E.'s. This may be generalized to vector equations, asymptotic series in fractional powers of the amplitude, and higher order terms (see for example [Guillemin and Sternberg, 1977] p. 50), but I want to focus on the bare essentials here.

In the mathematical literature on this subject (such as0 [Hörmander, 1983]), one usually is looking for asymptotic solutions to an *asymptotic* partial differential equation on a manifold. One makes geometric sense of a partial differential operator as a certain class of mappings between spaces of sections of bundles over the manifold. One introduces a small parameter ϵ and introduces the usual asymptotic equivalence classes of ϵ-dependent operators to define asymptotic operators. An asymptotic P.D.E. is given by requiring an asymptotic operator to vanish on an asymptotic function. One usually assumes that the higher derivative terms have coefficients with higher powers of ϵ, so that to make the terms balance, a solution must oscillate more and more as ϵ vanishes. The resulting class of solutions are of the form of an asymptotic amplitude times an ever more quickly oscillating exponential. One finds a Hamilton-Jacobi equation for the phase. In addition, one obtains series of transport equations, defined along the characteristics of the Hamilton-Jacobi equation, for the terms in the amplitude's expansion. The Hamiltonian is the so-called principal symbol of the operator, which *invariantly is* a function on the cotangent bundle. Because the wavelength is getting ever smaller, the local asymptotic behavior of the waves is unaffected by the global structure of the manifold.

We are interested in scaling our system the other way. We want to make the

coefficients of our equation slowly varying as $\epsilon \to 0$. If we are on a manifold, then the manifold should "grow asymptotically, to become more and more like $\Re^n$ locally. This type of scaling is often the physically relevant one. Most of the manifolds used in physics are really only manifolds in some asymptotic sense. Even if spacetime really is a manifold (which is very unlikely on scale lengths of order of the Planck length), the state spaces for our systems come from large products of space with itself (representing the state of many particles) followed by projections and constraints. There is always some "width" in the constraint direction, and the manifold picture breaks down on this scale. Similarly, quantum mechanics imposes finest scales on which it is reasonable to look at the eikonal state of a wavepacket as being a point in a manifold. Instead of letting this physically determined small scale shrink, we often mean to say that the large scale structure of the state space is not strongly affecting local behavior. We may represent this asymptotically as in section 7.1.1.

We also want to say only that the arbitrary scale lengths in our problem get large, and not to change the physical relations of the equation. For example, assume we are studying internal waves in the ocean and want to consider slow salt gradient variations. We introduce asymptotics which makes the gradient variation more and more gradual asymptotically. If we were to shrink the wavelength instead, we would be changing the *physics* of internal waves.

We are therefore interested in equations of the form

$$P(\epsilon x, \frac{1}{i}\frac{\partial}{\partial x}) \cdot \psi_\epsilon(x) = 0, \tag{7.44}$$

where $P(y,k)$ is a smooth function on the cotangent bundle of y space, and we

assume its growth in k is bounded by some power of k (so we have a finite order equation). Notice that we are treating y as an element of a vector space for the reasons discussed in section 7.1.1.

As an example, we will consider the Klein Gordon equation in 3-dimensions: This could have slowly varying coefficients, but for simplicity we demonstrate only the effect of slowly varying initial conditions. The base space is 4-dimensional and is parametrized by (x, y, z, t). The equation may be written

$$\left(-\left(\frac{1}{i}\frac{\partial}{\partial t}\right)^2+\left(\frac{1}{i}\frac{\partial}{\partial x}\right)^2+\left(\frac{1}{i}\frac{\partial}{\partial y}\right)^2+\left(\frac{1}{i}\frac{\partial}{\partial z}\right)^2+1\right)\psi=0 \tag{7.45}$$

So the function P on the cotangent bundle is given by

$$P(\epsilon x, \epsilon y, \epsilon z, \epsilon t, k_x, k_y, k_z, k_t) = -k_t^2 + k_x^2 + k_y^2 + k_z^2 + 1. \tag{7.46}$$

We look for solutions of the form

$$\psi_\epsilon = A(\epsilon x)e^{i\theta(\epsilon x)/\epsilon}. \tag{7.47}$$

In taking a derivative of this, we get some terms that come from differentiating $A(\epsilon x)$, and these will have as many powers of ϵ as there were orders of differentiation. In contrast, derivatives of $\exp(i\theta(\epsilon x)/\epsilon)$ do not bring down any extra ϵ's (though once a $\theta'(\epsilon x)$ has come down, any higher derivatives of it will get extra ϵ's). If $P\cdot\psi$ is to vanish to all orders in ϵ, it must vanish term by term. The lowest order term has all derivatives hitting the exponential and looks like

$$P(y, d\vartheta(y))\psi_\epsilon = 0, \tag{7.48}$$

where as before $y = \epsilon x$ and $d\theta$ is the exterior derivative of θ on y space.

If our solution doesn't vanish, then θ must satisfy the Hamilton-Jacobi equation:

$$P(y, d\theta(y)) = 0. \tag{7.49}$$

The analysis of such equations leads to a rich theoretical structure. To see where this comes from, in the next section we consider arbitrary first order P.D.E.'s and then specialize to those of Hamilton-Jacobi type.

For the Klein-Gordon example, the Hamilton-Jacobi equation is

$$-\left(\frac{\partial\theta}{\partial t}\right)^2 + \left(\frac{\partial\theta}{\partial x}\right)^2 + \left(\frac{\partial\theta}{\partial y}\right)^2 + \left(\frac{\partial\theta}{\partial z}\right)^2 + 1 = 0. \tag{7.50}$$

7.2.1. The Structure of First Order P.D.E.'s

Most of the symplectic structures in physics may trace their origin to a natural structure that arises with any (nonlinear) first order P.D.E. By a first order P.D.E. we mean that we are given an equation of the form

$$F(x, u, u_x) = 0, \tag{7.51}$$

where x represents a point in the n-dimensional manifold M on which the P.D.E. lives, u is the function on this manifold we are trying to solve for, u_x represents all its first derivatives, and F is a smooth function of $2n + 1$ variables. Geometrically the space on which F lives is the first jet space of M ([Arnold, 1983] p. 66). This is a manifold whose points are equivalence classes of functions defined on neighborhoods of points in M. Two functions have the same 1-jet at a point in M iff they have the same value and the same differential there. Let us use the coordinates p to

represent the differential directions of which there are n and u to represent the value direction.

Given a function, the graph of it and its first differential determine an n-dimensional surface in this first jet space. Just as we saw in section 7.1.4 that the property of being the graph of the differential of a function in the cotangent space was equivalent to being a Lagrangian submanifold, being the graph of a function and its differential is locally equivalent to being a *Legendre* manifold in this jet space. Just as there is a natural *symplectic* structure on the cotangent bundle, there is a natural *contact* structure on the first jet space ([Arnold, 1978] p. 349). This may be defined as a smooth choice of a hyperplane in each tangent space of a $(2n+1)$-dimensional manifold, that is locally annihilated by a one-form θ (called a *contact form*), with the property that $\theta \wedge d\theta^n$ is a volume form (here $d\theta$ is a two-form and $d\theta^n$ means $d\theta \wedge \ldots \wedge d\theta$ with n factors). If we think of the jet space as the cotangent bundle with an extra direction tacked on to represent the value of functions, then its natural contact form is a one-form that is the differential of the value coordinate minus the canonical one-form on the cotangent bundle. In the coordinates (x^i, p_i, u), it is given by

$$\theta = du - p_i \, dx^i. \tag{7.52}$$

A *Legendre* submanifold ([Arnold, 1978] p. 365) is an n dimensional submanifold on which the contact form vanishes (the hyperplanes defined by the contact form just contact the surface). It is easy to see that this is exactly the condition that the p coordinates really represent the derivatives of a function whose jet graph is supposed to be the surface in question.

The $2n$-dimensional distribution (smooth field of subspaces chosen from the tangent spaces at each point) defined by the vectors that θ annihilates is *maximally non-integrable*. It would be *integrable* if the contact space could be filled with smooth $2n$-dimensional submanifolds (which together define a *foliation*) that were tangent to the specified planes at each point. It requires very special circumstances that are spelled out in Frobenius' theorem (see [Spivak, 1979] p. 257) for θ to define such $2n$-dimensional surfaces (in fact, θ must be the differential of a function locally for these surfaces to exist). It turns out, however, that there always exists an n-dimensional foliation everywhere tangent to θ's annihilator. For the contact planes, this is in fact the largest dimensional foliation you can find (this is the meaning of maximally non-integrable). The two-form $d\theta$, acting on vectors in a characteristic plane, gives a measure of the nonintegrability there. If $d\theta$ doesn't vanish on two vectors, then it is not possible to deform the parallelogram they form to be tangent to the contact planes. That a contact structure is *maximally nondegenerate* says that $d\theta$ is nondegenerate on each contact plane (i.e. is a symplectic bilinear form on each plane).

The P.D.E. in this picture simply states that the jet graph of the solution function u must lie in the $2n$-dimensional set given by $F = 0$. That an n-dimensional surface is a jet graph of a function says that it is tangent to the contact planes at each of its points. If both of these conditions are satisfied and the surface projects down to M diffeomorphically, then we have a local solution to the P.D.E..

Any such surface must include a certain direction in its tangent plane at each point, called the characteristic direction. This direction is defined as follows. The

$2n$-dimensional tangent plane to the surface $F = 0$ at each point intersects the $2n$ dimensional contact plane generically in a $(2n-1)$-dimensional subspace. When $d\theta$ is restricted to this subspace it has a single degenerate direction (just as in Hamiltonian mechanics, where the degenerate direction of the symplectic structure restricted to an energy level surface gives the direction of the dynamics). This direction is the characteristic direction that must be included in any solution (since $d\theta$ is nondegenerate on the $(n-1)$-dimensional quotient by the characteristic direction and must vanish on our n-dimensional surface, that surface must include the characteristic direction).

It is now easy to see how to solve the Cauchy initial value problem for our P.D.E. The initial surface on which the value of u is given, is an $(n-1)$-dimensional submanifold of M. The initial data plus the constraint that it lie in $F = 0$, determines an $(n-1)$-dimensional submanifold in the jet space which our solution surface must include. As it must also include the characteristic directions, if these are not tangent to the intial manifold, we get a local solution by just flowing the intial manifold along the integral curves of the characteristic direction field (see [Guillemin and Sternberg, 1977] p. 34).

7.2.2. Hamilton-Jacobi Theory and Symplectic Manifolds

In the special case of a first order P.D.E. which does not explicitly depend on the value of the function u, we obtain the Hamilton-Jacobi theory (see [Abraham and Marsden, 1978] p. 381). We saw above that it is exactly this kind of equation that arises from eikonal solutions of linear P.D.E.'s. The whole point to the separation

of scales is that the value of the fast phase itself is irrelevant and only its slow derivatives contribute to the average dynamics in the eikonal limit. It is for this reason that mechanics may be formulated in terms of symplectic manifolds instead of contact manifolds.

In this case, the function $H(x,p)$ whose zero set gives our P.D.E. may be called the Hamiltonian. Since $H = 0$ includes the entirety of the u fibers, and the contact form θ is invariant under translation in the u direction, we may forget about the u direction completely in our theory. What is left is the cotangent bundle T^*M with its canonical one-form and corresponding symplectic form. The characteristic directions project to a line field on the set $H = 0$ in T^*M (since both H and θ were symmetric along u, so were the characteristic directions). The Hamiltonian vector field determined from H and the symplectic structure in the usual way is along this direction. Under this projection, Legendre submanifolds transverse to the u fibers project down to Lagrangian submanifolds. As we have seen in section 7.1.4, this is the condition in T^*M that an n-dimensional submanifold be the graph of the differential of a function. The Cauchy problem now becomes like Hamiltonian mechanics. Given an $(n-1)$-dimensional intitial surface in M with the intial data of u's value on it, we get an $(n-1)$-dimensional initial surface in T^*M as the only submanifold in $H = 0$ consistent with the differential of the intial data. The solution surface is then made up of the integral curves of H's Hamiltonian vector field that pass through the initial manifold.

We saw in section 8.2 that the Hamilton-Jacobi equation for eikonal solutions

of the Klein-Gordon equation is defined by the function

$$H = -k_t^2 + k_x^2 + k_y^2 + k_z^2 + 1 \tag{7.53}$$

on $(x, y, z, t, k_x, k_y, k_z, k_t)$ space. The characteristics are the orbits of the Hamiltonian vector field defined by H. This vector field is given by

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial k_x} = k_x \qquad \dot{y} = \frac{\partial H}{\partial k_y} = k_y \\ \dot{z} &= \frac{\partial H}{\partial k_z} = k_z \qquad \dot{t} = \frac{\partial H}{\partial k_t} = -k_t \\ \dot{k}_x &= -\frac{\partial H}{\partial x} = 0 \qquad \dot{k}_y = -\frac{\partial H}{\partial y} = 0 \\ \dot{k}_z &= -\frac{\partial H}{\partial z} = 0 \qquad \dot{k}_t = -\frac{\partial H}{\partial t} = 0. \end{aligned} \tag{7.54}$$

We only use these curves in the surface $H = 0$.

The physical P.D.E.'s we are often interested in considering (such as the Klein-Gordon example) have a distinguished time direction. The phase space discussed above is really an "extended" phase space in that it includes the time direction. If the Hamiltonian (and so the original P.D.E.) is time independent, we may apply reduction along the time direction to get a symplectic manifold of dimension $2n-2$. Now our intial wave is given over the entire base manifold. This is the picture we have been using in the earlier sections of this chapter. The time evolution of the Lagrangian manifold is given by letting it flow along the Hamiltonian trajectories of the reduced Hamiltonian. As in ordinary mechanics, these trajectories are now important on the whole phase space, not just on a subset like $H = 0$. As we have discussed earlier, the Hamiltonian dynamics can bend the Lagrangian submanifold over and make its projection singular.

For the Klein-Gordon example, we eliminate t and set the generator of t-translation: k_t, to a constant ω. The characteristics are now defined on the full (x, y, z, k_x, k_y, k_z) space and are the orbits of the Hamiltonian vector field of

$$\omega = \sqrt{1 + k_x^2 + k_y^2 + k_z^2}. \tag{7.55}$$

7.2.3. Cotangent Bundles, Contact Spaces, and Jet Spaces

Contact spaces are associated with odd dimensions and symplectic spaces with even dimensions. We have seen that the symplectic cotangent space has a natural generalization to the first jet space, and that the canonical one-form generalizes to the contact form. We may also obtain a contact space of one dimension lower than the cotangent bundle by forgetting about the magnitude of a covector. A covector without its length is given by saying only which vectors it annihilates (and not what it does to other vectors). This is a hyperplane called a contact element in each tangent space of our original manifold (see [Arnold, 1978] p. 354). The space of tangent hyperplanes of a manifold, is itself a contact manifold. The name "contact" makes the most sense here, since these planes represent elements of first order contact with surfaces in the manifold. The contact structure on this space is given quite analogously to the definition of the canonical one-form. The $(2n-1)$-dimensional space of contact elements naturally projects to the base manifold by sending a contact element to the point it is based at. A tangent vector to the space of contact elements is in the contact plane at that point if its projection lies in the contact element it is based at. We will use this in our study of thermodynamics in chapter 16.

7.2.4. The Contact Bundle and the Conormal Bundle

An important class of Lagrangian and Legendre submanifolds arises from submanifolds N of the base manifold M. The contact bundle of N is just the set of all contact elements which are tangent to N. This is clearly a Legendre submanifold since any vector tangent to it projects to a vector tangent to N and so is in the contact plane at that point. The conormal bundle of N is the set of all covectors that annihilate the tangent space of N. This is a Lagrangian submanifold of the cotangent bundle T^*M since a vector tangent to it must project to a vector tangent to N which is annihilated by the form it was based at and so by the canonical one form. Thus we see that the zero section of a cotangent bundle is Lagrangian in one limit (N equals M) and the fibers of the cotangent bundle are as well, in the other limit (N equals a point). If N is a source of light, then the conormal bundle gives the rays that are emanating from N. Thus a point source radiates in all directions while a plane radiates only normal to itself. The Huygens construction gives wavefronts as the envelope of the manifold formed by projecting those points that are a given distance along the rays.

7.3. Limitations of WKB Theory

It is well known that ordinary WKB theory breaks down when diffraction occurs (i.e. where the medium scale length is as small as the wavelength) and near turning points (where the wavelength goes to zero). Keller has developed a beautiful theory of geometric diffraction theory which uses geometric optics (i.e. WKB) away from the bad regions in the medium and glues in the extra rays due to diffraction emanating from these regions using matched asymptotics. The great simplification is that as the wavelength becomes smaller, any "edge" (or any other type of bad region) affects a ray over a smaller and smaller portion of it. Asymptotically, the effect of a discontinuity is identical to one of only a few "canonical problems" (eg. the edge of an infinite half plane, the tip of a cone, etc.). These are solved once and for all and it is their solution that is glued into the problem. Turning point problems were dealt with classically in a similar way in one dimension. Near the turning point the potential asymptotically becomes more and more like a linear potential as far as the asymptotic wave can tell. The exact solution for a linear potential (i.e. an Airy function) is glued in using matched asymptotics again. As we discussed in section 7.1.4, Maslov generalized WKB theory to situations with caustics in higher dimensions. Associated with each of the elementary catastrophes of Thom (which classify the generic caustics) is a special function which is a higher dimensional analog of the Airy function. The wave field around a caustic may be obtained by glueing in these special functions at the caustics of the WKB solution. One important phenomenon that has not to my knowledge received a geometric treatment is tunneling. This is the propagation of real waves into regions that the

WKB rays cannot get to. The problem is that eikonal waves decay exponentially in classically forbidden regions (i.e. regions without rays) as the eikonal parameter vanishes. Straight WKB cannot detect this, but by asymptotically matching solutions on opposite sides of the barrier, one can estimate the transmission coefficient.

As in the case of perturbation theory for mechanical systems, there has been little discussion of the *time* of validity of the WKB asymptotic expansion (i.e. the time-scale on which it is uniform). We saw in the mechanical case that using special techniques such as Kruskal's method could give us expansions uniform over time $1/\epsilon$ but that getting longer times was problematic. Based on the picture of WKB as reduction by an approximate symmetry in wave space, we expect exactly the same phenomena for waves. Long times have become important in recent years as much study has been devoted to *quantum chaos* (i.e. the behavior of eikonal waves when the corresponding rays are chaotic). Chaos is an infinite time concept.

In fact it is easy to construct examples where all the requirements of WKB are satisfied everywhere and for all time and yet the WKB solution becomes invalid on times of order $1/\epsilon$. Jeff Lerner has suggested to me the example of a translation invariant dispersive wave equation with initial condition $A(\epsilon x)e^{ikx}$ where k does not depend on x and $A(x)$ is a single hump. Because k is constant, all rays are parallel and WKB predicts no spreading of the wave packet with time. However, we might have alternatively represented our wave packet in terms of its Fourier transform. Because A varies, the Fourier spectrum will be a smooth hump containing a band of wavevectors centered at k. This shows that the dispersive character of the equation will indeed be noticed by the packet which then in fact will spread with time. Let

us estimate how long it takes for the packet to spread to twice its width. Because Fourier transforms live on the dual space, the peak in k-space has width of order ϵ. The greatest difference in group velocity at different points of the wave packet can then be at most of order ϵ. Thus our packet takes time $1/\epsilon$ to spread 1 unit in x-space. But the packet has width $1/\epsilon$. To spread to twice its width, we must let the packet evolve for time $1/\epsilon^2$. Given a real wave, there is always ambiguity in the choice of splitting into amplitude and wave-vector parts. The differences between these alternative scalings lead to a significant difference in the corresponding WKB predictions on time scales of order $1/\epsilon^2$. This may be seen by an argument exactly analogous to the one used in this example.

Chapter 8: A Hamiltonian Approach to Wave Modulation

"In mechanics the setting is the theory of slow modulations for vibrating systems... The classical theory is usually developed by Hamiltonian methods, which are not directly applicable to waves, but we may instead derive the simplest of the classical results by the methods developed here."— [Whitham, 1974] p. 506.

8.1. Introduction

If we are given a wave system with a Hamiltonian structure, we would like to find a Hamiltonian structure for the evolution of slow modulations of the amplitude and wave number. We want an algorithmic procedure completely independent of any previous knowledge or special features of the system (for example, nothing should depend on linearity). We work out the case of the linear Klein-Gordon equation with this constraint in mind. The same procedure should work for any system: nonlinear, multi-field, integral equations, etc. We wish to extend the beautiful work of Whitham and later Lighthill which is based on *Lagrangian* averaging (see

the discussion and references in the excellent text [Whitham, 1974]) to Hamiltonian systems, which we have seen are more general. [Dubrovin and Novikov, 1983] gives a Hamiltonian treatment for a special class of systems (though their perspective is quite different from ours) and may be of interest to readers as well.

Here we will work with the one-dimensional Klein-Gordon equation:

$$\varphi_{tt} - \varphi_{xx} + \varphi = 0. \tag{8.1}$$

To represent this as a Hamiltonian system we introduce the conjugate field Π with the equations of motion:

$$\begin{aligned} \varphi_t &= \Pi \\ \Pi_t &= \varphi_{xx} - \varphi. \end{aligned} \tag{8.2}$$

This is Hamiltonian with the Poisson bracket

$$\{F, G\} = \int \left(\frac{\delta F}{\delta \varphi} \frac{\delta G}{\delta \Pi} - \frac{\delta F}{\delta \Pi} \frac{\delta G}{\delta \varphi} \right) dx \tag{8.3}$$

and the Hamiltonian

$$H = \int h \, dx, \tag{8.4}$$

where h is the Hamiltonian density

$$h = \frac{1}{2}(\Pi^2 + \varphi_x^2 + \varphi^2). \tag{8.5}$$

8.2. Periodic Solutions

There is a three dimensional submanifold of periodic solutions in this φ, Π space. We may label points in this submanifold by the three coordinates (A, k, θ) via

$$\begin{aligned} \varphi_{A,k,\theta}(x) &\equiv A\sin(kx+\theta) \\ \Pi_{A,k,\theta}(x) &\equiv \sqrt{k^2+1}A\cos(kx+\theta). \end{aligned} \tag{8.6}$$

For convenience we define

$$\omega \equiv \sqrt{k^2+1}. \tag{8.7}$$

The dynamics restricted to this manifold is given by

$$\begin{aligned} \begin{pmatrix} \dot{\varphi}_{Ak\theta}(x) \\ \dot{\Pi}_{Ak\theta}(x) \end{pmatrix} &= \begin{pmatrix} \omega A\cos(kx+\theta) \\ -\omega^2 A\sin(kx+\theta) \end{pmatrix} \\ &= \begin{pmatrix} \omega\frac{d}{d\theta}[A\sin(kx+\theta)] \\ \omega\frac{d}{d\theta}[\omega A\cos(kx+\theta)] \end{pmatrix} \\ &= \omega\frac{d}{d\theta}\begin{pmatrix} \varphi_{Ak\theta}(x) \\ \Pi_{Ak\theta}(x) \end{pmatrix}. \end{aligned} \tag{8.8}$$

Thus the dynamics on A, k, θ space is

$$\begin{pmatrix} \dot{A} \\ \dot{k} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \omega \end{pmatrix}. \tag{8.9}$$

8.2.1. The Hamiltonian Restricted to Periodic Solutions

Now restrict the Hamiltonian to the submanifold and integrate over a large volume V:

$$H_V(\varphi_{Ak\theta}, \Pi_{Ak\theta}) \sim \frac{1}{4}A^2V(k^2+1) + \frac{1}{4}A^2Vk^2 + \frac{1}{4}A^2V = \frac{1}{2}A^2\omega^2V. \tag{8.10}$$

Notice that this is asymptotically true as $V \to \infty$ and neglects an order 1 contribution at the ends. If we were only interested in the periodic case it would perhaps

be more elegant to introduce the average here and not to introduce the V at all. In fact, we are studying the periodic case only as a model for the eikonal case. We want to think of the eikonal waves as limiting on the submanifold of exactly periodic waves as the eikonal parameter vanishes. It is for this reason that we do not take the mean over the whole line and are therefore stuck with the inelegant V's.

8.2.2. The Symplectic Structure Restricted to Periodic Solutions

Poisson brackets can be pushed forward along projections but cannot in general be pulled back along injections like we have here. We therefore work with the symplectic structure, which can be pulled back:

$$\Omega([\delta\varphi_1, \delta\Pi_1], [\delta\varphi_2, \delta\Pi_2]) \equiv \int dx\ (\delta\varphi_1(x)\delta\Pi_2(x) - \delta\Pi_1(x)\delta\varphi_2(x)). \tag{8.11}$$

This two-form is the differential of a one-form α, which is easier to work with:

$$\Omega = -d\alpha, \tag{8.12}$$

$$\alpha([\delta\varphi, \delta\Pi]) \equiv \int \Pi(x)\delta\varphi(x)\ dx. \tag{8.13}$$

To pull this back to (A, k, θ) space, we push forward a vector $(\delta A, \delta k, \delta\theta)$ to $(\delta\varphi, \delta\Pi)$. We only need the $\delta\varphi$ component:

$$\delta\varphi_{Ak\theta}(x) = \delta A\ \sin(kx + \theta) + \delta k\ A\ x \cos(kx + \theta) + \delta\theta\ A \cos(kx + \theta). \tag{8.14}$$

Thus

$$\hat{\alpha}([\delta A, \delta k, \delta\theta]) \equiv \alpha(\delta\varphi_{Ak\theta}, \delta\Pi_{Ak\theta}) \tag{8.15}$$

$$= \int_V A\,\omega\,\cos(kx+\theta)\big[\delta A\,\sin(kx+\theta)+$$
$$+ x\,\delta k\,A\,\cos(kx+\theta)+\delta\theta\,A\,\cos(kx+\theta)\big]dx \qquad (8.16)$$
$$= \frac{V}{2}A^2\,\omega\,\delta\theta.$$

So the two-form ω on A, k, θ space is

$$\tilde{\Omega} = -d\theta = -V\,A\,\omega\,dA\wedge d\theta - \frac{V}{2}\,A^2\,\frac{d\omega}{dk}dk\wedge d\theta. \qquad (8.17)$$

$\tilde{\Omega}$ is closed but degenerate (as it must be in three dimensions), and it annihilates the vector

$$A\,k\,\frac{\partial}{\partial A} - 2\omega^2\frac{\partial}{\partial k}. \qquad (8.18)$$

8.2.2.1. A Degenerate Poisson Structure on the Periodic Solutions

If we choose a function whose level sets are transversal to this vector field, we may make this function a Casimir for a Poisson structure agreeing with ω on the level sets. If, furthermore, the function is a constant of the motion, then H restricted to the level sets must give the correct dynamics. Here we take this function to be k, which we now hold constant. Then on A, θ space,

$$\tilde{\Omega} = V\,A\,\omega\,d\theta\wedge dA \qquad (8.19)$$

and

$$H_V = \frac{V}{2}A^2\omega^2. \qquad (8.20)$$

But

$$dH = A\,V\,\omega^2\,dA, \qquad (8.21)$$

so

$$X_H = \omega \frac{\partial}{\partial \theta}, \tag{8.22}$$

which is the correct dynamics.

8.2.3. The Action of Periodic Orbits

We would like to work in canonical coordinates. From the expression for $\tilde{\Omega}$, we see that θ is conjugate to

$$J \equiv \frac{1}{2} V\ A^2\ \omega. \tag{8.23}$$

In terms of J, θ the structure is quite nice:

$$\begin{aligned} \tilde{\Omega} &= d\theta \wedge dJ \\ H &= J\omega. \end{aligned} \tag{8.24}$$

This gives the dynamics

$$X_H = \omega \frac{\partial}{\partial \theta} \tag{8.25}$$

or equivalently

$$\dot{J} = 0 \quad \dot{\theta} = \omega. \tag{8.26}$$

8.3. Modulations

Modulations essentially "gauge" this theory. We introduce the space of pairs of functions: $(A(x), \theta(x))$. For each value of ϵ we map this into (φ, Π) space via

$$\begin{aligned} \varphi_{\epsilon, A(x), \theta(x)}(x) &\equiv A(\epsilon x)\sin(\frac{1}{\epsilon}\theta(\epsilon x)) \\ \Pi_{\epsilon, A(x), \theta(x)}(x) &\equiv \omega(\epsilon x)\ A(\epsilon x)\ \cos(\frac{1}{\epsilon}\theta(\epsilon x)), \end{aligned} \tag{8.27}$$

where we have defined for convenience

$$k(x) \equiv \theta'(x) \qquad \omega(x) \equiv \sqrt{k^2(x) + 1}. \tag{8.28}$$

This time the dynamics leaves the submanifold invariant only asymptotically as $\epsilon \to 0$ (but does so to all orders). We would like to consider this asymptotic dynamics on (A, θ) space, which represents the modulational equations, as a Hamiltonian system. We do this by pulling back the Hamiltonian and the symplectic structure for each ϵ and then do asymptotics in $\epsilon \to 0$.

8.3.1. Stationary Phase Integrals

The method of stationary phase tells us that, as long as θ' doesn't vanish, integrals of the form

$$\int f(\epsilon x)\ \cos(\frac{1}{\epsilon}\theta(\epsilon x))\ dx \sim 0 \tag{8.29}$$

vanish to all orders in ϵ as $\epsilon \to 0$. This easily implies that

$$\int f(\epsilon x)\ \cos(\frac{1}{\epsilon}\theta(\epsilon x))\ \sin(\frac{1}{\epsilon}\theta(\epsilon x))\ dx \to 0 \tag{8.30}$$

and

$$\int f(\epsilon x)\ \cos^2(\frac{1}{\epsilon}\theta(\epsilon x))\ dx \sim \frac{1}{2}\int f(\epsilon x)\ dx \tag{8.31}$$

to all orders in ϵ provided f is smooth and dies rapidly at infinity.

8.3.2. The Modulational Poisson Brackets

The modulational Poisson brackets are obtained as before by converting (A, θ) perturbations to (φ, Π) perturbations. If we call the map from (A, θ) to (φ, Π), i_ϵ, then

$$\begin{aligned} i_{\epsilon *}[\delta A, \delta\theta] &= \begin{pmatrix} \delta\varphi \\ \delta\Pi \end{pmatrix} \\ &= \begin{pmatrix} \delta A(\epsilon x)\, \sin(\frac{1}{\epsilon}\theta(\epsilon x)) + \delta\theta(x)\, \frac{1}{\epsilon} A(\epsilon x)\, \cos(\frac{1}{\epsilon}\theta(\epsilon x)) \\ \text{not needed} \end{pmatrix}. \end{aligned} \tag{8.32}$$

All sin's and cos's have the argument: $(\theta(\epsilon x)/\epsilon)$. We will leave this argument out of our expressions for clarity. So

$$i_\epsilon^* \alpha([\delta A, \delta\theta]) = \alpha(i_{\epsilon *}[\delta A, \delta\theta]) = \int (\omega\, A\, \cos)(\delta A\, \sin + \delta\theta \frac{1}{\epsilon} A\, \cos) dx. \tag{8.33}$$

By stationary phase, assuming $d\theta \neq 0$ anywhere, we have that to all orders in ϵ:

$$i_\epsilon^* \alpha([\delta A, \delta\theta]) = \frac{1}{\epsilon} \int \frac{1}{2} \omega\, A^2\, \delta\theta\, dx. \tag{8.34}$$

Motivated by the periodic case, we introduce

$$J(x) \equiv \frac{1}{2}\, \omega(x)\, A(x)^2. \tag{8.35}$$

This is the wave action density. Thus the one form is $(1/\epsilon^2) \int J\, \delta\theta\, dx$ (the ϵ is squared since J and $\delta\theta$ are evaluated at ϵx) and the corresponding Poisson bracket is canonical:

$$\left\{F, G\right\} = \epsilon^2 \int dx \left(\frac{\delta F}{\delta\theta} \frac{\delta G}{\delta J} - \frac{\delta F}{\delta J} \frac{\delta G}{\delta\theta} \right). \tag{8.36}$$

8.3.3. The Modulational Hamiltonian

The modulational Hamiltonian is similarly obtained by pulling back H:

$$i^*H = \int dx(\frac{1}{2}\omega^2 A^2 \cos^2 + \frac{1}{2}A^2 k^2 \cos^2 + 2\epsilon A k A' \cos\sin + \epsilon^2 A'^2 \sin^2 + \frac{1}{2}A^2 \sin^2). \tag{8.37}$$

To all orders in ϵ, this is

$$= \int dx(\frac{1}{4}\omega^2 A^2 + \frac{1}{4}A^2 k^2 + \frac{1}{2}\epsilon^2 A'^2 + \frac{1}{4}A^2) = \int dx(\frac{1}{2}\omega^2 A^2 + \frac{1}{2}\epsilon^2 A'^2). \tag{8.38}$$

Or in terms of J:

$$= \int dx(J\omega + \frac{1}{2}\epsilon^2 A'^2). \tag{8.39}$$

8.4. Global Symmetry Implies Local Conservation Law

Let us now show that the presence of a global symmetry implies the existence of a local conservation law. Given a pair of canonically conjugate fields $(J(x), \theta(x))$, so

$$\{F,G\} = \int \left(\frac{\delta F}{\delta \theta} \frac{\delta G}{\delta J} - \frac{\delta F}{\delta J} \frac{\delta G}{\delta \theta} \right) dx \tag{8.40}$$

and a Hamiltonian $H(J,\theta)$ that is invariant under a global change of θ by a constant θ_0 everywhere:

$$\text{i.e.} \quad H(J, \theta + \theta_0) = H(J,\theta), \tag{8.41}$$

we may apply a generalization of Noether's theorem known as reduction. This entails rewriting everything in terms of

$$k(x) \equiv \nabla \theta(x). \tag{8.42}$$

which contains all information in θ except for a constant θ_0. We see

$$\int \frac{\delta G(J,k)}{\delta \theta} f(x)\, dx = \frac{d}{d\beta} G(J, k + \beta \nabla f) \tag{8.43}$$

$$= \int \frac{\delta G}{\delta k} \cdot \nabla f \, dx = - \int \left(\frac{d}{dx} \cdot \frac{\delta G}{\delta k} \right) f(x)\, dx. \tag{8.44}$$

So

$$\frac{\delta G}{\delta \theta} = - \frac{d}{dx} \cdot \frac{\delta G}{\delta k} \tag{8.45}$$

for G's independent of θ_0. Thus the reduced Poisson bracket is

$$\{F,G\} = \int dx \left(\frac{\delta G}{\delta J} \frac{d}{dx} \cdot \frac{\delta F}{\delta k} - \frac{\delta F}{\delta J} \frac{d}{dx} \cdot \frac{\delta G}{\delta k} \right). \tag{8.46}$$

The evolution of J is thus

$$\dot{J}(x) = \{J, H\} = -\frac{d}{dx} \cdot \left(\frac{\delta H}{\delta k}\right). \tag{8.47}$$

This is the desired conservation equation for J, with the flux of J given by $\delta H/\delta k$. Similarly

$$\dot{k}(x) = \{k, H\} = -\frac{d}{dx}\left(\frac{\delta H}{\delta J}\right) = -\frac{d}{dx}\omega(x). \tag{8.48}$$

In our case, to first order in ϵ we have

$$H = \int h \, dx \qquad h(x) = J(x)\, \omega(x), \tag{8.49}$$

which is independent of the value of θ (it depends only on the gradient). We may thus apply the previous theorem to obtain

$$\dot{J}(x) = \{J, H\} = -\frac{d}{dx}\left(J\frac{\partial \omega}{\partial k}\right) = -\frac{d}{dx}(U\ J) \tag{8.50}$$

where we have introduced the group velocity:

$$U \equiv \frac{\partial \omega}{\partial k}. \tag{8.51}$$

Thus the wave action J is conserved and the flux is UJ.

We have therefore successfully obtained the correct modulational equations using a purely Hamiltonian framework. Traditionally, obtaining modulational equations is a very complex task fraught with traps for the unwary. Whitham and Lighthill brought order to this task during the 1960's using Lagrangian methods. A fine account of this work occupies most of the second half of Whitham's book on waves: [Whitham, 1974]. The quote at the beginning of this chapter indicates

that Whitham did not think that this could be accomplished within a Hamiltonian framework. We see that it can indeed. This is of both theoretical interest (particularly because of the explicit connection with the process of reduction) and practical interest. As we discussed in chapter 2, many of the recent systems which have received a Hamiltonian formulation have done so only in the context of Poisson manifolds and therefore have no Lagrangian analog. Our theory should be applicable to waves in these systems as well.

8.5. The Nonlinear Klein-Gordon Equation

This section describes work done in collaboration with Richard Montgomery. Recently, Richard Montgomery has proven that the method described in previous sections of this chapter for the linear Klein-Gordon equation really is quite general and applies to arbitrary systems (preprint, March 1985). Let us here sketch the extension to the nonlinear Klein-Gordon equation:

$$\phi_{tt} = \phi_{xx} - V'(\phi). \tag{8.52}$$

The arguments are almost identical to the linear case. We first introduce a momentum Π conjugate to ϕ:

$$\begin{aligned} \phi_t &= \Pi \\ \Pi_t &= \phi_{xx} - V'(\phi). \end{aligned} \tag{8.53}$$

This system is Hamiltonian with the canonical bracket given in section 8.1 and Hamiltonian

$$H = \int (\frac{1}{2}\Pi^2 + \frac{1}{2}\phi_x^2 + V(\phi))dx. \tag{8.54}$$

We search for periodic functions of the form

$$\tilde{\phi}(x,t) = F(kx + \omega t) \tag{8.55}$$

where F is a function of one variable. Let us denote $kx + \omega t$ by θ. We want F to be 2π-periodic:

$$F(\theta + 2\pi) = F(\theta). \tag{8.56}$$

The corresponding momentum will have the form

$$\tilde{\Pi}(x,t) = \omega F'(\theta) \tag{8.57}$$

(where we use a prime to denote the θ derivative). Substituting this ansatz into the equations of motion gives us a simple O.D.E. for F

$$(\omega^2 - k^2)\frac{\partial^2 F}{\partial \theta^2}(\theta) + V'(F(\theta)) = 0. \tag{8.58}$$

Anytime one searches for travelling wave solutions (or similarity solutions) one obtains a simple O.D.E. even when the original system is a complex nonlinear P.D.E.. In this case the resulting O.D.E. is the equation of motion for a "particle" of "mass" $(\omega^2 - k^2)$ moving in a potential well defined by V if we treat θ as "time". The corresponding "energy" of this particle is

$$\frac{1}{2}(\omega^2 - k^2)\left(\frac{\partial F}{\partial \theta}\right)^2 + V(F) = A \tag{8.59}$$

which we set to the constant value A (because it varies with wave amplitude).

It is interesting to note that the actual energy of the wave is

$$\begin{aligned} H &= \int (\frac{1}{2}\tilde{\Pi}^2 + \frac{1}{2}\tilde{\phi}_x^2 + V(\tilde{\phi}))dx \\ &= \int (\frac{1}{2}\omega^2 \left(\frac{\partial F}{\partial \theta}\right)^2 + \frac{1}{2}k^2 \left(\frac{\partial F}{\partial \theta}\right)^2 + V(F(\theta)))dx \end{aligned} \tag{8.60}$$

which is the integral over x of the "particle" energy. (This is connected with the use of the Klein-Gordon equation in particle physics).

The energy equation gives

$$\left(\frac{\partial F}{\partial \theta}\right)^2 = \frac{2(A - V(F))}{\omega^2 - k^2} \tag{8.61}$$

which may be integrated to give the solution

$$\theta = \frac{1}{2}(\omega^2 - k^2)\int \frac{dF}{\sqrt{A - V(F)}}. \tag{8.62}$$

The constant of integration is the initial phase of the wave. When V is a simple harmonic oscillator, corresponding to the linear system, this gives sin's and cos's as solutions. When V is a pendulum potential (as in the sine-Gordon equation) this is an elliptic integral and when V is quartic we get the solution to the Duffing oscillator. For the linear case we see that the constant A is half the amplitude squared (we continue to use A to facilitate the comparison with Whitham).

The nonlinear dispersion relation is obtained by requiring that F be 2π periodic in θ:

$$\begin{aligned} 2\pi &= \int_0^{2\pi} d\theta \\ &= \int_{\text{period}} \frac{d\theta}{dF} dF \\ &= \frac{1}{2}\sqrt{\omega^2 - k^2} \int_{\text{period}} \frac{dF}{\sqrt{A - V(F)}}. \end{aligned} \tag{8.63}$$

This gives one relation among A, k, and ω. This is analogous to the fact that we could write ω as a function of k in the linear case. Here, though, ω depends on A as well, which leads to characteristically nonlinear effects (just as in a nonlinear oscillator where the frequency depends on the amplitude). The submanifold of periodic solutions is thus three-dimensional again and we may coordinatize it by A, k and θ as in section 8.2.

As in the linear case, we find the action for a periodic solution by integrating the canonical one-form

$$\int \Pi(x)\delta\phi(x)dx \tag{8.64}$$

around a periodic orbit. $\delta\phi(x)$ acting on a unit tangent vector to the orbit will give

$$\frac{1}{\omega}\frac{\partial}{\partial t}\tilde{\phi}(x,t) = \frac{\partial F}{\partial \theta}(kx + \omega t) \tag{8.65}$$

since c moves around its orbit in time $1/\omega$. The total action is therefore

$$\int \omega \frac{\partial F}{\partial \theta}(kx + \omega t)\frac{\partial F}{\partial \theta}(kx + \omega t)dx. \tag{8.66}$$

As in the linear case, we must really integrate only over a volume $\tilde{V}$, divide by $\tilde{V}$, and let $\tilde{V} \to \infty$, leaving us with a well defined action density. Since the integrand is periodic in x, we need only integrate over one period. The action density $J(\ )$ is then

$$\begin{aligned} J(x) &= \frac{1}{(2\pi/k)}\int_0^{2\pi/k} \omega\left(\frac{\partial F}{\partial \theta}(kx+\omega t)\right)^2 dx \\ &= \frac{1}{(2\pi/k)}\int_0^{2\pi} \omega\left(\frac{\partial F}{\partial \theta}(\theta)\right)^2 \frac{dx}{d\theta}d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} \omega\left(\frac{\partial F}{\partial \theta}(\theta)\right)^2 d\theta. \end{aligned} \tag{8.67}$$

Let us denote the average over a period of θ by $\langle\rangle$.

$$\langle\cdot\rangle \equiv \frac{1}{2\pi}\int_0^{2\pi} \cdot d\theta. \tag{8.68}$$

The action density may then be written

$$J(x) = \omega\langle F'^2\rangle. \tag{8.69}$$

We may use the equation satisfied by F to write this as

$$\begin{aligned} J(x) &= \frac{\omega}{2\pi}\int_{\text{period}} \frac{\partial F}{\partial \theta}dF \\ &= \frac{\omega}{2\pi}\int_{\text{period}} \frac{\sqrt{2(A - V(F))}}{\sqrt{\omega^2 - k^2}}dF \\ &= \frac{\omega}{2\pi\sqrt{\omega^2 - k^2}}\int_{\text{period}} \sqrt{2(A - V(F))}dF. \end{aligned} \tag{8.70}$$

When we look at the special case where $V(F) = -\cos F$, this agrees with the expression given in [Forest and McLaughlin, 1982].

Let us now jump right into the modulational equations. We again consider the space of functions $(A(x), \theta(x))$ and for each ϵ, map this into (ϕ, Π) space via

$$\begin{aligned} \phi_{\epsilon, A(x), \theta(x)} &= F_{A(\epsilon x)}(\frac{1}{\epsilon}\theta(\epsilon x)) \\ \Pi_{\epsilon, A(x), \theta(x)} &= \omega(\epsilon x) F'_{A(\epsilon x)}(\frac{1}{\epsilon}\theta(\epsilon x)) \end{aligned} \tag{8.71}$$

where F_A is the periodic F with constant set to A and ω is the known function of $k = \theta_x$ and A. We use F'_A to denote the θ derivative of F_A and $\partial F_A/\partial A$ the A derivative. Exactly as in section 8.3.2, the canonical one-form α is pulled back to (A, θ) space by

$$\begin{aligned} i_\epsilon^* \alpha(|\delta A, \delta\theta|) &= \alpha(i_{\epsilon *}|\delta A, \delta\theta|) \\ &= \int \omega F'_A(\delta A \frac{\partial F_A}{\partial A} + \delta\theta \frac{1}{\epsilon} F'_A) dx. \end{aligned} \tag{8.72}$$

Now we use stationary phase to replace integrals over periodic quantities like F and its derivatives by integrals over the corresponding quantities averaged over $[0, 2\pi]$. We change to the slow scale $X \equiv \epsilon x$:

$$\begin{aligned} i_\epsilon^* \alpha(|\delta A, \delta\theta|) &= \frac{1}{\epsilon^2}\int \omega(X)(F'^2_A)(X)\delta\theta(X)dX + \frac{1}{\epsilon}\int \omega(X)(\frac{\partial F_A}{\partial A}F'_A)(X)\delta A(X)dX \\ &= \frac{1}{\epsilon^2}\int J(X)\delta\theta(X)dX + \frac{1}{\epsilon}\int \omega(X)(\frac{\partial F_A}{\partial A}F'_A)(X)\delta A(X)dX. \end{aligned} \tag{8.73}$$

Similarly, we see that the Hamiltonian is

$$\begin{aligned} H &= \int (\frac{1}{2}\pi^2 + V(\phi) + \frac{1}{2}\phi_x^2)dx \\ &= \int (\frac{1}{2}\omega^2 F'^2_A + V(F) + \frac{1}{2}\theta_x^2 F'^2_A \frac{\partial F_A}{\partial A}A_x + \epsilon^2(\frac{\partial F_A}{\partial A}A_x)^2)dx. \end{aligned} \tag{8.74}$$

Now we use stationary phase again and change to $X = \epsilon x$ to see that to leading

order

$$
\begin{aligned}
H &\sim \frac{1}{\epsilon}\int (\frac{1}{2}(\omega^2+k^2)F_A'^2(X)+V(F_A(X)))dX \\
&= \frac{1}{\epsilon}\int (\frac{1}{2}(\omega^2+k^2)\frac{2(A-V(F_A))}{\omega^2-k^2} - \frac{\omega^2-k^2}{\omega^2-k^2}(A-V(F_A)-A)dX \\
&= \frac{1}{\epsilon}\int (k^2\frac{2(A-V(F_A))}{\omega^2-k^2}+A)dX \\
&= \frac{1}{\epsilon}\int (k^2F_A'^2+A)dX \\
&\sim \frac{1}{\epsilon}\int (k^2(F_A'^2)(X)+A(X))dX \\
&= \frac{1}{\epsilon}\int (\frac{k^2}{\omega(X)}J(X)+A(X))dX.
\end{aligned}
\tag{8.75}
$$

Again we see that H is asymptotically independent of θ. Since J generates θ to leading order (as may be seen by looking at the asymptotic expression for the canonical one-form α), $J(X)$ is asymptotically a locally conserved quantity as for the linear system.

Chapter 9: A Lie Poisson Bracket for Wave Action Density

9.1. Explicit Calculation of the Lie Poisson Bracket

This section represents joint work of Allan Kaufman, Steve McDonald, and myself. J and ψ are real valued functions of $\mathbf{x} \in \Re^3$. We consider the canonical Poisson bracket on functionals of J and ψ:

$$\Big\{ A(J,\psi),\, B(J,\psi) \Big\} = \int \left[\frac{\delta A}{\delta J}(\mathbf{x}'')\, \frac{\delta B}{\delta \psi}(\mathbf{x}'') - \frac{\delta A}{\delta \psi}(\mathbf{x}'')\, \frac{\delta B}{\delta J}(\mathbf{x}'') \right] d^3\mathbf{x}''. \qquad (9.1)$$

We shift attention to functionals of the distribution I on $(\mathbf{x}, \mathbf{k})$ space. We obtain the Poisson bracket of such functionals Υ_1 and Υ_2 by requiring that it reduce to the above canonical bracket on distributions of the form

$$I(\mathbf{x},\mathbf{k}) = J(\mathbf{x})\, \delta^3\big(\mathbf{k} - \nabla\psi(\mathbf{x})\big). \qquad (9.2)$$

This can really only give the value of the wave Poisson bracket evaluated at distributions of this form. We may implicitly assume linearity in I, however, (and the

Lie Poisson bracket is linear) and thereby extend the definition to multiple eikonal waves and then by continuity to all waves. Calling $(\mathbf{x},\mathbf{k}) \equiv y$ and $(\mathbf{x}',\mathbf{k}') \equiv y'$, the chain rule gives

$$\begin{aligned}\Big\{\Upsilon_1(I),\Upsilon_2(I)\Big\} &= \int d^6y\, d^6y'\, \frac{\delta\Upsilon_1}{\delta I}(y)\,\frac{\delta\Upsilon_2}{\delta I}(y')\,\Big\{I(y),I(y')\Big\} \\ &= \int d^6y\, d^6y'\, d^3\mathbf{x}''\, \frac{\delta\Upsilon_1}{\delta I}(y)\,\frac{\delta\Upsilon_2}{\delta I}(y') \\ &\qquad \left[\frac{\delta I(y)}{\delta J}(\mathbf{x}'')\,\frac{\delta I(y')}{\delta\psi}(\mathbf{x}'') - \frac{\delta I(y)}{\delta\psi}(\mathbf{x}'')\,\frac{\delta I(y')}{\delta I}(\mathbf{x}'')\right].\end{aligned} \tag{9.3}$$

We calculate the needed functional derivatives:

$$\begin{aligned}\int \frac{\delta I(y)}{\delta J}(\mathbf{x}'')\cdot f(\mathbf{x}'')\,d^3\mathbf{x}'' &= \lim_{\epsilon\to 0}\frac{I(y)[J+\epsilon f]-I(y)[J]}{\epsilon} \\ &= f(\mathbf{x})\,\delta^3\big(k-\nabla\psi(\mathbf{x})\big).\end{aligned} \tag{9.4}$$

So

$$\frac{\delta I(y)}{\delta J}(\mathbf{x}'') = \delta(\mathbf{x}-\mathbf{x}'')\delta\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big). \tag{9.5}$$

Similarly,

$$\begin{aligned}\int \frac{\delta I(y)}{\delta\psi}(\mathbf{x}'')\cdot\phi(\mathbf{x}'')d^3\mathbf{x}'' &= \lim_{\epsilon\to 0}\frac{I(y)[\psi+\epsilon\phi]-I(y)[\psi]}{\epsilon} \\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}J(\mathbf{x})\Big(\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})-\epsilon\nabla\phi(\mathbf{x})\big)-\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big)\Big) \\ &= -J(\mathbf{x})\nabla\phi(\mathbf{x})\cdot\frac{\partial}{\partial\mathbf{k}}\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big).\end{aligned} \tag{9.6}$$

So

$$\frac{\delta I(y)}{\delta\psi}(\mathbf{x}'') = \frac{\partial}{\partial\mathbf{x}''}\cdot\delta(\mathbf{x}-\mathbf{x}'')J(\mathbf{x}'')\frac{\partial}{\partial\mathbf{k}}\delta\big(\mathbf{k}-\nabla\psi(\mathbf{x}'')\big). \tag{9.7}$$

First substitute in $\frac{\delta I}{\delta J}$ and do the $\mathbf{x}''$ integral:

$$\begin{aligned}\Big\{\Upsilon_1,\Upsilon_2\Big\} = \int d^3\mathbf{x}\,d^3\mathbf{k}\,d^3\mathbf{x}'\,d^3\mathbf{k}'\,d^3\mathbf{x}''\, &\frac{\delta\Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k})\frac{\delta\Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}') \\ &\Big[\delta^3(\mathbf{x}-\mathbf{x}'')\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big)\frac{\delta I(\mathbf{x}',\mathbf{k}')}{\delta\psi}(\mathbf{x}'') \\ &\quad -\delta(\mathbf{x}'-\mathbf{x}'')\delta\big(\mathbf{k}'-\nabla\psi(\mathbf{x}')\big)\frac{\delta I(\mathbf{x},\mathbf{k})}{\delta\psi}(\mathbf{x}'')\Big].\end{aligned} \tag{9.8}$$

Thus

$$\left\{\Upsilon_1, \Upsilon_2\right\} = \int d^3\mathbf{x}\, d^3\mathbf{k}\, d^3\mathbf{x}'\, d^3\mathbf{k}'\ \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k})\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}') \left[\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big)\frac{\delta I(\mathbf{x}',\mathbf{k}')}{\delta \psi}(x) - \delta^3\big(\mathbf{k}'-\nabla'\psi(\mathbf{x}')\big)\frac{\delta I(\mathbf{x},\mathbf{k})}{\delta \psi}(\mathbf{x}).\right. \tag{9.9}$$

Now substitute in $\frac{\delta I}{\delta \psi}$:

$$= \int d^3\mathbf{x}\, d^3\mathbf{k}\, d^3\mathbf{x}'\, d^3\mathbf{k}'\ \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k})\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}') \left[\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big)\frac{\partial}{\partial \mathbf{x}}\cdot\delta^3(\mathbf{x}'-\mathbf{x})J(\mathbf{x})\frac{\partial}{\partial \mathbf{k}'}\delta^3\big(\mathbf{k}'-\nabla\psi(\mathbf{x})\big)\right. - \delta^3\big(\mathbf{k}'-\nabla\psi(\mathbf{x}')\big)\frac{\partial}{\partial \mathbf{x}'}\cdot\delta^3(\mathbf{x}-\mathbf{x}')J(\mathbf{x}')\frac{\partial}{\partial \mathbf{k}}\delta^3\big(\mathbf{k}-\nabla'\psi(\mathbf{x}')\big)\Big]. \tag{9.10}$$

Now integrate by parts with respect to $\mathbf{k}'$ in the first term and $\mathbf{k}$ in the second:

$$= \int d^3\mathbf{x}\, d^3\mathbf{k}\, d^3\mathbf{x}'\, d^3\mathbf{k}'\ \left[-\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k})\left(\frac{\partial}{\partial \mathbf{k}'}\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}')\right)\right. \delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big)\cdot\frac{\partial}{\partial \mathbf{x}}\delta^3(\mathbf{x}'-\mathbf{x})J(\mathbf{x})\delta^3\big(\mathbf{k}'-\nabla\psi(\mathbf{x})\big) + \left(\frac{\partial}{\partial \mathbf{k}}\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k})\right)\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}') \delta^3\big(\mathbf{k}'-\nabla\psi(\mathbf{x}')\big)\cdot\frac{\partial}{\partial \mathbf{x}'}\delta^3(\mathbf{x}-\mathbf{x}')J(\mathbf{x}')\delta^3\big(\mathbf{k}-\nabla'\psi(\mathbf{x}')\big)\Big]. \tag{9.11}$$

Now change the variables of integration: exchange $\mathbf{k}$ and $\mathbf{k}'$ in the first term and $\mathbf{x}$ and $\mathbf{x}'$ in the second:

$$= \int d^3\mathbf{x}\, d^3\mathbf{k}\, d^3\mathbf{x}'\, d^3\mathbf{k}'\left[-\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k}')\left(\frac{\partial}{\partial \mathbf{k}}\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k})\right)\right. \delta^3\big(\mathbf{k}'-\nabla\psi(\mathbf{x})\big)\cdot\frac{\partial}{\partial \mathbf{x}}\delta^3(\mathbf{x}'-\mathbf{x})J(\mathbf{x})\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big) + \left(\frac{\partial}{\partial \mathbf{k}}\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x}',\mathbf{k})\right)\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k}') \delta^3\big(\mathbf{k}'-\nabla\psi(\mathbf{x})\big)\cdot\frac{\partial}{\partial \mathbf{x}}\delta^3(\mathbf{x}'-\mathbf{x})J(\mathbf{x})\delta^3\big(\mathbf{k}-\nabla\psi(\mathbf{x})\big)\Big]. \tag{9.12}$$

Collect terms:

$$= \int d^3\mathbf{x}\, d^3\mathbf{k}\, d^3\mathbf{x}'\, d^3\mathbf{k}' \left[-\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k}') \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}) \right) + \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x}',\mathbf{k}) \right) \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k}') \right] \cdot \delta^3(\mathbf{k}' - \nabla\psi(\mathbf{x})) \frac{\partial}{\partial \mathbf{x}} \delta^3(\mathbf{x}' - \mathbf{x}) J(\mathbf{x}) \delta^3(\mathbf{k} - \nabla\psi(\mathbf{x})). \tag{9.13}$$

Now recall $J(\mathbf{x})\delta^3(\mathbf{k} - \nabla\psi(\mathbf{x})) = I(\mathbf{x},\mathbf{k})$ and integrate by parts in $\mathbf{x}$:

$$\begin{aligned} = \int d^3\mathbf{x}\, d^3\mathbf{k}\, d^3\mathbf{x}'\, d^3\mathbf{k}' \Bigg\{ & \left[\left(\frac{\partial}{\partial \mathbf{x}} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k}') \right) \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}) \right) - \right. \\ & \left. - \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x}',\mathbf{k}) \right) \left(\frac{\partial}{\partial \mathbf{x}} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k}') \right) \right] \\ & \delta^3(\mathbf{k}' - \nabla\psi(\mathbf{x})) \delta^3(\mathbf{x} - \mathbf{x}') I(\mathbf{x},\mathbf{k}) + \\ & \left[-\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k}') \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}) \right) + \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x}',\mathbf{k}) \right) \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k}') \right] \\ & \times \nabla\nabla\psi \cdot \frac{\partial}{\partial \mathbf{k}'} \delta^3\left(\mathbf{k}' - \frac{\partial}{\partial \mathbf{x}}\psi(\mathbf{x})\right) \delta^3(\mathbf{x}' - \mathbf{x}) I(\mathbf{x},\mathbf{k}) \Bigg\}. \end{aligned} \tag{9.14}$$

In the first term, the $\delta^3(\mathbf{k} - \nabla\psi(\mathbf{x}))$ in $I(\mathbf{x},\mathbf{k})$ lets us replace $\delta^3(\mathbf{k}' - \nabla\psi(\mathbf{x}))$ by $\delta^3(\mathbf{k}' - \mathbf{k})$. In the second term we integrate by parts in $\mathbf{k}'$ and then do the same replacement.

$$\begin{aligned} \left\{ \Upsilon_1, \Upsilon_2 \right\} = \int d^3\mathbf{x}\, d^3\mathbf{k}\, d^3\mathbf{x}'\, d^3\mathbf{k}' \Bigg\{ & \left[\left(\frac{\partial}{\partial \mathbf{x}} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k}') \right) \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}) \right) \right. \\ & \left. - \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x}',\mathbf{k}) \right) \left(\frac{\partial}{\partial \mathbf{x}} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k}') \right) \right] \delta^3(\mathbf{k}' - \mathbf{k}) \delta^3(\mathbf{x} - \mathbf{x}') I(\mathbf{x},\mathbf{k}) \\ & + \left[\left(\frac{\partial}{\partial \mathbf{k}'} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k}') \right) \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x}',\mathbf{k}) \right) \right. \\ & \left. - \left(\frac{\partial}{\partial \mathbf{k}} \frac{\delta \Upsilon_1}{\delta I}(\mathbf{x}',\mathbf{k}) \right) \left(\frac{\partial}{\partial \mathbf{k}'} \frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k}') \right) \right] : \nabla\nabla\psi \delta^3(\mathbf{k}' - \mathbf{k}) \delta^3(\mathbf{x}' - \mathbf{x}) I(\mathbf{x},\mathbf{k}) \Bigg\}. \end{aligned} \tag{9.15}$$

We do the $\mathbf{x}'$ and $\mathbf{k}'$ integrals using the delta functions. The second term vanishes

and the first becomes the Lie-Poisson bracket:

$$\begin{aligned}\left\{\Upsilon_1(I),\Upsilon_2(I)\right\} &= \int d^3\mathbf{x}\, d^3\mathbf{k}\; I(\mathbf{x},\mathbf{k})\left[\left(\frac{\partial}{\partial \mathbf{x}}\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k})\right)\cdot\left(\frac{\partial}{\partial \mathbf{k}}\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k})\right)\right. \\ &\qquad \left.-\left(\frac{\partial}{\partial \mathbf{k}}\frac{\delta \Upsilon_1}{\delta I}(\mathbf{x},\mathbf{k})\right)\cdot\left(\frac{\partial}{\partial \mathbf{x}}\frac{\delta \Upsilon_2}{\delta I}(\mathbf{x},\mathbf{k})\right)\right] \\ &= \int d^6y\; I(y)\left[\frac{\delta \Upsilon_1}{\delta I},\frac{\delta \Upsilon_2}{\delta I}\right]_y.\end{aligned} \tag{9.16}$$

9.2. The Geometrical Picture

Alan Weinstein has given the following geometric interpretation of this result. On the one hand, as in section 2.7.10 we may consider the group of symplectomorphisms of a symplectic manifold M. An infinitesimal symplectomorphism is represented by a Hamiltonian vector field (at least locally). We may think of the Lie algebra of the group of symplectomorphisms as being all Hamiltonian vector fields on M with the Lie bracket of vector fields being the Lie algebra bracket (perhaps with a minus sign, depending on conventions). Equivalently (up to a constant), we may view it as the space of functions (Hamiltonians) on M with the Lie algebra bracket being the Poisson bracket of two functions (defined using M's symplectic structure). The coadjoint action of a symplectomorphism on such a function is simply given by pullback. The dual of the Lie algebra is then the space of distributions on M. We therefore have the natural Lie-Poisson bracket (given above) on functionals of such distributions.

The coadjoint action of a symplectomorphism on a distribution is given by pushing the distribution forward along the symplectomorphism. The coadjoint orbits, which have a natural KKS symplectic structure, consist of all distributions obtainable from a given one by canonical transformations. In particular, the orbit of a δ-function looks just like M and has the same symplectic structure (points of M correspond to the δ-function at that point). Lagrangian submanifolds (i.e. half-dimensional submanifolds on which the symplectic form vanishes) of M are taken to other Lagrangian submanifolds under canonical transformations (since the symplectic form is preserved). Locally, any small enough piece of a Lagrangian sub-

manifold can be taken to a corresponding piece of any other Lagrangian submanifold by a canonical transformation. So up to global issues, the space of measured Lagrangian submanifolds on M is itself a symplectic manifold (see section 2.7.10 for more details).

Above, we saw that the phase function which determines the manifold and the amplitude which determines the distribution on that manifold, may be treated as canonically conjugate variables in a way that is consistent with the Lie-Poisson structure on all distributions. We would like to understand this fact geometrically. We will show that the cotangent bundle of the space of Lagrangian submanifolds of M may be naturally identified with the space of distributions with Lagrangian support. The cotangent space at a given Lagrangian submanifold is identifiable with the the space of distributions defined on that Lagrangian submanifold. This is then identifiable with the corresponding space of δ-like distributions on M supported on that manifold (see section 2.7.10). Thus the canonical conjugacy of the distribution and the manifold arises in a natural way.

What is the tangent space to a Lagrangian submanifold of M in the space of Lagrangian submanifolds? A tangent vector will be a small deformation of the manifold which is itself Lagrangian. All such deformations come from Hamiltonian vector fields. The tangent space may thus be identified with the space of Hamiltonian vector fields on M modulo those which leave the Lagrangian submanifold invariant. A Hamiltonian vector field leaves a Lagrangian manifold invariant if and only if it is constant on it.

We see this as follows. If a Hamiltonian vector field X_H is tangent to our

Lagrangian manifold, then $dH = \omega(X_H, \cdot)$ must annihilate all tangent vectors to the manifold, since it is Lagrangian and so its tangent spaces are symplectically orthogonal to themselves. This says that H is constant on it (if it is connected). If H is constant on the Lagrangian submanifold, then it annihilates all tangent vectors. Thus X_H is symplectically orthogonal to the whole tangent space. But on a Lagrangian submanifold, the only such vectors are themselves tangent to it. Thus X_H preserves the Lagrangian manifold.

We may therefore identify the tangent space to the space of Lagrangian submanifolds at a given such manifold with the space of functions on M modulo those functions which vanish on the manifold (again ignoring constants). But this is exactly the space of functions on the Lagrangian submanifold. Taking the dual we see that the cotangent space to the space of Lagrangian submanifolds at a given Lagrangian manifold is indeed isomorphic to the space of distributions on that manifold.

This space of distributions is identifiable with the delta-like distributions on M supported on the manifold. The cotangent bundle of the space of Lagrangian submanifolds obtainable by deforming a given one is thus isomorphic to the coadjoint orbit of the symplectomorphism group which contains any delta-like distribution supported by the given Lagrangian submanifold. The KKS coadjoint orbit symplectic structure is exactly the canonical structure obtained from looking at the orbit as a cotangent bundle.

PART III:

DISSIPATION

and

STATISTICS

"These two branches, mechanics and thermodynamics, can be joined only from a higher standpoint, that of the statistical mechanics of molecular systems."—[Klein, 1928] p. 203

Chapter 10: Imbedding and Projection Theorems

In this chapter we will demonstrate a number of relations between dissipative and non-dissipative systems. We will show, in particular, that any system can be embedded in a Hamiltonian system of twice the dimension and in a Poisson system of only one dimension more. These constructions are of interest because they help to delineate what operations lead one to artificial structures. Since they can arise from any system, the Hamiltonian structures of this section cannot be expected to give any new insights into the original system. They may be useful in understanding how dissipative dynamics may arise from an underlying Hamiltonian system, however, since this always involves some kind of projection. We will also give examples of some seemingly harmless "dangerous operations" which can completely destroy the physical content of a model.

10.1. Imbedding in a Hamiltonian System

Consider an arbitrary dynamical system given by a vector field X on a manifold Q. We will construct a Hamiltonian system of twice the dimension which has an invariant submanifold diffeomorphic to Q and on which the restricted dynamics

is exactly that given by X. The manifold will be the cotangent bundle of Q, i.e. T^*Q. The dynamics will be the cotangent lift of X to T^*Q. This is defined as follows: the flow of X is a one parameter family of diffeomorphisms of Q to itself. A one-form may be pushed forward along a diffeomorphism (just pull it back along the inverse of the diffeomorphism). Thus at every moment of time each point in T^*Q is mapped to another such point. Taking these together gives us a natural one-parameter family of diffeomorphisms from T^*Q to itself which covers the flow of X (this is a point transformation). This family is actually the flow of a vector field on T^*Q which is Hamiltonian with respect to the natural symplectic structure on T^*Q. The Hamiltonian is the function on T^*Q given by

$$H(q,p) \equiv \langle p, X(q) \rangle. \tag{10.1}$$

Notice that H is linear on each fiber of T^*Q and that the zero section is in the zero set of H. It is easy to check that this gives the correct dynamics. With the symplectic structure

$$\omega = dq \wedge dp, \tag{10.2}$$

the dynamics is

$$\begin{aligned} \dot{q}^i &= X^i(q) \\ \dot{p}^i &= -p^j \frac{\partial X^j(q)}{\partial q^i}. \end{aligned} \tag{10.3}$$

The zero section is indeed invariant (since if p starts out zero, it remains so) and the dynamics restricted to it is the original dynamics. Thus by restricting the class of initial conditions, we may have dynamics inside a Hamiltonian system looking like a dissipative system of any type.

Notice further that under the natural projection

$$\pi : T^*Q \to Q \quad \text{by} \quad (q,p) \mapsto q, \tag{10.4}$$

the Hamiltonian dynamics projects down to the given dynamics. Thus if we observe only certain variables in a Hamiltonian system, it may look just like a dissipative system of any type.

Linear Hamiltonian vector fields must have their eigenvalues ditributed in the complex plane so as to be symmetric under reflection about both the real and imaginary axes. Thus if we start with a linear attracting fixed point, like

$$X^i(q) = -q^i, \tag{10.5}$$

which has all of its eigenvalues in the left half-plane, we must double the dimension to imbed it in a Hamiltonian system. Thus the above construction is the smallest Hamiltonian space we might construct in general.

10.2. Projection from a Hamiltonian System

This argument for a lower bound on dimension does not apply to the case of projection, however. A fixed point of projected dynamics need not be a fixed point of the unprojected dynamics, so the eigenvalue argument doesn't hold. In fact, we may construct a Hamiltonian system that projects to our system by adding only a single dimension if the original space is odd dimensional, or two more dimensions if it is even dimensional. The construction sheds light on a dangerous operation one sometimes sees being performed in the physics literature. If one has unbounded motion, one may smoothly untwist the orbits and hide any features of the dynamics that one desires.

Let us extend our given space Q by a single dimension to get $Q \times \Re$. Let us denote the time t flow of the given dynamics X on Q by $\mathcal{F}_t$. It is quite natural to consider the extra dimension in $Q \times \Re$ as time and to consider the original dynamics on Q augmented by a time which flows uniformly:

$$\tilde{X}(q,t) \equiv (X(q), 1). \tag{10.6}$$

Because this has no fixed points and no recurrent orbits, we may now perform the following seemingly harmless operation. We define a diffeomorphism from $Q \times \Re$ to itself by

$$(q,t) \mapsto (\mathcal{F}_{-t}(q), t). \tag{10.7}$$

By the definition of flow,

$$\frac{\partial}{\partial t}\mathcal{F}_t(q) = X(q), \tag{10.8}$$

so the image of our dynamical vector field $\tilde{X}$ under this diffeomorphism is simply:

$$(0,1). \tag{10.9}$$

Thus we have completely trivialized the dynamics by a coordinate change that unraveled the original orbits. In particular all dynamics can be made to look the same on extended state spaces. For our Hamiltonian construction, we may choose a symplectic structure and Hamiltonian that give the trivial dynamics of (10.9) and pull them back to the original extended state space by our diffeomorphism. There they give the extended dynamics as a Hamiltonian system and so project onto the original dynamics. The upshot is that one must beware of time-dependent coordinate changes because they can easily do many non-physical things, such as making any system look Hamiltonian.

10.3. Dangerous Operations with Unbounded Variables

This same construction is sometimes used in disguise to show that all dynamical systems (or any particular system an author is interested in) are integrable. Through a series of suitably complex coordinate changes, the author succeeds in adding an extra variable with unbounded dynamics (the next two sections present explicit examples). This is then used as above to trivialize the system, where constants of the motion are pi.ntiful. The constants are pulled back and appear magically to have simplified the original system.

Other dangerous things one can do are to make attractive systems look repulsive (as in figure (10.1)) and in fact get Liapunov exponents to be anything one wishes. (Oseledec's theorem shows that they are well defined for bounded systems, in the sense that all smooth metrics give them the same values as discussed on p. 284 of [Guckenheimer and Holmes, 1983], but they have no intrinsic meaning for unbounded systems).

10.3.1. Eg.: Surreptitiously Changing Damping to Driving

Let us convert a damped harmonic oscillator to one with negative friction. We start with the system

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -x - kv. \end{aligned} \tag{10.10}$$

We make the change of coordinates

$$\begin{aligned} \tilde{x} &= e^{kt}x \\ \tilde{v} &= e^{kt}v + ke^{kt}x. \end{aligned} \tag{10.11}$$

We see that the new coordinates satisfy the equations

$$\begin{aligned}
\frac{d}{dt}\tilde{x} &= ke^{kt}x + e^{kt}\dot{x} \\
&= ke^{kt}x + e^{kt}v \\
&= \tilde{v} \\
\frac{d}{dt}\tilde{v} &= ke^{kt}v + e^{kt}\dot{v} + k^2e^{kt}x + ke^{kt}\dot{x} \\
&= ke^{kt}v - e^{kt}x - ke^{kt}v \\
&= -\tilde{x} + k\tilde{v}
\end{aligned} \tag{10.12}$$

as desired.

In perturbation theory, if one makes time dependent coordinate changes, one must make certain that they remain close to the identity for large times (for example by requiring that they be periodic or quasiperiodic in time) or else one may sweep any undesirable dynamics under the rug.

10.3.2. Eg.: Pitfalls in the Use of Lie Transforms

Let us illustrate some of these dangers with a simple example using the method of Lie transforms. It is becoming fashionable to use Lie transforms in very complex situations, where it is hard to keep one's physical intuition about the problem (eg. in systems with many-particle dynamics or even infinite-dimensional wave systems). It is therefore worthwhile to point out some pitfalls for the unwary in a simple example. Such examples will hopefully help us to avoid misapplying Lie transforms in more complex circumstances.

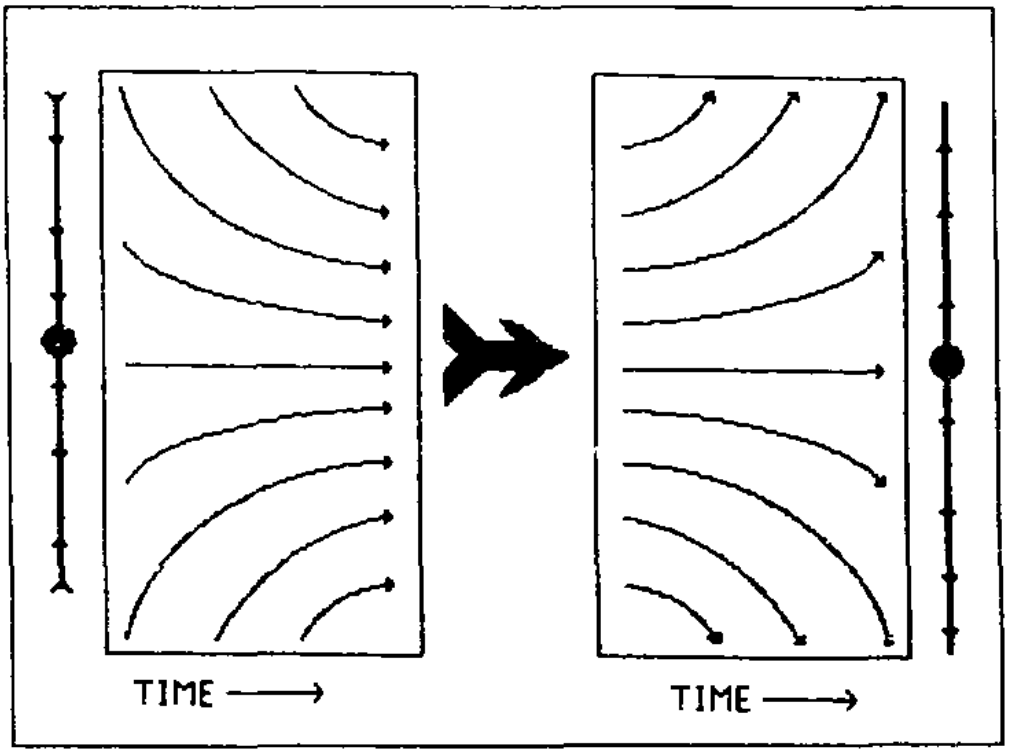

Figure 10.1: Attraction changed to repulsion by a time dependent change of coordinates.

In this method we assume given a Hamiltonian as an asymptotic series in ϵ whose zero order term has been brought to action-angle form and depends only on the actions. Let us examine the example:

$$H = H_0 + \epsilon H_1 = \omega_1 J_1 + \omega_2 J_2 + \epsilon \cos(\theta_1 - \theta_2), \tag{10.13}$$

on the phase space $\theta_1, J_1, \theta_2, J_2$ with the canonical bracket: $[,]$. The method of Lie transforms seeks to find an ϵ-dependent canonical transformation which becomes the identity as $\epsilon \to 0$ and which converts H to a new Hamiltonian K, which depends only on the actions. We represent the canonical transformation as the time-one flow generated by the ϵ-dependent Hamiltonian:

$$-W = \epsilon W_1 + \epsilon^2 W_2 + \dots. \tag{10.14}$$

$-W$'s Hamiltonian vector field, viewed as a differential operator on functions on phase space, is denoted by

$$L = [W, \cdot]. \tag{10.15}$$

The action of the flow on functions is then the operator:

$$e^L. \tag{10.16}$$

The transformed Hamiltonian is then given by

$$K = e^L \cdot H = H_0 + \epsilon([W_1, H_0] + H_1) + \dots. \tag{10.17}$$

We want to choose W order by order to eliminate the θ dependence in K. At first order we would like to solve for W_1:

$$[W_1, H_0] = -H_1. \tag{10.18}$$

This just says that the derivative of W_1 along the unperturbed dynamics (generated by H_0) should be equal to $-H_1$. One sometimes sees the solutions written formally:

$$W_1 = -\int_{\text{unperturbed orbits}} H_1. \tag{10.19}$$

This leads one to think that the transformation is always possible and that the qualitative behavior of $K(\tilde{J}_1, \tilde{J}_2)$'s dynamics, which has all orbits periodic or quasi-periodic, is representative of the behavior of H's. In the example we are considering, this is indeed true if ω_1 and ω_2 are irrationally related. In this case, the Lie transform remains close to the identity for all time (and gets closer as $\epsilon \to 0$) and so the change of variables doesn't do much damage.

Consider the case $\omega_1 = \omega_2$, however. For this system

$$\dot{\theta}_1 = \omega_1 = \dot{\theta}_2, \tag{10.20}$$

and so $\theta_1 - \theta_2$ is constant in time. This means that

$$\dot{J}_1 = \epsilon \sin(\theta_1 - \theta_2) \tag{10.21}$$

is constant as well, and so J_1 grows steadily in time. This is in great variance with the prediction of K which says that J_1 remains constant in time. How is it that by a change of variables we converted a system with monotonically changing action to one with constant action? The seemingly harmless integral over unperturbed orbits created a canonical transformation which was unbounded in time. As we have discussed, such transformations allow one to convert any system to any other, regardless of the actual physics involved. As this example demonstrates, one must be especially careful when dealing with systems that have resonances (different unperturbed degrees of freedom with the same frequency or with rationally related frequencies). Notice that in this example, the error over bounded time is only of size ϵ and so we have done as well as naïve perturbation theory. The whole point of Lie transforms, though, is to get behavior on a time-scale of order $1/\epsilon$. On this scale the method has failed in the example above.

10.4. Imbedding in Poisson Systems

Because Hamiltonian dynamics on Poisson manifolds is more general than on symplectic manifolds, we may imbed arbitrary dynamics into Poisson dynamics of only one dimension higher. As above, assume given an arbitrary manifold Q and dynamical vector field X. We construct the manifold $\Re \times Q$ (this is really $(T^*\Re \times Q)/\Re$ as discussed in section 1.5, question 66). Let h be a function on $\Re$ defined by

$$h(x) = x. \tag{10.22}$$

Define the vertical vector field Y on $\Re \times Q$ to be

$$Y = \frac{\partial}{\partial h}. \tag{10.23}$$

We will use h pulled up to $\Re \times Q$. We may define a Poisson bracket on $\Re \times Q$ as follows (this Poisson structure may be viewed as the bivector $X \wedge Y$):

$$\{f, g\} \equiv (X \cdot f)(Y \cdot g) - (X \cdot g)(Y \cdot f). \tag{10.24}$$

This satisfies the Jacobi identity, as may be routinely checked:

$$
\begin{aligned}
&\{f,\{g,h\}\}+\{g,\{h,f\}+\{h,\{f,g\}\}= \\
&Xf\,Y(Xg\,Yh-Xh\,Yg)-X(Xg\,Yh-Xh\,Yg)Yf \\
&+Xg\,Y(Xh\,Yf-Xf\,Yh)-X(Xh\,Yf-Xf\,Yh)Yg \\
&+Xh\,Y(Xf\,Yg-Xg\,Yf)-X(Xf\,Yg-Xg\,Yf)Yh \\
&=Xf\,YXg\,Yh+Xf\,Xg\,YYh-Xf\,YXh\,Yg-Xf\,Xh\,YYg \\
&-XXg\,Yh\,Yf-Xg\,XYh\,Yf+XXh\,Yg\,Yf+Xh\,XYg\,Yf \\
&+Xg\,YXh\,Yf+Xg\,Xh\,YYf-Xg\,YXf\,Yh-Xg\,Xf\,YYh \\
&-XXh\,Yf\,Yg-Xh\,XYf\,Yg+XXf\,Yh\,Yg+Xf\,XYh\,Yg \\
&+Xh\,YXf\,Yg+Xh\,Xf\,YYg-Xh\,YXg\,Yf-Xh\,Xg\,YYf \\
&-XXf\,Yg\,Yh-Xf\,XYg\,Yh+XXg\,Yf\,Yh+Xg\,XYf\,Yh \\
&=0
\end{aligned}
\tag{10.25}
$$

since the Lie bracket of X and Y:

$$[X,Y]=XY-YX=0, \tag{10.26}$$

since Y is constant. Because

$$Y\cdot h=\frac{\partial}{\partial h}h=1, \tag{10.27}$$

we see that for any f

$$\{f,h\}=X\cdot f. \tag{10.28}$$

Thus with this Poisson bracket and h as Hamiltonian we obtain the original dynamics given by X. Here, each of the level sets of h has the original dynamics and so

both each injected level set and the projection along $\Re$ give the original X dynamics on Q. The symplectic leaves are the products of $\Re$ and the non-fixed-point orbits of X. Each fixed point of X yields an interval of point symplectic bones along $\Re$. We have shown that by adding one extra dynamical variable, we can make any system into a Hamiltonian system on a Poisson manifold.

Let us give an explicit example. The simplest dissipative dynamical system is given by

$$\dot{x} = -x, \tag{10.29}$$

where x is a point on the real line. This type of system is used as a model for linear relaxation in non-equilibrium thermodynamics, where $x = 0$ is x's equilibrium value. If we introduce another variable y, then the (x, y) plane is a Poisson manifold with a Poisson bracket given by

$$f, g = x\left(\frac{\partial f}{\partial y}\frac{\partial g}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial g}{\partial y}\right). \tag{10.30}$$

If we consider the Hamiltonian

$$H = y, \tag{10.31}$$

then the Hamiltonian dynamics is

$$\dot{f} = f, H = -x\frac{\partial f}{\partial x}. \tag{10.32}$$

This gives the dynamical equations

$$\dot{x} = -x \qquad \dot{y} = 0. \tag{10.33}$$

The original dissipative system: $\dot{x} = -x$, is both imbedded as the dynamics on any of the submanifolds: y =constant, and the result of projecting along y (i.e. forgetting the value of y).

Chapter 11: Projected Area and Canonical Transformations

Tourist: Can you give me the directions to Omaha?

Farmer, scratching his head: You can't get there from here.—Anon.

Much of this thesis has been about simplifying the description of physical systems by projecting their dynamics down to various subspaces. A classical theorem of Liouville states that the dynamics of a Hamiltonian system preserves a certain canonical volume element in phase space. In modern parlance this theorem is very easy to prove. The Hamiltonian flow preserves a symplectic structure ω on phase space (i.e. it is equal to its pullback along the flow for any time). Since the operation of pullback of differential forms commutes with the operation of wedging them together, the Hamiltonian flow also preserves $\omega \wedge \omega$, $\omega \wedge \omega \wedge \omega$, ..., and finally in N dimensions: ω^N. It is easy to see that this last is a volume element, and in fact is the generalization to manifolds of the one described by Liouville. This volume preservation property of Hamiltonian flows has many significant consequences for physical systems. It is the key ingredient for the Poincaré recurrence theo-

rem [Arnold, 1978]and is responsible for the absence of attractors in Hamiltonian systems. It is also the basis for much of statistical mechanics.

11.1. Application to Particle Accelerators

When we project our dynamics to a smaller space, the volume preservation is in general lost, but some remnant remains to impose constraints on the projected dynamics. There are many physical situations where these constraints lead to very interesting consequences. Andy Sessler has described to me the relevance to the design of free-electron lasers of any theorems constraining the ability of a Hamiltonian system to change the volume of projection of a region. The time evolution of the particles moving through an FEL is often described by linearizing about some known orbit, leading to a time dependent linear canonical transformation on single particle phase space. If the particles in the beam are sufficiently noninteracting, the device applies this Hamiltonian transformation to all particles simultaneously. The phase space for a particle is a product of a 2-dimensional longitudinal phase space and a 2 or 4 dimensional transverse phase space. The source of the beam going into the FEL typically produces a particle distribution with longitudinal and transverse distributions uncoupled, and a uniform spread over some range of momenta and some range of positions. Thus we can think of the incoming beam as a uniformly filled in parallelepiped lined up along the transverse and longitudinal phase spaces. We send these particles through the FEL and get a distorted and bent over parallelepiped in phase space out from the other end. Quite often we are more interested in some of the phase space coordinates than the other ones. For instance, we may

want a beam with very small transverse spread in space and momentum, while not caring about the longitudinal spatial spread, to get a well defined beam axis. (Or sometimes the other way around if we want well defined buckets to increase the efficiency of energy transfer between a wave and the beam). While we know that the total volume in the parallelepiped must remain constant, we are interested in the possibility of trading some transverse phase space volume for longitudinal phase space volume.

11.1.1. Courant's Theorem

[Courant, 1966] has looked at this question and arrived at the following theorem: "An ellipsoid in phase space whose principal axes are the canonical coordinate and momentum axes can be transformed by a linear canonical transformation into another such normally oriented ellipsoid only if the areas of the projections of the first ellipsoid on each of the (q_i, p_i) planes are separately equal, one by one, to the corresponding projections of the second ellipsoid. The transformation is then the direct sum of N separate area-preserving, two-dimensional transformations." This statement of his theorem assumes that all the projected areas are different. When two are the same, the symplectic transformation may couple them. It is also important to note that the "corresponding" projections that have the same area need not be along the same sets of axes. In particular, we may exchange two sets of axes. This may be all that is needed for certain circumstances in the accelerator setting. If the phase space is (q_1, p_1, q_2, p_2), then the Hamiltonian

$$H = q_1 p_2 - q_2 p_1, \tag{11.1}$$

has a flow which just rotates the (q_1, q_2) and (p_1, p_2) planes (this is the angular momentum for a point particle in the plane and we know its action is just rotation). Thus after an appropriate amount of time we will have exchanged q_1 with q_2 and p_1 with p_2. Courant then states that the same theorem is true for the rectangular parallelepipeds tangent to and surrounding the ellipsoids. Unfortunately, this theorem is restricted to the case where the final set is lined up the same way as the initial set is, which is likely to be a rather rare occurence. We would like to generalize it to the projections under arbitrary transformations.

(Note added in proof: In a very recent preprint, M. Gromov has shown that the unit ball in $\Re^{2N}$ cannot be mapped into a cylinder over a disc in (x, p_x) space with radius less than one by any canonical transformation. This landmark work resolves many unsolved classical problems in symplectic geometry using very sophisticated arguments combining minimal surface theory from Riemannian geometry, elliptic P.D.E. theory, complex analysis in many variables, and the Atiyah-Singer index theorem. It is available as a preprint dated January 1985: "Pseudo-holomorphic Curves in Symplectic Manifolds" from *Institut des Hautes Etudes Scientifiques*, 35, route de Chartres, 91440–Bures-sur-Yvette, France.)

11.2. Relation to the Uncertainty Principle

Semi-classical mechanics allows us to construct certain asymptotic relations between eikonal waves, say evolving under the Schrödinger equation, and Hamiltonian dynamics on a corresponding classical phase space. Every wave function has a corresponding Wigner distribution on the classical phase space. We may use this to asymptotically assign to wave functions, those regions in phase space which contain most of the density of their Wigner function. It is a folk theorem in physics that in N dimensional systems, these regions are of volume h^N (where h is Plank's constant). A classical theorem of Weyl shows that as the energy E approaches infinity, the number of eigenstates of a Hamiltonian corresponding to a bounded classical system, with energy less than E, asymptotically approaches the volume of the region in classical phase space with energy less than E, measured in units of h^N. The exact relation between the classical regions and the quantum states is being clarified with the techniques of micro-local analysis [Fefferman, 1983]. Even though the volume of a wave-packet is always the same in phase space, we are usually interested in its extent in position space or in its Fourier transform's extent in momentum space. As the wave packet evolves, it typically stretches out in a "diagonal" direction in phase space, making both of these projections grow. This is the well known quantum spreading of wave packets (or any other waves with dispersion). There is a rigorous lower bound on how tightly we can compact a wave packet in both a spatial direction and the corresponding momentum direction, given by the uncertainty principle. This says that the product of the q dispersion and the p dispersion must be greater than h for each coordinate. In particular, we may not arbitrarily

give up phase volume in one set of coordinates at the behest of another set. Thus to the extent that the semiclassical connection is valid, we would expect there to be limitations on shrinking projected volumes under Hamiltonian flows. If we could find such a shrinking flow, we could apply the corresponding quantum Hamiltonian to a wave packet and make measurements that give more information than allowed by the uncertainty principle about position and momentum at an earlier time.

11.3. Weinstein's Approach

The basic idea of Weinstein's approach [unpublished] is to reduce the question of reducing projected area to a known theorem about Lagrangian tori. He finds a Lagrangian torus in the region of interest which projects to a circle. If the area enclosed by this circle shrinks under a canonical transformation, w may cause the image of the circle to be disjoint from the circle itself. This implies that the Lagrangian torus is also disjoint from its image. This is disallowed by a known theorem for sufficiently small canonical transformations. Therefore the projected area cannot shrink for small enough canonical transformations. Let us now go through this argument more precisely.

Consider the symplectic manifold S, formed as the product of two symplectic manifolds S_1, ω_1 and S_2, ω_2. Take S_1 to be two-dimensional and S_2 to be arbitrary. For the product symplectic structure we use $\omega = \omega_1 + \omega_2$ (here we are identifying ω_1 and ω_2 with their pullbacks along the natural projections). Consider connected, simply-connected regions $R_1 \subset S_1$ and $R_2 \subset S_2$. Their product

$$R \equiv R_1 \times R_2, \tag{11.2}$$

will be our initial region. Its projection down to S_1 has area equal to the area of R_1. Weinstein has shown that under arbitrary (i.e. nonlinear) canonical transformations, that are sufficiently close to the identity, the projection of the transformed R to S_1 has an area that is greater than or equal to the initial area. Furthermore, there appears to be a "rigidity" theorem which says that if the area is the same, then, in fact, our transformation is a product of a symplectic transformation on S_1

and one on S_2. Any coupling between the two sets of degrees of freedom must cause the projected area to increase. This is reminiscent of the increase of coarse-grained or projected entropy in statistical mechanics.

The proof of the first part rests on some Lagrangian intersection theory. Let us sketch the basic idea here. The boundary of R_1 is topologically a circle. Since S_1 is 2 dimensional, we may apply a canonical transformation that takes any such region to any other one with the same area as was shown in [Banyaga, 1977]. The circle is a Lagrangian submanifold of S_1 (i.e. ω_1 vanishes on it) as are all 1 dimensional submanifolds. Choose a Lagrangian torus in R_2 (say by taking a product of sufficiently small circles in each of the canonical planes formed by a canonical basis in some little region). The product of these two tori will be a Lagrangian torus in R with respect to ω. This torus projects down to the circular boundary of R_1 in S_1. Assume we could apply a near identity canonical transformation to R such that its projection to S_1 had a smaller area than R_1. By a canonical transformation of S_1 we can force this image to be strictly inside the boundary of R_1 (say by making it look like a smaller circle concentric to the boundary of R_1 in some coordinates). But this means we have a near-identity canonical distortion of a Lagrangian torus whose projection doesn't intersect the initial torus. If the projections don't intersect, then neither do the tori. But this is known to be impossible.

In fact, recent technical advances have shown that even C^0 small canonical deformations *of* Lagrangian tori must intersect the initial torus [Conley and Zehnder, 1983 and [Chaperon, 1983]. This result is one of a number of related results about the existence of periodic orbits (and fixed points and intersection points) arising

from a combination of topological and symplectic properties. These ideas had their beginning in Poincaré's conjecture that an area preserving map of an annulus to itself that twisted the bounding circles in opposite directions must have at least two fixed points [Poincaré, 1912]. Poincaré used his theorem to show the existence of infinitely many periodic orbits in the neighborhood of an elliptic periodic orbit in celestial mechanics. Poincaré's theorem was proved in [Birkhoff, 1913] and partially generalized to certain compact symplectic manifolds in [Arnold, 1965]. Since then many fascinating developments have occurred as surveyed in [Weinstein, 1984a].

Let us sketch the proof for C^1 small deformations (as in [Arnold, 1978] p.420). A small neighborhood of any Lagrangian submanifold is symplectomorphic to a neighborhood of the cotangent bundle of that Lagrangian submanifold with the canonical cotangent symplectic structure [Weinstein, 1971]. In this representation, the initial Lagrangian torus is the zero section. A sufficiently close Lagrangian torus will project diffeomorphically onto this under the canonical cotangent projection. In fact it is actually the graph of the differential of a smooth function on the initial torus. Because it is Lagrangian it must be the graph of a closed form. (It is easy to check that the pullback of the canonical one-form on a cotangent bundle from the graph of a one-form α back to the zero section is exactly that one-form α. Being Lagrangian means that the differential of the canonical one-form vanishes on the manifold, and so its pullback's differential must also vanish, which just says that α is closed.) Because the canonical transformation extends to an open ball containing the torus, this form must actually be exact.

Poincaré's lemma tells us that locally every closed one-form is the differential

of a function. The only problem might be that when we go around the fundamental uncontractable loops of the torus, the function might not return to the initial value. If we think of the initial torus and its deformation in the original phase space, the canonical transformation is defined over a topological ball containing them. We may thus find 2-dimensional discs whose boundaries are the fundamental loops on the torus. The integral of the symplectic form over the initial disc is therefore equal to its integral over the deformed disc. If we make a closed cylinder (no boundary) from the two discs and the sheet swept out by a loop under the deformation, Stokes' theorem says that the integral of ω over the cylinder is zero. Since the two disc contributions cancel, the integral of ω over the cylinder swept out by the loop is itself zero. Stokes' theorem then tells us that the integral of a one form whose differential is ω in the region the sheets sweep out, must have the same integral over the loop and its deformation. In general, canonical deformations with this property are known as exact deformations. The canonical one-form in our cotangent representation has ω as its differential in this region. But since the intial torus is the zero section, the integral of $p\ dq$ around any loop must be zero. We may then conclude that its integral around any loop of the deformed torus is also zero. There is therefore no obstruction to finding a function on the deformed torus whose differential is the canonical one-form.

The points of intersection of the deformed torus with the initial torus are exactly the places where this differential vanishes (and so its graph hits the zero section). But these are the critical points of the function on the torus. But being compact, the torus forces any function to have a maximum and a minimum (and

at least one possibly degenerate saddle by Lusternik–Shnirelman category theory) which forces intersection. Chaperon gives the criterion that any exact deformation of the torus on which the $q_i^2 + p_i^2$ all remain positive (and so there is some loop that isn't pulled through zero) must cause the image to intersect the initial torus.

11.4. Theorem for Linear Canonical Transformations

We would like to generalize this result from near-identity transformations to arbitrary transformations. Unfortunately, it cannot be true in general as the following example shows. Let us consider $\Re^2 \times \Re^2$ and take our region to be a large area in S_1 times a small area in S_2. There is a canonical transformation which exchanges S_1 and S_2 and so makes the projection onto S_1 smaller. This exchange is a "large transformation" and by the above, the projection must first increase and then decrease. We can get around this example in a number of ways. We may consider regions whose projected areas are the same or we may study the sum of the areas or the minimal area under the different projections. Below we prove a theorem for arbitrary linear transformations on the increase of the sum of the projected areas. This generalizes Courant's theorem to arbitrary linear canonical transformations (but so far only in $\Re^4$) and is not restricted to be near-identity. One would like to extend it to higher dimensions, non-parallelepiped initial conditions, and nonlinear transformations. We develop the theory of projected parallelepipeds for arbitrary dimensions. The nonlinear situations are locally linear and we may chop our region into parallelepipeds to which the theorem applies. Unfortunately, the little parallelepipeds obscure one another under projection and this must be understood for a nonlinear theory.

Another tack to take is to not look at the projected area, but rather the projected measure (i.e. we want the volume in the region that sits over each little area in the two-dimensional space). Physically, it is often important to know not just that *some* particle's state projects to a given region, but also *how many* particles.

Perhaps the entropy (i.e. integral of $(-p \log p)$ over the projected region) of this distribution function must increase. In this situation as well the geometry of parallelepipeds appears relevant. One place to see the possible connection is through the central limit theorem. If we project a cube to a real line parallel to its longest diagonal, then the projected measure along the line is the multiple convolution of a rectangular pulse, with one pulse for each dimension of the cube. In the limit of large dimension, this approaches a Gaussian. The measure of lines on which the projection is close to a Gaussian asymptotes to one as the dimension grows. Futhermore, the entropy of a Gaussian is a maximum for all distributions with the same dispersion. A cube projected to its diagonal is in some sense maximally sensing each of the degrees of freedom (perturbing a point along any of the orthogonal axes of the cube is reflected by a perturbation in the image of the projection; if the projection isn't diagonal, then some axes are short-changed because their projections are less important than others). Any rotation which lines the projection up more along an edge will decrease the entropy of the projected measure.

If such a theorem guaranteeing the increase of entropy under coupling is true in infinite dimensions, then it might shed light on the increase of entropy under the evolution of Boltzmann's equation. Bogoliubov derived Boltzmann's equation as the first order term in an asymptotic expansion of the BBGKY hierarchy. The essential part of the argument obtains the evolution of the two particle distribution function as the Liouville equation for two interacting particles (the contribution of the three-particle and higher distributions is higher order in the ordering scheme). The evolution of the two particle distribution function is then Hamiltonian and

linear (albeit infinite dimensional). In Bogoliubov's argument, we are given a spatially uniform one-particle distribution whose evolution we wish to find (its evolution equation depends on the two particle distribution). We take the two particle distribution which is simply an uncorrelated product of this one-particle one with itself as initial conditions for the 2-particle Liouville equation. Under evolution the two particles will collide creating a correlation and the infinite time asymptotics of this (i.e. when the two particles become widely separated) is used as the 2-particle driving term in the 1-particle evolution, which becomes the Boltzmann collision term. Boltzmann shows that under this evolution, the entropy of the one-particle distribution must increase.

11.4.1. The Geometry of Projected Parallelepipeds

When we project a parallelepiped down to a lower dimensional space, the result is no longer a parallelepiped as in figure (11.1). We show here, though, that it may be decomposed into parallelepipeds in a uniform way, allowing us to get a formula for the volume of the projection. We demonstrate this decomposition using induction and a couple of simple initial lemmas. Let us be given n vectors V_i, $1 \leq i \leq n$ in the k-dimensional linear space $\Re^k$. We are interested in the region R consisting of points of the form

$$\sum_{i=1}^{n} a_i V_i \quad \text{for} \quad 0 \leq a_i \leq 1. \tag{11.3}$$

Let us call this region the $[0,1]$ span of the V_i's.

Lemma 11.2. *For any n and k, R is convex.*

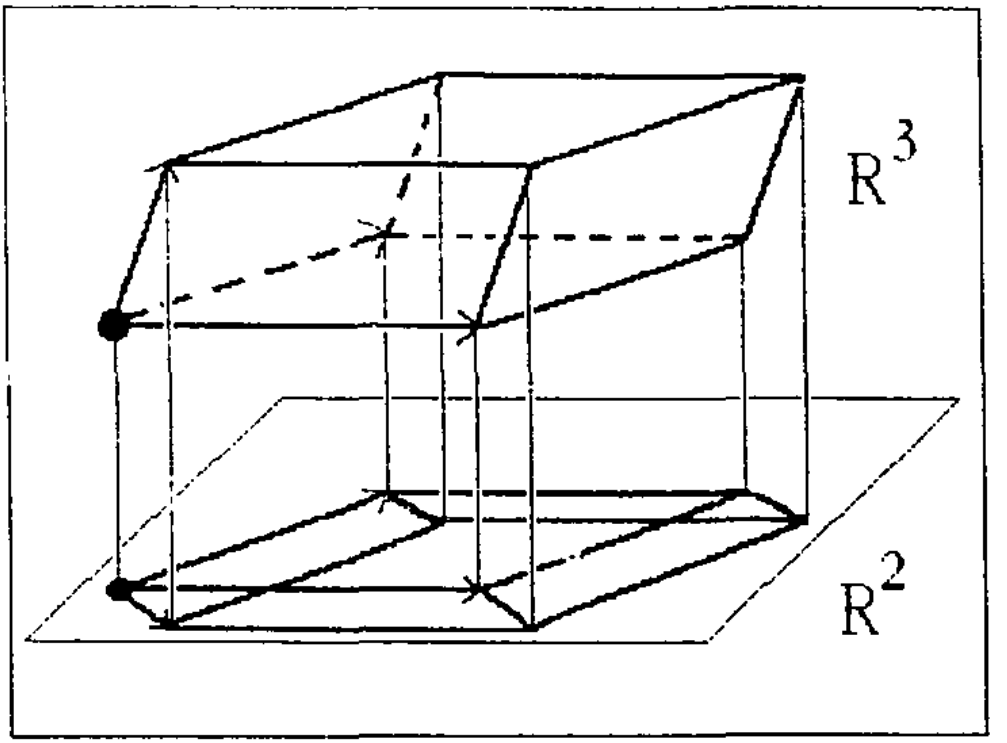

Figure 11.1: A three-dimensional parallelepiped projects to a hexagon in two dimensions.

Proof. Consider any two points $\sum_{i=1}^{n} a_i V_i$ and $\sum_{i=1}^{n} b_i V_i$ in the region R. The line between them consists of all points of the form

$$\sum_{i=1}^{n} a_i V_i + t\left(\sum_{i=1}^{n} b_i V_i - \sum_{i=1}^{n} a_i V_i\right) = \sum_{i=1}^{n} [(1-t)a_i + tb_i]V_i \tag{11.4}$$

where $0 \leq t \leq 1$. But since t and $1-t$ are non-negative and a_i and b_i are less than one, we see that

$$(1-t)a_i + tb_i \leq (1-t)\cdot 1 + t \cdot 1 = 1. \tag{11.5}$$

But this shows that all points on the line between the two given points satisfy the defining criterion for R and therefore belong to R. Hence R is convex. Q.E.D.

Definition 11.1. *Let us define a vertex of a subset of $\Re^k$ to be any point of the set such that there does not exist any open interval of a straight line in $\Re^k$ which contains the point and lies completely in the set.*

Lemma 11.3. *All vertices of the set R have a unique representation as a sum $\sum_{i=1}^{n} a_i V_i$ and all of the a_i's are either 1 or 0.*

Proof. Assume we can represent a vertex as $\sum_{i=1}^{n} a_i V_i$ where some a_j is not 1 or 0. There is then an open interval around a_j which is contained in [0,1], and therefore the corresponding vectors form an interval of a straight line lying in R, violating our assumption of verticity. Now assume that there are two representations: $\sum_{i=1}^{n} a_i V_i$ and $\sum_{i=1}^{n} b_i V_i$ for the vertex, where all the a_i's and b_i's are neccessarily 0 or 1. The point may then also be represented as

$$(1-t)\sum_{i=1}^{n} a_i V_i + t\sum_{i=1}^{n} b_i V_i, \tag{11.6}$$

where $0 \leq t \leq 1$ or equivalently as

$$\sum_{i=1}^{n} |(1-t)a_i + tb_i| V_i. \tag{11.7}$$

But unless all the b_i are equal to the corresponding a_i, we can thereby get a coefficient which is not 1 or 0, violating the above. Thus vertices have unique representations. Q.E.D.

Lemma 11.4. *a) Let us assume that the number of vectors n is greater than or equal to the dimension k of the ambient space. The region R may then be decomposed as a union of parallelepipeds formed by rigid translations of the [0,1] spans of k-element subsets of the set of n vectors, such that they intersect only in*

$(k-1)$ and lower dimensional sets (which are of Lesbegue measure zero). There are $n!/k!(n-k)!$ such parallelepipeds.

b) Let us assume that R is k-dimensional. The boundary of R may then be decomposed as a union of parallelepipeds formed by rigid translations of the $[0,1]$ spans of $(k-1)$-element subsets of the set of n vectors, where each such span is included twice and the intersections of the translates are of dimension $k-2$ and lower.

Proof. We prove these two parts together using a double induction in k and n. The essence of the proof is shown in figure (11.2). We have shown the effect of adding an extra vector to the projection of a 3-dimensional parallelepiped's projection onto 2-dimensions (to give the projection of a hyper-cube). It adds an extra parallelogram to half of the boundary edges formed from the extra vector and 1 each of the original 3 vectors. We may generalize this same construction to arbitrary dimensions as follows.

Both a) and b) have assumptions that prevent n from being less than k. To start the induction, assume $n = k$, where k is arbitrary. a) is now trivial since there are exactly k vectors and R is indeed equal to their $[0,1]$ span. In case b), R is actually a parallelepiped and its faces are obtained by taking the $[0,1]$ span of each set of $k-1$ vectors and basing it at zero and at the end of the remaining vector, which agrees with the statement in the lemma. We also need the case $k = 1$ for arbitrary n (see figure (11.3)). This case is also easy, since we may just line up the n vectors starting at zero, each representing its own $[0,1]$ span and the boundary is just two endpoints which we may take as the union of the ends of the vectors.

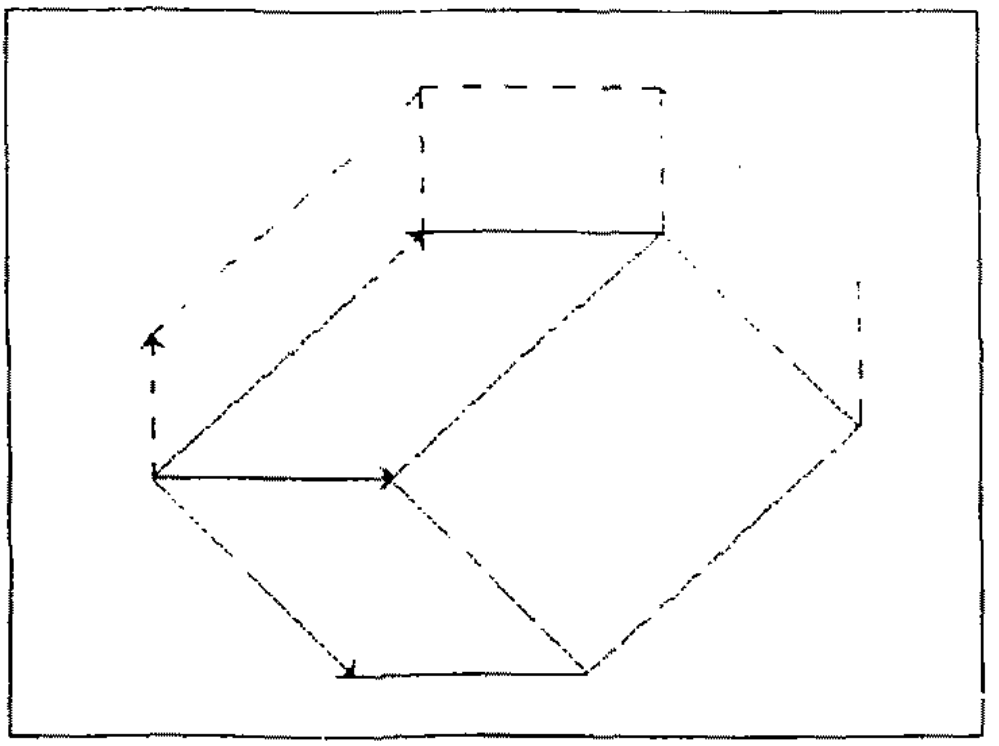

Figure 11.2: The essence of the lemma. Adding a fourth vector adds three parallelograms formed from that vector and each of the original three vectors.

Let us now assume the theorem true for all n's when k is less than k_0 and for n's less than or equal to n_0 when k equals k_0 and prove it for k_0, $n_0 + 1$. By induction this will prove the lemma as stated for all allowed k's and n's.

Let us begin with part a). We may assume that $n_0 + 1$ is strictly greater than k (since we already did the equal case). Thus R is the $[0,1]$ span of $n_0 + 1$ vectors. Consider the $[0,1]$ span of the first n of these and call it S. R is obtained from S by adding $[0,1]$ multiples of the $(n_0 + 1)$st vector to each point in S. In fact, we get all the extra points in R that aren't in S by adding such multiples to only boundary points of S (since to get a new point the vector must intersect the boundary and we can get the same point by adding a smaller multiple to that point

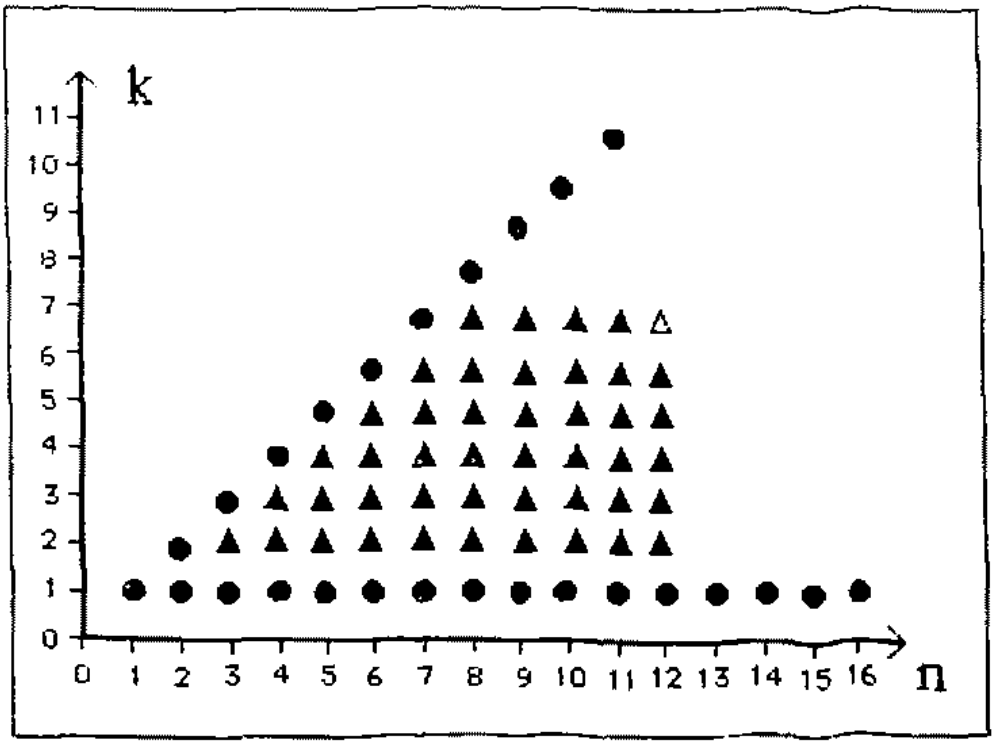

Figure 11.3: The n, k plane.The circles are the base induction cases. The open triangle is an example case. The closed triangles are the cases we must use in its proof by induction in addition to the circles.

of intersection). By induction, S decomposes into translates of the $[0,1]$ spans of the k-element subsets of the first n_0 vectors. We need only show that the region in R : ot contained in S is the union of translates of $[0,1]$ spans of the (n_0+1)st vector with all $k-1$ element subsets of the first n_0 vectors (up to sets of smaller dimension) to prove part a). But we may now apply part b) to S by the inductive hypothesis. Consider each set of $k-1$ vectors from the first n_0 in turn. By part b) the boundary of S has two regions that are translates of their span, which are parallel (since they are rigid translates of each other). By the convexity of S (from the lemma above), S lies entirely in and on one side of the hyperplane in $\Re^k$ which

contains such a piece of the boundary. If the two regions are distinct, S must lie on opposite sides of their repective hyperplanes. The vector V_{n_0+1} either lies in such a hyperplane or points to one side of it when its basepoint is contained in it. If it lies in the hyperplane or if the $k-1$ vectors span a space of lower than $k-1$ dimensions, then the parallelipiped it forms with the $k-1$ vectors is degenerate and so of lower dimension than k and irrelevant for our union. Otherwise, take the piece of S's boundary such that S is on the opposite side of the hyperplane as the vector. This parallelepiped and the vector $[0,1]$ span a k-parallelepiped which is in R but whose intersection with S is of lower dimension (since they can only intersect in the hyperplane because they are on opposite sides of it). Because S's boundary is filled out by such $k-1$-parallelepipeds, and R is obtained by adding $[0,1]$ multiples of the (n_0+1)st vector to the boundary, R is the union of these k-parallelepipeds as desired. It remains only to show that they intersect each other in sets of dimension lower than k. Since these parallelepipeds are obtained by sweeping the $(k-1)$-parallelepipeds in half the boundary of S along the last vector, if two of them intersect in a region of dimension k, the corresponding $(k-1)$-parallelepipeds must intersect in a region of dimension $k-1$, but this is not allowed by the statement of b).

Let us now prove b) under the same inductive assumption as above and using the result of a) (this is ok since the proof of a) only used b) on lower inductive cases). Since we have added k-parallelepipeds to S to get R, half of the original $k-1$ parallelepipeds in S's boundary have remained and the other half have just been rigidly translated by the last vector. Thus we have two parallel copies of each

of the $(k-1)$-parallelepipids generated by the first n_0 vectors, in the boundary of R as desired. Extra boundary sufaces arose when we translated one parallelepiped by the last vector but not its neighbor. Consider the projection from $\Re^k$ to $\Re^{k-1}$ along the last vector. Both S and R have the same image and the extra part of R's boundary is exactly formed by translating by the last vector that part of S which projects to the boundary of the image. We may apply part b) to the projection of S to see that this boundary may be decomposed into pairs of translates of all $[0,1]$ spans of the projection of $k-2$ vectors (since the induction hypothesis applies to this lower dimensional space). The subset of S that projects onto this boundary is then decomposed into unions of $[0,1]$ spans of $k-2$ vectors up to things which point along the last vector and are therefore degenerate with it. So indeed the extra part of the boundary of R is decomposable into translates of parallelepipeds formed from the last vector and all $k-2$ element sets of vectors from the first n_0. Together with the original decomposition of S's boundary these give us the desired decomposition of part b). Q.E.D.

11.4.2. The Case of Linear Canonical Transformations

The way we wish to use this lemma is as follows. We get subsets that are $[0,1]$ spans of vectors as the projected image of a parallelepiped in a lower dimensional space. The volume of this projection is the sum of the volumes of all parallelepipeds formable by these vectors, by the lemma. For $2N$-dimensional symplectic vector spaces, the volume of a $2N$-dimensional parallelepiped is given by inserting the generating vectors into the Nth wedge product of the symplectic form with itself.

giving a natural volume element. Let us now use this to prove the theorem of interest.

Let $S_1 \approx \mathbb{R}^2$ have the coordinates (q_1, p_1) and the symplectic form: $\omega_1 = dq \wedge dp$. Let S_2, ω_2 be defined similarly. Consider the symplectic manifold

$$S \equiv S_1 \times S_2 \tag{11.8}$$

with the symplectic structure

$$\omega \equiv \omega_1 + \omega_2. \tag{11.9}$$

Here the ω_i are thought of as pulled back to S. We are interested in how the area of the projection of a set down to S_1 changes under canonical transformations. We may prove the following theorem for product parallelepipeds under linear symplectic transformations. This is a physically relevant setting for many physical situations and is the local picture for the general case.

Theorem 11.5. *Let us be given two vectors u_1 and v_1 in S_1 whose $[0,1]$ span is a parallelogram of area A_1 and similarly for u_2 and v_2 in S_2 with area A_2. Consider the parallelepiped P in S which is the product of these two parallelograms. Under an arbitrary linear canonical transformation L of S, the sum of the projected areas of $L \cdot P$ in S_1 and S_2 is always greater than or equal to $A_1 + A_2$. Furthermore equality only holds if a certain restrictive condition listed below holds.*

Proof. We have seen above that the projected area of $L \cdot P$ in S_1 is equal to the sum of the areas of the parallelograms spanned by the projections of $(L \cdot u_1, L \cdot u_2)$, $(L \cdot u_1, L \cdot v_1)$, $(L \cdot u_1, L \cdot v_2)$, $(L \cdot u_2, L \cdot v_1)$, $(L \cdot u_2, L \cdot v_2)$, and $(L \cdot v_1, L \cdot v_2)$. The area is just the absolute value of the result of inserting the projected vectors into

ω_1. But this is the same as inserting the vectors into the pullback of ω_1 along the projection, which we are denoting again by ω_1. Thus the sum of the projected areas in S_1 and S_2 is given by

$$\begin{aligned}
&|\omega_1(L\cdot u_1, L\cdot u_2)| + |\omega_1(L\cdot u_1, L\cdot v_1)| + |\omega_1(L\cdot u_1, L\cdot v_2)| + \\
&|\omega_1(L\cdot u_2, L\cdot v_1)| + |\omega_1(L\cdot u_2, L\cdot v_2)| + |\omega_1(L\cdot v_1, L\cdot v_2)| + \\
&|\omega_2(L\cdot u_1, L\cdot u_2)| + |\omega_2(L\cdot u_1, L\cdot v_1)| + |\omega_2(L\cdot u_1, L\cdot v_2)| + \\
&|\omega_2(L\cdot u_2, L\cdot v_1)| + |\omega_2(L\cdot u_2, L\cdot v_2)| + |\omega_2(L\cdot v_1, L\cdot v_2)|.
\end{aligned} \tag{11.10}$$

Applying this formula in the original situation, where L is the identity, and using the fact that ω_1 annihilates u_2 and v_2, we see that

$$A_1 + A_2 = |\omega_1(u_1, v_1)| + |\omega_2(u_2, v_2)|. \tag{11.11}$$

Because L is symplectic on S, ω is preserved by it. This means that for any two vectors w_1 and w_2 in S, we have

$$\begin{aligned}
\omega(L\cdot w_1, L\cdot w_2) &= \omega(w_1, w_2) \\
&= \omega_1(L\cdot w_1, L\cdot w_2) + \omega_2(L\cdot w_1, L\cdot w_2) \\
&= \omega_1(w_1, w_2) + \omega_2(w_1, w_2).
\end{aligned} \tag{11.12}$$

Let's assume that u_1, v_1 and u_2, v_2 are in the right order so that ω_1 and ω_2 give positive answers when acting on them. Then we see that

$$\begin{aligned}
A_1 + A_2 &= \omega_1(u_1, v_1) + \omega_2(u_1, v_1) + \omega_1(u_2, v_2) + \omega_2(u_2, v_2) \\
&= \omega_1(L\cdot u_1, L\cdot v_1) + \omega_2(L\cdot u_1, L\cdot v_1) + \omega_1(L\cdot u_2, L\cdot v_2) + \omega_2(L\cdot u_2, L\cdot v_2) \\
&\le |\omega_1(L\cdot u_1, L\cdot v_1)| + |\omega_2(L\cdot u_1, L\cdot v_1)| + |\omega_1(L\cdot u_2, L\cdot v_2)| + |\omega_2(L\cdot u_2, L\cdot v_2)|.
\end{aligned} \tag{11.13}$$

But in the expression above for the new sum of projected areas, this last expression appears with some positive or zero terms added to it. Thus the sum of the new

areas is greater than or equal to the original sum. We can get a necessary condition for the sum to be equal, since all the extra terms in absolute value must separately vanish. So $\omega_1(L \cdot u_1, L \cdot u_2)$, $\omega_1(L \cdot u_1, L \cdot v_2)$, $\omega_1(L \cdot u_2, L \cdot v_1)$, $\omega_1(L \cdot v_1, L \cdot v_2)$, $\omega_2(L \cdot u_1, L \cdot u_2)$, $\omega_2(L \cdot u_1, L \cdot v_2)$, $\omega_2(L \cdot u_2, L \cdot v_1)$, and $\omega_2(L \cdot v_1, L \cdot v_2)$ must all vanish. Q.E.D.

Courant's paper contains an apparent counterexample to this kind of theorem for nonlinear dynamics. A technique used in certain accelerators is to insert a knife-edged septum into a recirculating beam to strip off a thin outer band and cause the inner region to continue to recirculate. In this way one can make the spatial width of the beam smaller, without increasing its momentum spread. The knife edge is really a very steep potential. Even though the number of particles that hit it is small, they are given a very large momentum kick (this sounds like some of the classical arguments for the uncertainty principle). Thus the actual region in phase space is probably very spread out. The physically important quantity in this example, however, is the particle distribution function. If designed properly, one could presumably make the measure of the wild particles small. This kind of example indicates that one must be very careful in any distribution function version of these theorems.

Another construction that such a theorem will have to beat is as follows. It is well known that there exists a canonical transformation that moves any N distinct points of phase space to any other N distinct points. If we force most of the measure to be near these points, then we can line them up to project to any kind of distribution we desire. How do we get the measure to be near points in a product

distribution? Consider the phase space to be $\Re^6$. All canonical transformations will take place on only the $\Re^4$ formed by the first 4 coordinates. The last two coordinates are needed only to make the measure large near the specified points in the first four dimensions. We choose points in the first and second $\Re^2$ so that the points in the product $\Re^4$ don't all lie over a single point in the first $\Re^2$ (for example we can choose 9 points arranged in three rows of three). We choose a volume in $\Re^6$ which is very extended in the last two dimensions near these points. This may be constructed as a product of a region in the first 2 dimensions times a region in the last 4. We now use a canonical transformation in the first 4 coordinates which makes the points line up along a single fiber of the projection to the first two coordinates, letting the last two dimensions just go along for the ride. The projected distribution can then be made to vary in any way we wish.

An even simpler example can be constructed in $\Re^4$ which we assume to be coordinatized by (x,p_x,y,p_y). Let our initial set be the unit ball in $\Re^4$. We chop the ball into 3 pieces: the part B_1 with $-1 \leq x \leq -\epsilon$, the part B_2 with $-\epsilon \leq x \leq \epsilon$ and the part B_3 with $\epsilon \leq x \leq 1$. By making ϵ small enough, we can make the measure of B_2 arbitrarily small. By a simple canonical translation and rotation defined on a region including B_3 but in the complement of B_1, we can make B_3 "sit above" B_1 so that they project to the same region in the (x,p_x) plane. This is easily extended to a canonical transformation on all of $\Re^4$ which leaves B_1 fixed (say by extending the time dependent Hamiltonian vector field defining the transformation using a partition of unity). By Gromov's theorem, we know that B_2 must be stretched in some wild way so that its projected area is large (or at

least its image surrounds a large area). But we have constructed B_2 so that its measure in $\Re^4$ is small. By repeating this construction, we can make an arbitrarily large percentage of the measure of the ball project to arbitrarily small regions by a canonical transformation.

Chapter 12: Reversibility vs. Irreversibility

The seeming paradox of reversible underlying dynamics leading to irreversible macroscopic behavior has been wrestled with since the time of Boltzmann. For the case of gas dynamics, Boltzmann derived his famous equation on the basis of the questionable statistical assumption of "Stosszahlansatz". From the Boltzmann equation one may derive the "H theorem" which says that the time derivative of the entropy is greater than or equal to zero. The "Stosszahlansatz" or "molecular chaos assumption" says that the probability distributions for colliding molecules should be uncorrelated. As has been pointed out many times (see for example: pp. 46-88 of [Chapman and Cowling, 1958] and pp. 28-32 of the *Statistical Physics* volume in [Landau and Lifshitz, 1960-1981]), because the underlying dynamics is reversible, for every state with its entropy increasing, there is a corresponding state with its entropy decreasing. In fact, one may show from this argument that the stosszahlansatz can hold only when the time derivative of entropy is zero. There is no intrinsically special direction in time: If one starts with a random state with low entropy, the entropy increases if one follows the evolution of the state either backwards or forwards in time. There are many more states with high entropy than

low, and so a system is likely to be in a high entropy state regardless of the time.

There is a simple dynamical system which exhibits some of the seemingly paradoxical features of these complex statistical systems. Consider the differential equations

$$\begin{aligned} \dot{x} &= -xy \\ \dot{y} &= x^2 - y^2. \end{aligned} \tag{12.1}$$

A sketch of the dynamics in the x-y plane is shown in figure (12.1).

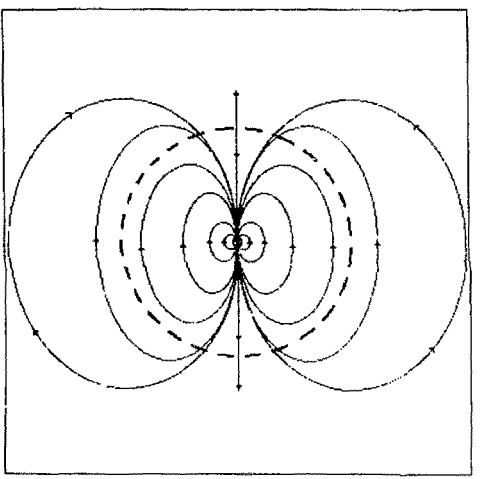

Figure 12.1: The phase portrait of the equations in the x-y plane. The origin acts like an attractor both forward and backward in time.

Consider the evolution of the radius: $\sqrt{x^2 + y^2}$. For every point (x,y) where the radius is getting smaller, there is a corresponding point (x,-y) where it is getting larger. That is, if you look on a circle of radius r, the length of arc where the vectorfield points outward is the same as where it points inward. Nonetheless almost every initial condition (excluding the y axis, which you have zero probability of hitting) leads to the eventual decrease of the radius asymptotically to zero. Notice that this decrease takes place not only forward in time but backwards. The origin is almost an attractor for both the true and the time-reversed dynamics.

Chapter 13: Hamiltonian Dissipation in Infinite Dimensions

"We now have an example of a so-called irreversible thermodynamic process, such as does not occur in mechanics: in all natural processes the entropy increases. ... Irreversible processes can in no way be simulated by purely mechanical processes."— [Klein, 1928] p.203.

In this section we will examine some interesting facets of the dynamics of infinite dimensional Hamiltonian systems. One of the key simplifying aspects of dissipative dynamics is the presence of attractors whose behavior dominates the time asymptotics of all nearby initial conditions. In finite dimensions, Liouville's theorem regarding the volume preservation of Hamiltonian dynamics leads to the Poincaré reccurence theorem. This says that under the time-one map of a bounded Hamiltonian system every neighborbood of every initial condition has points that return to that neighborhood and furthermore almost every point comes back infinitely often arbitrarily close to its starting point. The proof rests on the pigeon hole principle which says that with N pigeon holes and $N + 1$ pigeons, there must

be at least one pigeon hole with more than one pigeon. In a finite volume region, there is only room enough for a finite number of images of a non-zero volume region under a volume preserving mapping. Eventually two images must intersect and by mapping the intersection region back to the starting region, we find points that recur. This precludes the presence of attractors in finite dimensional Hamiltonian systems (no small region can be special since almost all points go back to where they came from). We will see that infinite dimensional Hamiltonian systems can have attractors of a certain kind.

13.1. Poincaré Recurrence and Attractors

The fact of recurrence is behind Zermelo's objection to Boltzmann's H theorem proving the increase of entropy for the Boltzmann equation. The fact of the matter is that the recurence time is quite long. For merely 10 harmonic oscillators with frequencies about 1 cycle per second and irrationally related, the room in the corresponding 10-torus is so vast that the typical time to return to within one percent of their initial conditions is 10^{20} seconds which is longer than the age of the universe. Boltzmann's retort to Zermelo was purportedly: "You should live so long", [Kac, 1959]p. 62 (in response to Loschmidt who objected to getting an irreversible equation from reversible underlying dynamics he replied: "Go ahead, reverse them!"). These long times also indicate that true ergodicity is not responsible for the experimental validity of statistical mechanics (if we consider a state of a cubic meter of air to be the number of molecules in each cubic millimeter and if

the dynamics is ergodic, the gas visits every state only once in about 10^{100} times the age of the universe).

13.1.1. The Lack of Recurrence in Infinite Dimensions

Reccurrence is a characteristically finite dimensional phenomenon. As the dimension gets higher the recurrence time-scale grows exponentially. If we are interested in finite (but perhaps long) times, it is often physically valid to introduce asymptotics even if it leads to infinite dimensional models. Sometimes the analysis of infinite dimensional systems is easier because the structure "at infinity" organizes the dynamics. The wave equation is easier to deal with than the underlying molecular dynamics. When we make such models we must be careful not to believe them when they depend crucially on asymptotic aspects beyond the value of the limiting parameter at which the modelled system really is. (For an analysis of some very interesting phenomena that occur near molecular wavelengths in a system asymptotically approximated by the wave equation, see [Maslov, 1976] p. 58). The time to recur grows with the dimension and infinite dimensional systems need not recur. There is no general Liouville theorem in infinite dimensions and orbits need not recur (they can "head off to infinity" along higher and higher dimensions still staying close to the origin). While we have argued that when the system is in too high a dimension the model becomes unphysical, the behavior on the way there will reflect the real behavior and properties of the infinite dimensional system may be properties of the real system for long times.

13.2. Asymptotic and Liapunov Stability

For finite dimensional dynamical systems there are two standard notions of stability of an invariant subset. The weaker of the two is *Liapunov stability* which says that given any neighborhood of the set we can find a (possibly) smaller neighborhood all of whose points have future time orbits lying inside the given neighborhood. This says that if we perturb a point slightly from our set, it hangs around forever. An example is the equilibrium of a simple harmonic oscillator. The stronger notion is that of *asymptotic stability*. This says that there is some neighborhood V of our set whose forward time images all lie in V and such that each orbit asymptotes into our set. Asymptotically stable sets are called *attractors* and the points which limit on them form the *basin of attraction*. If an attracting set is a fractal (with respect to your favorite definition of dimension), then it is a *strange attractor*.

13.2.1. Almost Attractors

Recent work of Grebogi, Ott and Yorke has shown that strange attractors can exist without chaotic dynamics [Grebogi, Ott, and Yorke, 1984]. For this they needed a definition of attractor which we shall also use. We call a set an "almost attractor" if for small enough neighborhoods, almost every point eventually asymptotes to our set. We have seen in the example of a vector field on the plane along dipole field lines that it is possible for a point to be an almost attractor both forward and backward in time. By Poincaré reccurrence, this is not possible for finite dimensional Hamiltonian systems.

13.3. Reversible Attractors and Infinite Dimensional Hamiltonians

Infinite dimensional Hamiltonian systems, on the other hand, can have almost attractors and if the system is time reversible (as most physical systems are) then it is of the time reversible type. This type of structure is responsible for most of the dissipative models in physics that I know of. The resolution of Loschmidt's paradox here lies in the fact that the infinite dimensional model is valid only asymptotically and becomes a bad representative of the underlying finite dimensional system after a long but finite time. Until that time it represents the system well and its dissipation represents real tendencies of the underlying system (when they ultimately break down due to recurrence, we call it a *fluctuation*).

13.3.1. Reversible Almost Attractor in the Wave Equation

A simple example to think about merely consists of the wave equation on a one-dimensional string. It is well known that this system is Hamiltonian. Let us define the state space of the system, to be those displacement and velocity fields which die off exponentially at infinity. This class is preserved by the time evolution for finite time (which just translates a wave). Let us assume we are more interested in what the wave is doing under our noses near the origin than what it is doing far away. We will thus put a norm on our space which says that the size of a wave is the integral of the sum of the absolute values of the displacement and velocity fields over the line weighted by a Gaussian centered at the origin:

$$\int_{-\infty}^{\infty} (|f(x)| + |v(x)|)e^{-x^2}dx. \tag{13.1}$$

Since the dynamics is just translation, eventually every wave in our class gets "pushed into the wings" far enough so that its norm decays monotonically to zero. Note that this happens both forward and backward in time. Thus the quiescent string is a reversible almost attractor for this system. In higher dimensions, compactly supported waves decay even in the uniform norm (i.e. the size of a function is the maximum of its absolute value over all space).

This example has many of the characteristics of the dipole vector field in an infinite dimensional Hamiltonian context. While everything eventually comes into the origin, for any given time there are always states which haven't come close yet.

13.3.2. The Liouville Equation and Koopmanism

A natural situation in which this sort of system arises occurs when we consider the Liouville equation of a dynamical system, which is known to be Hamiltonian (albeit with respect to a Lie Poisson bracket). This looks at the evolution on the space of probability distributions on a manifold given by pushing a distribution forward along a measure preserving map. Koopman suggested studying dynamical properties of the underlying system by the spectral properties of the Liouville operator(see [Cornfeld, Fomin, and Sinai, 1982] p. 323). John Cary and John David Crawford have looked at the Liouville evolution of probability distributions on Arnold's cat map. What one finds is that the map's effect on distributions is merely to shift Fourier components along hyperbolas in Fourier space except for the zero wavevector component which is invariant. If the original distribution is smooth, then its Fourier series dies off exponentially. If we put a norm on the distributions which

weighs the low Fourier components more heavily (say our data is run through a low pass filter, i.e. we perform some kind of averaging or binning), then as our distribution's spectrum goes trotting off to infinity, its norm goes to zero leaving only the constant part and we again have a reversible attractor up to a constant. They show that this leads to the exponential decay of the autocorrelation of smooth functions.

13.3.3. Landau Damping

Similar phenomena are responsible for Landau damping (which damps both forward and backward in time) of waves in plasmas. We start with a distribution of particles (as in figure (13.1)) in the bottom half of the "eye" of the pendulum dynamical vector field (as in starting with particles in a strip in velocities slower than the wave, and evenly spread out in phase relative to the wave). As time goes on the dynamics inside the eye is like a shearing harmonic oscillator, which soon smears the inital distribution uniformly through the eye. The effect of this is that the average energy of the particles has gone up (and so the energy of the wave must go down causing damping). This same shearing will occur backwards in time.

13.3.4. The Boltzmann Equation and the BBGKY Hierarchy

The same kind of phase mixing is responsible for the dissipative aspects of the Boltzmann equation as derived from the Hamiltonian and reversible BBGKY hierarchy (see volume 10, *Physical Kinetics* of [Landau and Lifshitz, 1960- 1981]). We rewrite the Liouville equation for the distribution function of all the particles

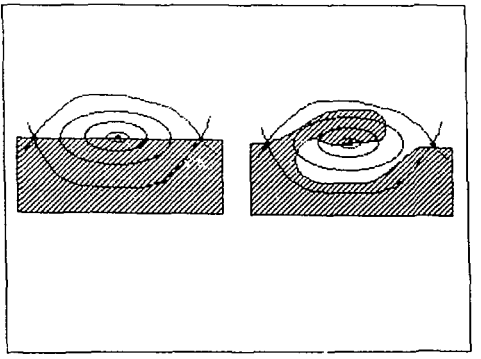

Figure 13.1: The evolution of the distribution function near the wave velocity.

as a hierarchy of equations for 1-particle, 2-particle, ... etc. distribution functions. If we truncate at any stage, we obtain the Liouville equation on the phase space of that many particles. The contribution of the higher order correlations phase mixes away as above leaving only the Boltzmann collision term to first order.

13.3.5. Dissipation from Resonance

The same scenerio applies to the random phase approximation for interacting waves. The essential idea here may be seen by considering resonances. One of the most interesting examples is the simple resonantly driven harmonic oscillator. The solution to this problem is given in most elementary mechanics texts but I have never seen any discussion of its rather paradoxical properties. Let us consider a

particle in one-dimension moving in a unit-frequency harmonic oscillator potential which is driven sinusoidally at the same frequency:

$$\ddot{x} = -x + \cos t. \tag{13.2}$$

For this kind of linear differential equation with an inhomogeneous term, the usual method of solution is to find a particular solution to the inhomogeneous problem and then to add to this an arbitrary solution of the homogeneous one. In this case this procedure leads to the general solution:

$$x(t) = A \sin t + B \cos t + \frac{1}{2} t \sin t.$$

This solution is really quite remarkable. Regardless of the initial conditions (which are specified by A and B), the $t \sin t$ term eventually dominates and the amplitude of oscillations grows indefinitely with time. But the equation is invariant under $t \to -t$! This means that the same behavior occurs as t approaches $-\infty$.

It is interesting to try to visualize the orbits in the three dimensional (x, v, t) space. On the one hand we know that each orbit winds to larger and larger x and v values as t goes toward both positive and negative infinity. This means that an individual orbit winds on what looks like a one-sheeted hyperboloid of revolution along centered on the t axis (or a cone if ever both the position and velocity vanish on its trajectory). On the other hand, the entire system is invariant under the translation $t \to t + 2\pi$. To get some feel for the way this orbits can fit together in this way, consider all orbits that intersect the t axis. These form an invariant 2-dimensional submanifold which includes the t axis. Imagine taking the t, v plane and twisting it about the t axis so that it makes one twist every 2π in t. The dynamics

on the unwound plane is made of parallel straight lines that intersect the t axis at some angle. They leave the t axis linearly for both positive and negative time and yet the whole field is translation invariant in time. When we wind this plane up, the orbits lie on cones. The other orbits do a similar thing on hyperboloids.

Let us think of the harmonic oscillator as being a piece of a larger system and the driving as arising from the rest of the system. Time asymptotically the oscillator only sucks energy from the rest of the system and this is true both forward and backward in time. Let us calculate the time variation of the average energy in the oscillator for long times:

$$\begin{aligned} E &= \frac{1}{2}\dot{x}^2 + \frac{1}{2}x^2 \\ &= \frac{1}{2}(\frac{1}{2}\sin t + \frac{1}{2}t\cos t + A\cos t - B\cos t)^2 \\ &\quad + \frac{1}{2}(A\sin t + B\cos t + \frac{1}{2}t\sin t)^2. \end{aligned} \tag{13.3}$$

For times large compared to A and B (we can make this comparison because we have made everything dimensionless), only terms with an unadorned t in them are important:

$$\begin{aligned} E &= \frac{1}{8}(t^2\cos^2 t + t^2\sin^2 t) \\ &= \frac{t^2}{8}. \end{aligned} \tag{13.4}$$

Thus the oscillator energy grows without bound as t goes to either positive or negative infinity.

In real closed Hamiltonian systems, the energy that the oscillator is sucking out of the rest of the system will eventually cause the damping to change. In nonlinear systems the frequency of driving will be pushed away from resonance. In linear systems the amplitude of driving will go down until eventually the oscillator

drives the rest of the system leading to oscillation on long time scales. The slowly changing amplitude of driving can also be viewed as having energy in frequencies slightly away from resonance. As long as we are in a situation where the effect of the oscillator on the rest of the system is small, however, the solutions will be close to the resonant one. In such cases we get time-reversible dissipative behavior for long times.

This is exactly the sense in which non-equilibrium thermodynamics is dissipative. If we are far from equilibrium then there are many more ways to be perturbed closer to equilibrium than away from it. Statistically the motion appears to go in one direction (eg. entropy increases). The same increase occurs backward in time, however, since the same statistics applies. If we are in a state with low entropy, then it is most likely that a few moments ago we had higher entropy and that a few moments hence we will also have higher entropy. As we get closer to equilibrium, the imbalance in the number of perturbations toward and away from it diminshes. Exactly at the highest entropy state the only way you can go is down. In equilibrium the dissipative properties dissappear and instead we have fluctuations. Since exactly the same mechanisms cause dissipation far from equilibrium and fluctuations in equilibrium, the corresponding rates are related. This is the content of the fluctuation-dissipation theorem. (As an example, a Brownian particle satisfying the Langevin equation:

$$\frac{dv(t)}{dt} = -\nu v(t) + \delta F(t) \tag{13.5}$$

in a heat bath at temperature T which causes the particle's motion to be damped at rate ν and driven with random force $F(t)$, the fluctuation-dissipation theorem

says that

$$\nu = \frac{M}{2T}\int_{-infty}^{\infty} dt\langle \delta F(0)\delta F(t)\rangle. \tag{13.6}$$

)

The mechanism of time-reversible dissipation due to resonance underlies many important physical processes. In many of these situations, the driver changes frequency so as to slowly pass through resonance. An example might be an eikonal light wave passing through a medium whose electrons are bound like harmonic oscillators. The light wave loses energy to the resonantly excited oscillators (note that here it might be that the wave is of constant frequency but the oscillator frequency changes slowly as we progress through the medium; the physics of this situation is identical). For frequencies of driving which are not exactly resonant, the oscillator begins by removing energy, but eventually gives it back (on a time scale of the reciprocal of the frequency deviation from resonance) leading to an oscillatory overall behavior which does not change the average energy of the driver. If we watch the system for time T, then all frequencies within order $1/T$ of the resonant frequency will behave as if they are resonant. As we slow the frequency variation of our eikonal wave asymptotically and watch the effect for longer time we have two effects: 1) we are in resonance for a longer time causing a larger dissipation and 2) since we are watching longer, less of the nearby frequencies contribute any dissipation. As time goes to infinity there is net dissipation which arises from arbitrarily small neighborhoods of the resonant frequency.

Let us see this explicitly in an example. We consider the driven oscillator:

$$\ddot{x} = -x + F(\epsilon t)\cos(\omega(\epsilon t)t), \tag{13.7}$$

where we assume that the resonance occurs at $t = 0$ (and nowhere else)

$$\omega(0) = 1 \tag{13.8}$$

and that we pass through resonance with non-zero speed:

$$\left.\frac{d\omega}{dt}\right|_{\epsilon=0} \equiv \omega' > 0. \tag{13.9}$$

Let us denote $\dot{x}$ by v. We wish to solve this asymptotically as $\epsilon \to 0$. We begin (as in the methods discussed in chapter 3) by going to a "rotating" system of coordinates in phase space:

$$X = x\cos t - v\sin t \qquad V = v\cos t + x\sin t. \tag{13.10}$$

The dynamics of these coordinates is

$$\begin{aligned}\dot{X} &= \dot{x}\cos t - x\sin t - \dot{v}\sin t - v\cos t \\ &= -F\cos(\omega t)\sin t \\ &= \frac{-F}{2}(\sin((\omega+1)t) + \sin((\omega-1)t))\end{aligned} \tag{13.11}$$

and

$$\begin{aligned}\dot{V} &= \dot{v}\cos t - v\sin t + \dot{x}\sin t + x\cos t \\ &= F\cos(\omega t)\cos t \\ &= \frac{F}{2}(\cos((\omega+1)t) + \cos((\omega-1)t)).\end{aligned} \tag{13.12}$$

We may integrate these equations using the method of stationary phase, which was discussed in section 7.1.3. Introducing the rescaled time $\tau \equiv \epsilon t$ and using stationary phase, we see that to leading order the change in X in passing through the resonance

is

$$\begin{aligned}
\Delta X &= C + \int_{-T/\epsilon}^{T/\epsilon} \frac{-F}{2}(\sin((\omega+1)t) + \sin((\omega-1)t))dt \\
&\sim C + \int_{-T/\epsilon}^{T/\epsilon} \frac{-F}{2}\sin((\omega-1)t)dt \\
&\sim C + \int_{-T/\epsilon}^{T/\epsilon} \frac{-F}{2}\sin(\omega'\epsilon t^2)dt \\
&\sim C - \frac{1}{\epsilon}\frac{F(0)}{2}\int_{-T}^{T}\sin(\omega'\frac{\tau^2}{\epsilon})d\tau \\
&\sim C - \frac{1}{\epsilon}\frac{F(0)}{2}\int_{-\infty}^{\infty}\sin(\omega'\frac{\tau^2}{\epsilon})d\tau \\
&\sim C - \frac{1}{\epsilon}\frac{F(0)}{2}\Im\int_{-\infty}^{\infty}\exp(\frac{i\omega'\tau^2}{\epsilon})d\tau \\
&\sim C - \frac{1}{\epsilon}\frac{F(0)}{2}\Im\int_{-\infty}^{\infty}\exp(\frac{i\omega'\tau^2}{\epsilon})d\tau \\
&\sim C - \frac{1}{\epsilon}\frac{F(0)}{2}\Im\left(\frac{\sqrt{2\pi\epsilon}}{\sqrt{2\omega'}}e^{i\pi/4}\right) \\
&= C - F(0)\sqrt{\frac{\pi}{8\omega'\epsilon}}.
\end{aligned} \tag{13.13}$$

Similarly,

$$\Delta Y = C_2 + F(0)\sqrt{\frac{\pi}{8\omega'\epsilon}}. \tag{13.14}$$

Thus asymptotically there is a net dissipation which grows as the reciprocal of the square root of the eikonal parameter. Notice that for any real system the initial conditions will determine whether the oscillator takes or gives up energy for a time independent of ϵ. The asymptotics we have introduced always beats this time for some value of ϵ. Thus the asymptotic system can have dissipation without constraint on the initial conditions whereas any real system may not have this behavior. I think this is a good example for seeing why an asymptotic approximation to a system (say Boltzmann's equation) may have properties like irreversibility and lack of reccurence

which are not possessed by any real system. These are new asymptotic concepts which validly apply to the system whenever the asymptotics is valid

13.3.6. Resonant Coupling of Eikonal Waves

For this kind of experiment we must really use eikonal driving since we are interested in the response as we turn on the oscillations. Traditionally in plasma physics one uses a linear analysis in a uniform system and so finds the normal modes for the evolution operator. These will be unphysical infinite plane waves and lead to subtleties like a singular spectrum (as in van Kampen modes) which obscures the physics of the analysis. It is interesting to rethink these analyses in terms of eikonal waves. For example, the response function of a harmonic oscillator at its natural frequency is infinite. If we excite it with a slowly varying frequency, we may use stationary phase to study the passage through resonance and we see that there is no infinity. Only in the asymptotic limit does the response lead to a pole on the natural frequency. In traditional analyses of the three wave interaction (which arises by retaining third order terms in the nonlinear coupling of waves) one studies the response of an infinite plane wave to the presence of two other infinite plane waves. In reality, given two eikonal waves in three dimensions, the resonance condition can only be satisfied on a 2-dimensional spatial surface which moves through space. The generated wave will include these geometrical aspects.

13.4. The String with a Spring

Let us now give an interesting example which shows how the effect of an infinite dimensional system on a finite dimensional piece may sometimes be modeled by dissipation. The idea is that, asymptotically in time, the infinite number of modes will only suck energy out of the system, leading to the appearance of dissipation.

The simple system we have in mind is a half infinite string whose end is attached to a mass on a spring (or any other shaking system). If we consider only string displacements which are damped at infinity, then asymptotically the string behaves as if it were quiescent (any energy going away from the end doesn't affect us, and all the energy going toward it reflects and is eventually leaving it).

Let the string's displacement w and its velocity v be functions of the position x along the string. Let the tension in the string be T (i.e. the force along the string) and the mass density be ρ. The transverse force that the string applies at its end is given by the transverse component of the tension, and for small displacements this is just the tension times the slope:

$$T\frac{\partial w}{\partial x}. \tag{13.15}$$

The wave equation arises from Newton's equation for a little piece of string, where the transverse force is the difference of the force due to the string on the left and the string on the right. Asymptotically this is proportional to the second derivative of the displacement, leading to the usual wave equation:

$$\rho\frac{\partial^2 w}{\partial t^2} = T\frac{\partial^2 w}{\partial x^2}. \tag{13.16}$$

This is known to be Hamiltonian. The solutions away from the end are sums of a function translating rigidly to the left and to the right with velocity

$$c = \sqrt{\frac{T}{\rho}}. \tag{13.17}$$

For a quiescent string, if we shake the end according to

$$w(0,t) = y(t), \tag{13.18}$$

then we get only a wave travelling away and the string displacement for all time is

$$w(x,t) = y(t - x/c). \tag{13.19}$$

What then is the force the string applies to our shaking apparatus? From the above it is

$$T \left.\frac{\partial w}{\partial x}\right|_{x=0} = T \frac{d}{dx} y(t - x/c)|_{x=0} = -\frac{T}{c}\dot{y}(t). \tag{13.20}$$

If the shaker was a harmonic oscillator

$$m\ddot{y} = -ky, \tag{13.21}$$

then we could forget about the string if we changed the equation to

$$m\ddot{y} = -ky - c\rho\dot{y}. \tag{13.22}$$

This is just a damped oscillator. Thus an infinite dimensional piece of a Hamiltonian system is replaced by an effective dissipation in the evolution of a finite dimensional piece. Notice that the exact form of the system to which the string is coupled is completely irrelevant. It is this fact which makes such replacement physically relevant (if it behaved dissipatively only in extremely special circumstances we would have no reason to think of it as a replaceable piece during simplification).

Chapter 14: Reinsertion in Area-Preserving Horseshoes

"It is said that Newton communicated to Leibniz an anagram somewhat like this: aaaaabbbeeeeeii, etc. Leibniz naturally understood nothing at all of this; but we who have the key know that this anagram meant, translated into modern language: I know how to integrate all differential equations; and this suggests to us that either Newton was very lucky or else he held a peculiar illusion."—Poincaré (1908) [Hirsch, 1984]

Over the last ten years, the idea of intrinsic stochasticity in the deterministic dynamics of low dimensional dynamical systems has joined the mainstream of physical thought. Every month physics journals carry many reports of chaos (as the phenomenon has come to be known) in new physical situations. Applications to plasma physics are given in [Smith and Kaufman, 1975], [Smith and Kaufman, 1978], and [Kaufman, 1979]. The chaos has always been there, but until recently the mathematical tools for noticing it were not well known in the physics community. Every example of chaos that I know of has within it a construction dubbed

by Smale the horseshoe in [Smale, 1967]. The essential phenomena may be seen in a taffy puller (and indeed Otto Rössler designed his attractor based on it). The key feature of a taffy puller is its continual stretching of the taffy. To keep it in a finite volume, the taffy must also be bent over. These two features lead to dynamics where the taffy is thoroughly mixed together. (This may be seen somewhat grotesquely by considering the fate of a fly which lands on the sweet smelling taffy. In on the order of ten cycles, there will be a piece of the fly in each piece of the taffy.)

14.1. The $2x$ mod 1 Map

The basic mathematical model of this stretching and mixing is the map from the circle to itself given by

$$x \mapsto 2x \mod 1. \tag{14.1}$$

This dynamics has periodic orbits of every period, dense orbits, and orbits which hop from the interval: $[0, 1/2]$ to the interval: $[1/2, 1]$ according to any desired (possibly random) sequence of L's and R's (p. 106 of [Cornfeld, Fomin, and Sinai, 1982]). This is easy to see by the method of symbolic dynamics. Call the left interval 0 and the right one 1. The interval in which a point lies is given by the first digit in the binary expansion of the number corresponding to that point. Since $2x$ mod 1 just shifts the binary point and lops off the integer part, we see that the sequence of hops is given by the sequence of digits in the binary expansion, but this is arbitrary. This map cannot appear in real dynamics since it is 2 to 1 (but variants

of it appear in projected dynamics all the time, for example, Lorenz's equations as in [Guckenheimer and Holmes, 1983] p. 276).

14.2. The Baker's Transformation

The trouble, of course, is that we are continually losing information about our intitial condition, so knowing where we are isn't enough to tell us where we came from. We can remedy this by adding another variable which keeps track of the lost information. The symbolic dynamics model is the space of *doubly* infinite sequences of 1's and 0's with the map being a shift. This clearly has the same stochastic properties, but is invertible. A dynamical realization may be found in the so called baker's transformation, named because of the similarity of the map to the experiences a baker puts dough through (particularly filo dough). The baker first rolls the dough out so that it is thinner and longer, he then cuts the elongated dough into two pieces, places one atop the other and repeats the process. The corresponding map is from the unit square to itself:

$$\begin{aligned} x &\mapsto 2x \mod 1 \\ y &\mapsto \begin{cases} \dfrac{y}{2} & \text{if } 0 \le x < \dfrac{1}{2} \\ \dfrac{y+1}{2} & \text{if } \dfrac{1}{2} \le x < 1. \end{cases} \end{aligned} \tag{14.2}$$

If we take y's binary expansion backward and adjoin it to x's binary expansion, then this map is exactly the shift on the doubly infinite sequence of the combined digits ([Cornfeld, Fomin, and Sinai, 1982] p. 9). Unfortunately, this still cannot appear as a return map because it is not continuous (due to the baker's knife).

14.3. The Horseshoe

If instead of pushing the two pieces of dough all the way together, we leave a gap, then we may imbed this in a smooth map which has all the interesting stochastic properties ([Guckenheimer and Holmes, 1983] p. 230). The horseshoe is such a map as shown in figure (14.1). Because of the gap (which under the map gets sent to a skinnier gap, etc.) the invariant set associated with the symbol shift is a Cantor set.

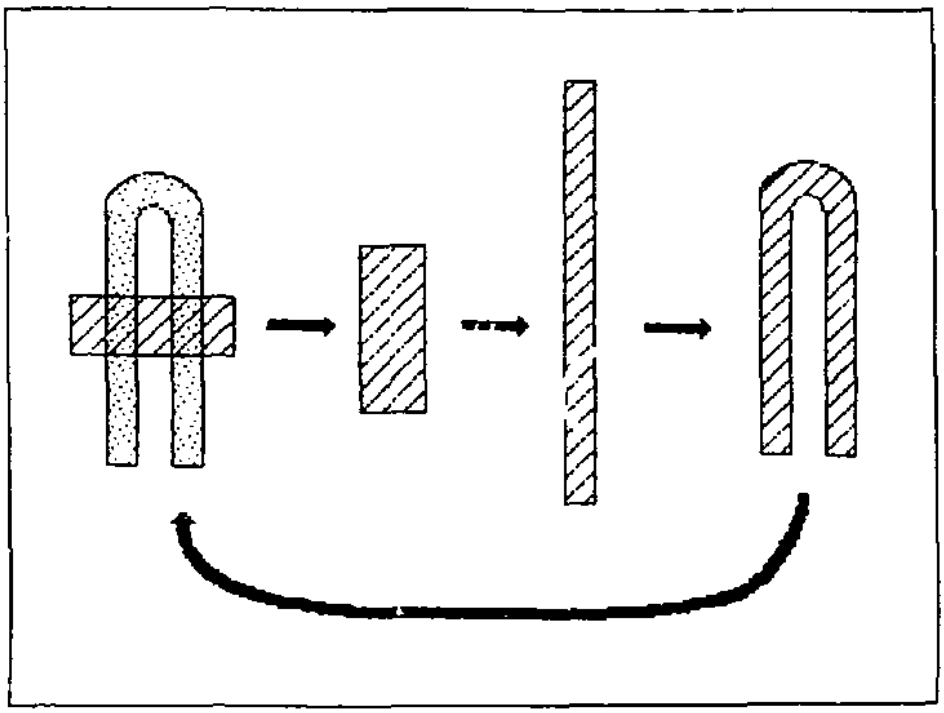

Figure 14.1: Smale's horseshoe map.

That such a map might appear in any natural systems is at first perhaps surprising, but in fact it is quite common as we now show. Consider a dynamical system whose state space is three dimensional (for example, the three dimensional

energy surface of a two degree of freedom Hamiltonian system). We may study the neighborhood of a periodic orbit by means of the two dimensional return map induced by the dynamics on a two dimensional surface transverse to the orbit. The orbit is a fixed point of the map and we consider the case where its linearization is hyperbolic with one eigenvalue inside the unit circle of the complex plane, and one outside. The stable manifold theorem states that there exists a smoothly imbedded two dimensional manifold in our state space consisting of all points which asymptote to our orbit as time goes to positive infinity. It must approach the orbit tangent to the linear attracting direction and the orbit. The corresponding unstable manifold theorem says a similar thing about the points that asymptote to our orbit under the time reversed dynamics. We will now see that the Poincaré return map must contain a horseshoe if the stable and unstable manifolds intersect transversally. The first image of a point of intersection must also asymptote to the orbit both forward and backward in time, implying that the stable and unstable manifolds must intersect again, as in figure (14.2) (repeating this argument shows that they must intersect an infinite number of times). As in the figure, if we choose an appropriate rectangle, some iterate of it gives us a horseshoe imbedded in our dynamics.

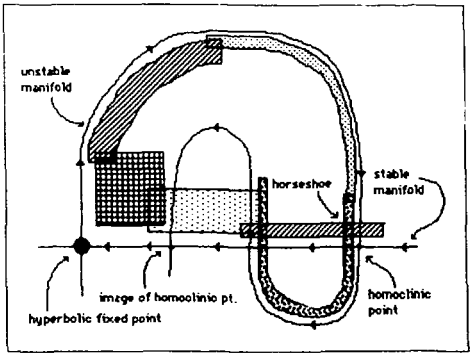

Figure 14.2: A transversal intersection of the stable and unstable manifolds implies the presence of a horseshoe for some iterate of the map.

14.4. Example of Horseshoes in Gyromotion

Let us give a plasma example which has horseshoes in the dynamics. Consider the two-dimensional x, y motion of a charged particle in a linear magnetic field

$$B_z = \alpha y, \tag{14.3}$$

near the region where it vanishes. Define the vector potential by taking $A_y = 0$, so that

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = -\frac{\partial A_x}{\partial y} = \alpha y, \tag{14.4}$$

leading us to take

$$A_x = -\frac{\alpha}{2} y^2. \tag{14.5}$$

As usual, let us set $e = m = c = 1$. The Hamiltonian is then

$$H = \frac{1}{2}((p_x - \frac{\alpha}{2}y^2)^2 + p_y^2). \tag{14.6}$$

and since it is x-translation symmetric, p_x is conserved. The equations of motion are

$$\dot{x} = \frac{\partial H}{\partial p_x} = p_x - \frac{\alpha}{2}y^2, \tag{14.7}$$

$$\dot{y} = \frac{\partial H}{\partial p_y} = p_y, \tag{14.8}$$

$$\dot{p}_x = -\frac{\partial H}{\partial x} = 0, \tag{14.9}$$

and

$$\dot{p}_y = -\frac{\partial H}{\partial y} = -(p_x + \frac{\alpha}{2}y^2)\alpha y. \tag{14.10}$$

Let us denote p_x's constant value by

$$p_x = p_x^0. \tag{14.11}$$

We know that by reducing by the x translation symmetry we may get equations involving only y. The y evolution is given by

$$\ddot{y} = \dot{p}_y = -\alpha p_x^0 y - \frac{\alpha^2}{2}y^3. \tag{14.12}$$

This is Duffing's equation, which has been well-studied. It represents the Hamiltonian dynamics of a particle moving in a potential given by

$$V(y) = \frac{\alpha}{2}p_x^0 y^2 + \frac{\alpha^2}{8}y^4. \tag{14.13}$$

Let us assume that α is positive. If

$$p_x^0 = \dot{x} - \frac{\alpha}{2}y^2 \tag{14.14}$$

is positive, then this potential has a single minimum. If p_x^0 is negative, then the potential well has two minima as shown in figure (14.3a). The phase portrait for this case is shown in figure (14.3b). Far from the origin, all orbits are periodic and encircle the origin. As we move in, we see that there are two stable elliptic fixed points and an unstable hyperbolic fixed point. The hyperbolic fixed point has two homoclinic separatrices which are susceptible to horseshoes under periodic driving. In fact, using the method of Melnikov, Holmes and Marsden have shown that the driven Duffing equation has transversal homoclinic points and therefore horseshoes ([Guckenheimer and Holmes, 1983] p. 184). If we force our particle in this magnetic field with a low amplitude wave in the y direction, it too will have horseshoes.

What do the orbits look like? For the unforced system, the particle can move in a straight line along the x axis with any velocity, as in figure (14.3c). Since the Duffing oscillator is bounded in y, every orbit has a maximal excursion in y, where $\dot{y} = 0$. Let us then study the orbits with initial condition

$$y = y_0 \qquad \dot{y} = 0 \tag{14.15}$$

for various values of p_x^0 and any x. When p_x^0 is positive, the y motion is represented by a Duffing oscillator with a single minimum in its potential. The x velocity

$$\dot{x} = p_x^0 + \frac{a}{2}y^2 \tag{14.16}$$

never goes negative and so the particle moves inexorably in the positive x direction as in figure (14.3d). When $p_x^0 = 0$, $\dot{x}$ vanishes at $y = 0$, but is positive elsewhere. This leads to particle motion which crosses the field reversal perpendicularly as in figure (14.3e). As p_x^0 becomes negative, the orbits actually move in the negative x

direction whenever

$$|y_0| < \sqrt{-\frac{2}{\alpha} p_x^0} \tag{14.17}$$

as in figure (14.3f). The Duffing equation now has two minima but we are in the oscillatory region that encircles them. As p_x^0 gets more negative, the loops overlap one another as in figure (14.3g and h) until successive loops actually coincide to give a "figure eight" orbit as shown in figure (14.3i). As p_x^0 gets still more negative, the net motion is in the negative x direction, while still in the positive direction at the tops and the bottoms of the loops. The particle path is made of alternating curliques about the field reversal line as shown in figure (14.3j and k). When

$$p_x^0 = -\frac{\alpha}{4} y_0^4, \tag{14.18}$$

we are on a separatrix of the Duffing equation and the particle makes but a single loop, asymptoting to $y = 0$ both forward and backward in time as shown in figure (14.3l). There is another orbit corresponding to the other separatrix which has y negative as in figure (14.3m). For

$$p_x^0 < -\frac{\alpha}{4} y_0^2, \tag{14.19}$$

we are inside the separatrix and the particle executes ordinary gyration as in figure (14.3n). When

$$p_x^0 = -\frac{\alpha}{2} y_0^2, \tag{14.20}$$

we are at a stable fixed point of the Duffing equation, $\dot{x} = 0$, and particle motion has stopped. For p_x^0 more negative, the particle gyrates in the region $y > y_0$.

Under oscillatory perturbation, there are orbits near the separatrix that go around one and then the other separatrix according to any (possibly random) sequence of 0's and 1's. The resulting particle orbit has corresponding upward and downward curliques, as shown in figure (14.3p).

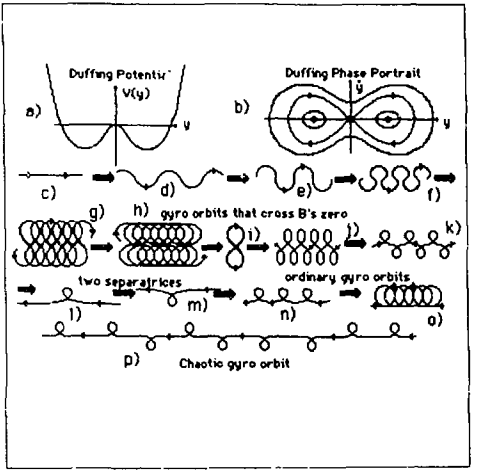

Figure 14.3: The Duffing potential, phase portrait, corresponding particle orbits, and a chaotic orbit.

14.5. Area Preserving Horseshoes and Reinsertion

The issue that we would like to address here concerns the structure of the stable and unstable manifolds in the area preserving (as arising from a Hamiltonian system) as opposed to the dissipative case. As we extend the stable and unstable manifolds in the dissipative case, we obtain the intricate pattern shown in figure (14.4). We will see that the area preserving case must be more complicated.

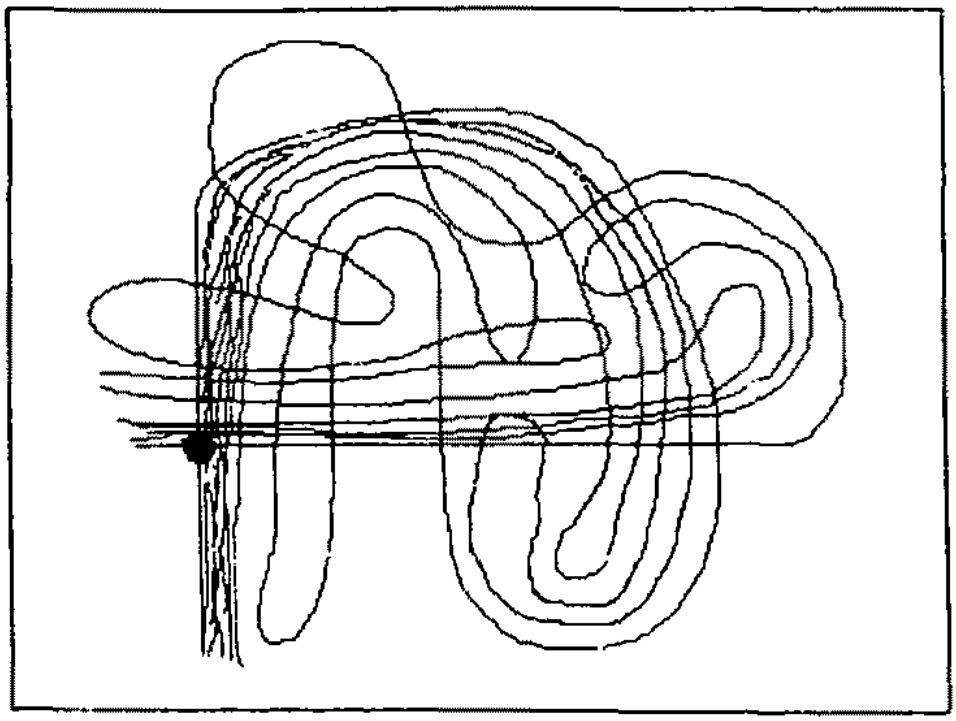

Figure 14.4: The structure of the stable and unstable manifolds in the dissipative case.

If we try this same construction in the area preserving case (figure (14.5)), we run into problems. As in that figure, the region A bounded by the stable and unstable manifolds is taken to the region A' of the same area. This is taken to A'' and so on. In real systems, this structure is often bounded between two curves

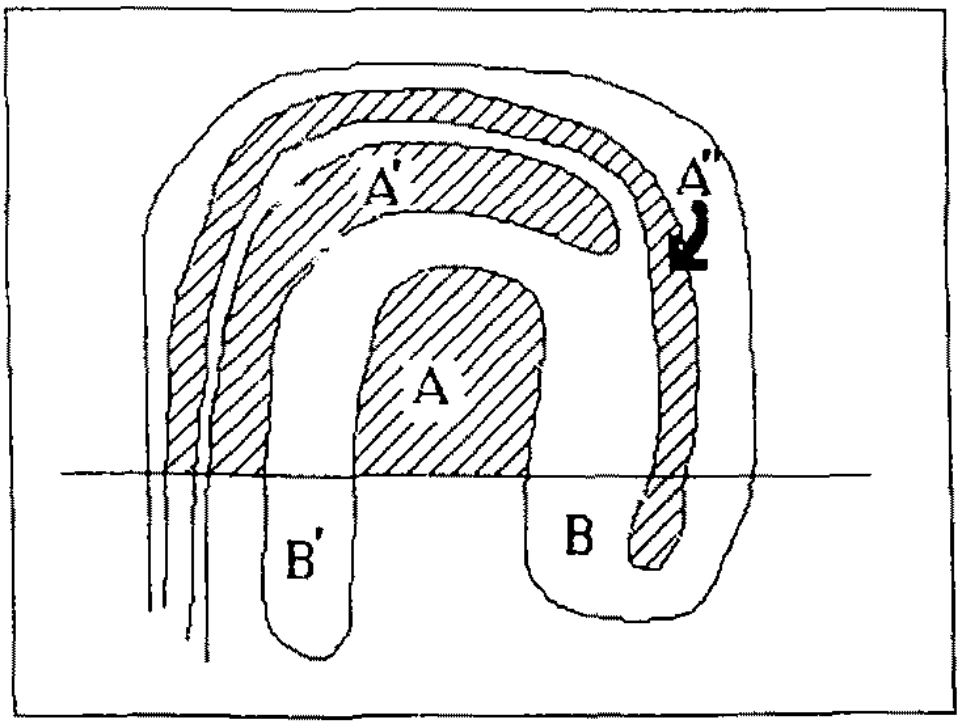

Figure 14.5: The seemingly paradoxical tongues which must intersect.

corresponding to Kolmogorov-Arnold-Moser tori ([Abraham and Marsden, 1978] p. 582) and so is constrained to a finite area. But because the tounges all have the same area and there are an infinite number of them, they must intersect one another (and in fact an infinite number of them must intersect). Because the whole picture is taken to itself under the map, each tongue must intersect an infinite number of others, including both an infinite number of images and preimages (if tongue T intersects $M^1 \cdot T$, then it must intersect $M^{-1} \cdot T$ as well as one can see by applying M^{-1} to both sets). Upon first examining the figure it is a mystery how these tongues can possibly intersect. Neither the stable or the unstable manifolds can cross themselves (since they are injectively immersed copies of the real line). One therefore concludes that if tongue A^N is going to intersect A, it cannot come

in through the top segment of A (which is a piece of the unstable manifold) but must come in through the bottom (which is a piece of the stable manifold and so is allowed to intersect the top of A^N's tongue). The question is: how does A^N get down to the bottom border of A? The unstable manifold makes essentially a complete loop (it includes the origin as it leaves vertically, and it limits on the origin on the right), leaving no room for A^N to sneak through. We quickly come to the conclusion that if A^N is going to get inside A, then one of the lower tongues B^M must as well and A^N gets in via B^M.

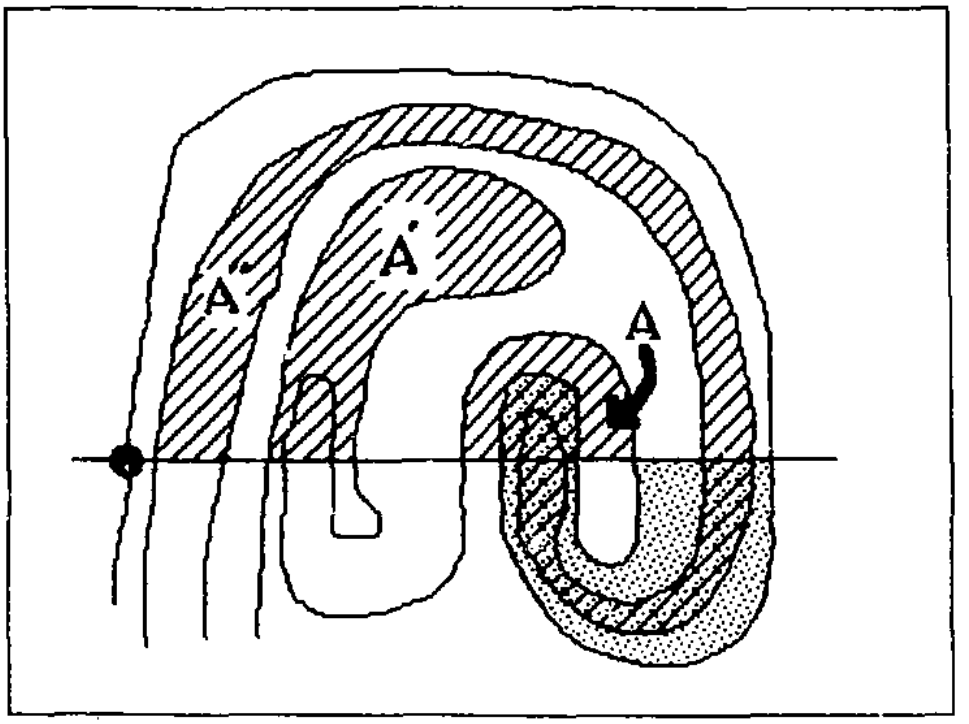

Figure 14.6: The Re-insertion of tongues forced by area preservation.

This implies the much more complex structure shown in figure (14.6) than I have seen described in the literature. The number of tongues between a tongue

and the first lower tongue to insert into it is an invariant of the map. Furthermore as we constrain the area within which this structure lies, relative to the area of a tongue, the point at which reinsertion must take place gets nearer and nearer (since a tongue must intersect more of its colleagues), until no structure is possible when the area of a tongue is larger than the total allotted area. Thus we see that Hamiltonian systems are forced to have a very intricate tongue structure. Recent work of [Holmes and Whitley, 1984] has studied the change in bifurcation sequences in "shoemaking" in going from the dissipative to the non-dissipative case. They find a continuum of different bifurcation sequences. It would be interesting to see how the intersection structure of the horseshoe tongues evolves in this sequence.

Chapter 15: Renormalization Group

Recently a number of previously intractable problems in several very different areas of physics have been successfully tackled using renormalization group techniques. Running through these treatments is a beautiful set of ideas which are fairly easy to understand, even though the actual calculations can become quite complex. The goal of this section is to present the essential concept in simple physical situations, where it is easy to see what is going on. We sketch the physical idea of renormalization in the original context of the statistical mechanics of spin lattices and then use it to prove the central limit theorem.

15.1. Scaling and Universality

It is a quite common physical occurrence that certain forms appear over and over in many different contexts. In statistical physics we see the Gaussian distribution e^{-x^2} and the Boltzman distribution $e^{-E/kT}$ appearing again and again. In electromagnetism we have the monopole and dipole potential functions, for example. In radiation theory there are the plane wave, dipole radiation, and the Airy

function near caustics. In quantum mechanics we continually see the free particle, the harmonic oscillator, and the rigid rotator. In dynamics there exist certain canonical forms for change (eg. the Hopf bifurcation, the saddle-node bifurcation, period-doubling cascades, and the breakdown of KAM tori) which appear in many contexts.

In many cases the effect of a physical system on the rest of the world is only felt on a large scale in space-time which averages over the detailed structure. It is a common tool in physics to study a particular instance of a system by thinking of it as a member of a family of related systems. Sometimes by looking at the family, as opposed to the individual, we see new regularities. This is the basis for perturbation theory which studies the case of interest by finding an easily soluble case nearby and studying how the differences change the behavior.

Often, one may formalize these heuristic notions by introducing a "space", which is often infinite dimensional, whose points represent physical systems. For the statistical systems, we "step back" from the system (i.e. look at it from a greater distance), and treat all behavior as if it were occuring on a smaller scale. For example, we might consider photographs of the ocean's surface taken at different heights. The rescaling is called *renormalization*. In the situations of interest, the renormalized system is again a member of our space, but corresponds to different values of th parameters. The ocean with a tidal wave may, from a distance, behave just like a ripple in the tide pools on the shore. The ripple may need to be in water with a different surface tension or viscosity, however. (Such ideas are actually used for special effects in movies. Since one usually has control of the scale of space and

time but not of viscosity, the small-scale waves in ocean scenes and the small-scale eddies in explosions often do not look right.) The renormalization map that rescales a system sometimes has a fixed point whose properties can be used to study nearby systems. For example, as we get further and further from the surface of the ocean, it looks calmer and calmer.

We shall not discuss the examples of renormalization in dynamical systems here. They work in much the same way, but the renormalization usually corresponds to looking at the system through a finer and finer microscope. For example, if we zoom in on a hyperbolic fixed point of a vector field, the flow becomes more and more like its linearization about the fixed point. A linear system is taken to itself under rescaling and so is a fixed point of the renormalization map.

15.2. Magnetic Spin Lattices

As an example, let us look at the Ising spin model for which Kadanoff introduced many of these scaling ideas (see, for example, the review article [Wilson and Kogut, 1974] and the references given there). We imagine an infinite checkerboard with a spin at the center of each square which may have a magnetic moment pointing either up or down. We imagine that nearby squares feel each other's magnetic fields and that it is energetically favorable for them to have their spins pointing in the same direction.

The whole system is in contact with a heat bath with temperature T. This tends to cause the spins to jostle between up and down. The higher the temperature, the more violent the jostling. The spin-spin interaction tends to prohibit the relative jostling of neighbors. If the temperature is very high, then the average kinetic energy of the degrees of freedom of the heat bath will be much greater than the alignment potential energy of a spin due to its interaction with its neighbors. In this case the spins will wiggle between up and down relatively independently of the state of their neighbors.

If the temperature is very low, then the average thermal energy will be much less than the interaction potential energy and the spins will only rarely be able to overcome it and flip relative to their neighbors.

When $T = \infty$ the interaction is completely negligible and the statistics of a spin are independent of its neigbors. The probabilities for up and down are the same and there is no correlation between the probability distributions for different spins. When $T = 0$ the thermal energy is completely negligible compared to the interaction

potential energy. The spins all point in the same direction, say up. The probability distribution for an individual spin has spin up with probability one and down with probability zero. Again the distributions for the individual spins are uncorrelated. Even if a spin happened to be pointing down (an infinitely improbable event), the probability distributions for the other spins would remain unchanged.

Now we know the statistics of the spins in two states: $T = 0$ and $T = \infty$. What do states near these two look like? If we are at a high but finite temperature, then spins are still kicked between up to down, the probability of each being $\frac{1}{2}$. Now, however, the statistics of one spin depend slightly on the state of its neighbors. There is some tendency for neighboring spins to flip together and one will see little patches of spins pointing the same direction. If we know a spin is up, then its nearest neighbors are slightly more likely to point up than down. Next nearest neighbors are affected even less, and one can show that the correlation decays exponentially as we move away from the upward pointing spin. The patches of coherent spins are finite in extent and get smaller as the temperature gets higher.

When the temperature is low but not zero, spins are still much more likely to point up than down. Now, however, if a spin is flipped to down this slightly affects the neighboring spins. Nearest neighbors have a slightly higher probability for flipping than usual. Again the effect dies off exponentially. The state looks like a sea of upward pointing spins with little islands of spins which point down. As the temperature is decreased, the islands get smaller.

We now introduce the idea of renormalization. If our eyes are blurry, when we step away from the infinite checkerboard we will only notice the average spin

over a region. If we are near $T = \infty$, then as we step back, neighboring spins become a single spin to us and we are ever more unable to resolve the last vestiges of correlation among neighbors. The further we go the more the system's statistics look like those of $T = \infty$. We may codify this idea, by introducing a (typically infinite dimensional) space whose points represent the statistical state of our spins (i.e. the space of probability distributions on the space of spins). The operation of "stepping back" or *rescaling* takes one distribution and gives us another one. We may therefore think of this as a mapping on this big space and we denote it by R which stands for "renormalization group operator". The infinite temperature state (i.e. the probability distribution is just an uncorrelated product of half up and half down distributions for each spin) is a fixed point of this mapping. We have just indicated that all nearby states get even closerto this fixed point under the renormalization map, and so it is a stable fixed point.

Similarly, the state representing zero temperature (again an uncorrelated product of distributions for each spin, each of which is probability one for being up (actually the real one is a sum of two uucorrelated states on all spins; one for up and one for down)) is a fixed point. If we are near zero, then as we step back, the "cloud" of spins that an errant spin takes with it on the rare occasion of a flip gets smaller and smaller until in the limit there is no such cloud. So zero temperature is also a stable fixed point.

There is clearly a path from zero to infinite temperature labeled by the temperature itself. At some point on this path we must leave zero's basin of attraction and enter infinity's (it is not hard to show that there are . r stable fixed points).

We expect there to be a codimension-one surface that forms the boundary between these two basins. There is and it is the stable manifold of a third fixed point with a 1-dimensional unstable direction. While the correlation length of neighboring spins was zero for the other two fixed points, it is infinite for this one (it must be either zero or infinity, since these are the only two numbers that are equal to themselves when multiplied by a rescaling factor). The corresponding temperature is called the critical temperature and the point on a path where it intersects the separatrix is called a critical point. Because there are fluctuations on all scales there, it is often associated with such phenomena as critical opalescence (where a normally clear substance becomes milky white due to fluctuations which can scatter all wavelengths of light). If we consider the place at which our path pierces this codimension 1 stable manifold under applications of R, we see that the path asymptotes to the 1-dimensional unstable manifold (see figure (15.1)). Thus this represents a universal path from zero to infinite temperature, and the way it crosses the separatrix will represent that of any path seen from far enough away. In particular, the unstable eigenvalue will tell how fast things scale as we cross the critical surface and may be used to calculate the universal critical exponents corresponding to this fixed point.

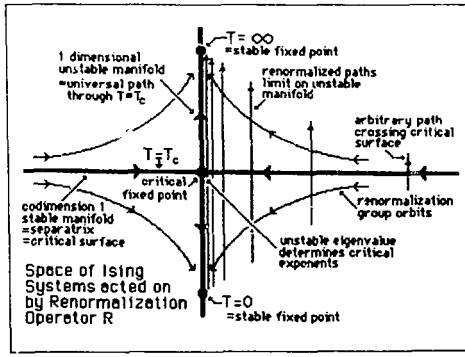

Figure 15.1: The space of spin probability distributions with two stable fixed points corresponding to zero and infinite temperature and a co-dimension one unstable fixed point corresponding to the critical point.

15.3. The Central Limit Theorem

One very old example of universality is the prevalence of Gaussian probability distributions in the statistical description of physical phenomena. The mathematical theorem describing the approach of the distribution of a sum of random variables to Gaussianity is the central limit theorem. Khinchin based his development of statistical mechanics on this theorem ([Khinchin, 1949]). It is of interest to look at this theorem from a renormalization group perspective.

A typical physical example, like Brownian motion, involves some distribution of random forces on a small time scale whose average effect on a long time scale

is of interest (see, for example, [Wax, 1954]). One may consider the problem to be that of finding the force distribution as a function of time scale starting with the given one on the smallest scale and hopefully asymptoting to a universal one on long time scales. In our renormalization group approach we define a mapping R from the space of distributions D to itself, which integrates out the next smaller time scale and renormalizes the variables so as to return the problem to one of the original form. We will show how a simple version of the central limit theorem fits into this context.

D will be the space of distributions $\rho(x)$ which are normalized:

$$\int_{-\infty}^{\infty} \rho(x)\, dx = 1, \tag{15.1}$$

have zero mean:

$$\int_{-\infty}^{\infty} x\, \rho(x)\, dx = 0, \tag{15.2}$$

and a constant nonzero finite dispersion:

$$\int_{-\infty}^{\infty} x^2 \rho(x)\, dx = \sigma^2. \tag{15.3}$$

The distribution of $x = x_1 + x_2$ where x_1 and x_2 are individually and independently described by ρ is the convolution

$$\rho^{(1)}(x) = \int_{-\infty}^{\infty} \rho(x - y)\, \rho(y)\, dy. \tag{15.4}$$

since $x = (x - y) + (y)$ for $-\infty < y < \infty$ describes all ways of decomposing x into a sum. It's easy to see that the dispersion squared of $\rho^{(1)}$ is twice that of ρ. Thus

$$\sigma^2 = \frac{1}{2} \int_{-\infty}^{\infty} x^2 \rho^{(1)}(x)\, dx = \int_{-\infty}^{\infty} x^2 \rho^{(1)}(\sqrt{2}x)\sqrt{2}\, dx. \tag{15.5}$$

To keep the normalization and the dispersion constant we define the renormalization operator as

$$R\rho(x) = \sqrt{2}\int_{-\infty}^{\infty} \rho(\sqrt{2}\,x - y)\,\rho(y)\,dy. \tag{15.6}$$

We would like to study R by looking for fixed points and studying their stability. By two changes of coordinates on D we will actually turn R into a linear operator.

We first label a distribution by its Fourier transform:

$$\hat{\rho}(k) = \int_{-\infty}^{\infty} e^{ikx}\rho(x)\,dx. \tag{15.7}$$

The moment conditions on ρ turn into

$$\hat{\rho}(0) = 1 \qquad \frac{d}{dk}\hat{\rho}(0) = 0 \qquad \frac{d^2}{dk^2}\hat{\rho}(0) = -\sigma^2. \tag{15.8}$$

We see that $\hat{\rho}$ has a quadratic maximum of value 1 at $k = 0$. Because convolution turns into multiplication under Fourier transform, in these coordinates the renormalization operator R becomes

$$R\hat{\rho}(k) = \left(\hat{\rho}(\frac{k}{\sqrt{2}})\right)^2. \tag{15.9}$$

It is already easy to see why Gaussians will arise. Squaring emphasizes large values compared to small ones. After repeated squaring, only the quadratic maximum of $\hat{\rho}$ will play any role and so $\hat{\rho}(k)$'s behavior will be the same as that of $1 - \frac{\sigma^2}{2}k^2$. But

$$\text{as} \quad N \to \infty, \qquad \left(1 - \frac{\sigma^2}{2}\left(\frac{k}{\sqrt{2}^N}\right)^2\right)^{2^N} \to e^{\frac{-\sigma^2 k^2}{2}}. \tag{15.10}$$

This is very similar to the case of period doubling in one dimensional maps where the Feigenbaum scaling is determined by the quadratic maximum.

To make R linear we now choose as our coordinates

$$\gamma(k) = \log(\hat{\rho}(k)). \tag{15.11}$$

γ satisfies the conditions

$$\gamma(0) = 0, \qquad \frac{d\gamma}{dk}(0) = 0, \qquad \frac{d^2\gamma}{dk^2}(0) = -\sigma^2 \tag{15.12}$$

and R takes the form

$$R\gamma(k) = 2\gamma\left(\frac{k}{\sqrt{2}}\right) \tag{15.13}$$

which is indeed linear. The "eigenfunctions" of rescaling are powers of k and so we expand $\gamma(k)$ in a Taylor series:

$$\gamma(k) = -\frac{\sigma^2}{2}k^2 + a_3k^3 + a_4k^4 + \dots \tag{15.14}$$

R takes $(a_3, a_4, \dots)$ into $(a_3/\sqrt{2}, a_4/\sqrt{2}^2, \dots)$. Thus $(a_3 = 0, a_4 = 0, \dots)$ is the unique attracting fixed point. This fixed point in the other coordinates is

$$\begin{aligned} \gamma^*(k) &= -\frac{\sigma^2}{2}k^2, \\ \hat{\rho}^*(k) &= e^{\frac{-\sigma^2k^2}{2}}, \\ \text{and} \quad \rho^*(x) &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}. \end{aligned} \tag{15.15}$$

Thus every distribution asymptotically approaches a Gaussian with the same dispersion under this mapping.

15.4. A Poor Man's Feigenbaum Number

One of the great discoveries of the last decade was the universality of certain aspects of period-doubling cascades. Many physical systems undergo a series of *bifurcations* or sudden changes in their behavior with the increase of some parameter. Typical parameters, like the Reynold's number, represent the strength of driving or energy input to the system. An example to think about is the behavior of the stream of water from a faucet as the water pressure is increased. Initially the flow is steady but at some point oscillations set in and eventually the stream dynamics becomes chaotic. Another example is a flag waving in the wind. As the wind gets stronger, the flag begins fluttering and then waving chaotically. A last example is river meanders. A high viscosity river (made of honey, say) goes straight down a hill. As the viscosity is lowered, the river's path meanders from side to side. As discovered by unfortunate farmers who built their farms on the rich soil between meanders of rivers, the meanders slowly move downstream. The state of a meandering river is therefore periodic in time (at any point on the hill the river position moves periodically from side to side). I don't know if any further bifurcations have been observed in this example.

For small enough values of the parameter, each of these systems is in a stable equilibrium state with no time variation. As the parameter increases, a *Hopf* bifurcation occurs at a particular value and the system undergoes oscillations described by a stable limit cycle. As the parameter is further increased in many systems, the period of the oscillation doubles at some point. On every other cycle, the system doesn't quite come back to the starting point causing the period to be twice

what it was originally. It is interesting that many systems continue to have these period-doubling bifurcations as the parameter increases. This phenomenon is called a *period-doubling* cascade. Successive doublings occur for closer and closer parameter values. There is a special parameter value, called the critical value, at which an infinite number of doublings have occured and beyond which the dynamics is very complicated and often chaotic. Feigenbaum discovered that the period doubling parameter values approach this limiting point geometrically, and that the rate of approach is a universal number, called Feigenbaum's number: 4.669...

[Crawford and Omohundro, 1984] gives a geometric picture of period-doubling in the state space of the system and discusses the phenomenon of knotted period-doubled orbits. Here we would like to sketch Feigenbaum's renormalization argument (for more information, see [Guckenheimer and Holmes, 1983], p. 346 and the references given there). Feigenbaum obtained his number to high accuracy using a Cray supercomputer. We will show how to obtain it to within 25 percent on the back of an envelope.

Period-doubling is usually studied by looking at the Poincaré return map for a periodic orbit. This is the mapping obtained by considering the effect of the dynamics on points near the periodic orbit that lie on a codimension-one sheet which cuts the orbit transversally. Each point on the sheet flows along until it hits the sheet again. We wish to study the mapping of the sheet to itself that this defines. A periodic orbit with a period near that of the one under study is represented by a fixed point of the mapping. Orbits of higher period are fixed points of some iterate of the mapping. Period doubling occurs when an eigenvalue of the linearization

of the mapping at a stable fixed point goes through -1. This has the effect of making the fixed point unstable and creating a periodic orbit of period-two nearby. All of the interesting behavior occurs along the eigendirection of the eigenvalue which goes through -1. It is for this reason that period-doubling of systems with many-dimensional state spaces may be effectively studied by considering only one-dimensional maps.

The key features of the period doubling cascade arise in any one-dimensional map with a quadratic maximum. We may consider

$$f_\mu(x) = 1 - \mu x^2 \tag{15.16}$$

on the interval $x \in [-1, 1]$. As μ varies, the width of the hump changes and period-doubling ensues. Let us use the term *hump map* to denote even maps of the interval $[0, 1]$ to itself, which have a single quadratic maximum at $x = 0$ with value 1.

The renormalization picture of period-doubling is based on the observation that for any hump map f, its first iterate $f \circ f$ again looks like a hump map when we consider only a smaller range of x's and invert and magnify the value. Furthermore, when f is undergoing the nth stage of period-doubling, $f \circ f$ is undergoing only the $n-1$st stage. We are therefore motivated to introduce a renormalization operator R on the space of hump maps that takes a map to a rescaled first iterate:

$$T[f](x) \equiv \alpha f(f(\frac{x}{\alpha})). \tag{15.17}$$

The rescaling parameter α is chosen so that given a hump map, T produces another one. Since $f(0) = 1$, we see that $f(f(0)) = f(1)$. For this to be one, we must define

$$\alpha = \frac{1}{f(1)}. \tag{15.18}$$

Since period-doubling requires only one eigenvalue to be pushed through -1, the set of hump maps undergoing any order of period doubling is codimension-one (i.e. one dimension less than the full space of hump maps). We noted that the renormalization operator T takes the sheet of order n period doubling to the sheet of order $n-1$. T has a fixed point f^* in the space of maps. Since the renormalization halves the period of periodic orbits, a fixed point either has no periodic orbits or orbits of every power of two period. It turns out that f^* has these periodic orbits and is at the accumulation point for period-doubling (where it has just doubled an infinite number of times). f^* has a 1-dimensional unstable manifold and a codimension-1 stable manifold. All maps that approach f under repeated application of the renormalization operator must also have just period-doubled an infinite number of times. Thus f^*'s stable manifold is the codimension-one sheet of maps at criticality. T^{-1} takes sheets of period-doubling to sheets of higher and higher order period-doubling, which must eventually limit on f^*'s stable manifold. The rate at which the approach to this manifold occurs is given by the unstable eigenvalue of the linearization of T at f^*, which is therefore Feigenbaum's number.

To actually calculate this number, one may employ various numerical tricks which amount to projecting the entire hump space onto some finite dimensional approximating space and carrying out the analysis there. We may actually carry this out by hand for an extremely crude one-dimensional approximation. We consider the family

$$f_\mu(x) = 1 - \mu x^2 \tag{15.19}$$

for different values of μ as the approximating one-dimensional space. We project

arbitrary hump maps to this space by truncating their Taylor series about 0 at the quadratic term. The action of the projected T on this space is then

$$\begin{aligned} T[f_\mu](x) =& \alpha f_\mu(f_\mu(\frac{x}{\alpha})) \\ =& \alpha(1-\mu(1-\mu(\frac{x}{\alpha})^2)^2) \\ =& \alpha - \alpha\mu + 2\alpha\mu^2\frac{x^2}{\alpha^2} - \alpha\mu^3\frac{x^4}{\alpha^4} \\ =& (\alpha - \alpha\mu) + \frac{2\mu^2}{\alpha}x^2 - \frac{\mu^3}{\alpha^3}x^4. \end{aligned} \tag{15.20}$$

To get this to have 1 as the constant term we must choose the renormalization factor as above:

$$\alpha = \frac{1}{f(1)} = \frac{-1}{\mu - 1}.$$

Using this and truncating away the cubic terms gives

$$T[f_\mu](x) = 1 + 2\mu^2(1-\mu)x^2.$$

Our space can be coordinatized by μ and T then has the form

$$T(\mu) = (\mu - 1)2\mu^2 = -2\mu^2 + 2\mu^3. \tag{15.21}$$

The fixed point is labelled by μ^* which satisfies

$$T(\mu^*) = \mu^* = 2\mu^{*3} - 2\mu^{*2}. \tag{15.22}$$

We find the solution to the resulting quadratic equation to be:

$$\mu^* = \frac{1+\sqrt{1+2}}{2} = \frac{1+\sqrt{3}}{2}. \tag{15.23}$$

The eigenvalue is then obtained by taking the derivative of T at this fixed point:

$$\left.\frac{dT}{d\mu}\right|_{\mu^*} = (6\mu^2 - 4\mu)|_{\mu^*} = 4 + \sqrt{3} \sim 5.7. \tag{15.24}$$

This isn't too bad considering the crudity of the approximation. We could have kept quartic or higher terms to eventually get any desired accuracy (unfortunately these require finding roots of quartic and higher order polynomials, which is hard without a computer).

Chapter 16: Symplectic Thermodynamics from Maximum Entropy

"The formulation is mathematically equivalent to the more usual foundations. There are, therefore, no fundamentally new results. However, there is a pleasure in recognizing old things from a new point of view. Also, there are problems for which the new point of view offers a distinct advantage."—R. P. Feynman introducing path integrals in [Feynman, 1948].

We have seen that the underlying geometry of classical mechanics is symplectic geometry and that many physically important structures result from this. We have also seen that these geometric structures arise in a natural way as the mathematics of the asymptotic limit of an underlying wave theory. When one looks at classical thermodynamics, one sees many tantalizing indications of symplectic geometry. The Legendre transform plays an essential role, thermodynamically conjugate variables remind us of canonically conjugate variables, differential one-forms and their integrals around loops (suggesting important two-forms) abound in the theory of Carnot cycles, etc. Thermodynamics also arises as the mathematical structure of

the asymptotic limit of an underlying statistical mechanical theory. We will show that indeed the structure of thermodynamics is intimately based on symplectic geometry and that this structure arises naturally from the underlying statistical mechanics in a way that is suprisingly analogous to the wave case. Let us begin by describing some earlier inquiries into the geometry of thermodynamics and possible relations with wave asymptotics.

16.1. Previous Approaches to Geometric Thermodynamics

In his book [Tisza, 1966] of collected papers, Tisza describes his lifelong attempt to develop a unified theory of thermodynamics. His sixth paper on p. 235 in this volume is entitled: "The Geometrical Interpretation of the Formalism of MTE" (Macroscopic Thermodynamics of Equilibrium). In this chapter he points out that there is no natural metric on thermodynamic state space but that there is a natural volume element. He claims that there is more structure than just a volume and so introduces an affine structure and attempts to find symmetry groups of the theory. He claims to be unable to find out anything about the group he finds. He gives the affine geometric interpretation of Legendre transforms due to Pluecker, in terms of representing a curve in the plane by either its points or by the tangent lines to it (i.e. its image in the the dual projective space as we have discussed). He finally attempts to relate a so called "stiffness moduli" to the curvature, but makes the comment that curvature in a theory with no underlying metric is puzzling (there is curvature in affine geometry, however).

[Gilmore, 1981] p. 229 attempts to introduce a metric structure into thermodynamics using the Hessian of a certain generating function as the metric. He makes some interesting connections, but the fundamental basis of his metric appears obscure to me. He ends with a section on page 247 entitled: "Additional Questions". There he mentions the classical limit of quantum mechanics and the reconstruction of quantum mechanics in terms of path integrals. He asks if there is a similar way to reconstruct statistical mechanics from thermodynamics. He then gives Hamilton's equations of motion, writes some thermodynamic equations with a similar form and asks: "Is there an intrinsic geometric structure in $\Re^n \times \Re^n$ associated with this variational formulation of thermodynamics? Is this geometry associated with the symplectic or orthogonal group $Sp(2n)$ or $SO(2n)$ or some related real form?" but does not go any further in their elucidation.

In [Poston and Stewart, 1978] p. 237 they discuss thermodynamics and phase transitions in terms of catastrophe theory (apparently with the consultation of Gilmore, as indicated in the preface). They do not discuss the underlying geometry of thermodynamics, but noting the similarity of their analysis of certain phase transitions to caustics in asymptotic optics, comment: "It is interesting to speculate on the possibility of a unified asymptotic analysis, treating phase transitions as caustics in the 'matter wave' everything is made of."

In a series of papers beginning with [Souriau, 1970b], Souriau has addressed certain aspects of statistical mechanics and thermodynamics from a geometric point of view. He focusses on relativistic and cosmological issues and does not appear to consider the questions addressed here.

Quantum statistical mechanics often leads to similarities between essentially wave ideas and essentially statistical ideas. One introduces a density matrix ρ (the Wigner function is the corresponding probability distribution on phase space and so was introduced with statistics in mind!). The time evolution of the density matrix is given by

$$\rho = i[\rho, H], \tag{16.1}$$

where H is the Hamiltonian operator [Feynman, 1972]. The canonical density matrix at given temperature $T = k/\beta$ is

$$\rho(\beta) = e^{-\beta H}/\mathrm{Tr}(e^{-\beta H}). \tag{16.2}$$

If we introduce an unnormalized ρ, then it satisfies the equation

$$\frac{\partial \rho}{\partial \beta} = [\rho, H]. \tag{16.3}$$

This looks like the evolution equation where i times the inverse temperature β plays the role of the time. This formal similarity is behind all the uses that I have seen of path integrals in statistical mechanics as in [Feynman and Hibbs, 1965], 20 [Feynman, 1972], and 0 [Schulman, 1981]. This is great for calculating partition functions, but it is hard to see any deep physical significance for their relation. We will present an alternative approach based on the maximum entropy formalism where it is very easy to see the physical significance.

The last connection between waves and statistics that I have seen mentioned by other authors is that the eikonal first amplitude transport equations may be written in a form that looks like the mass and momentum density fluid transport

equations of Euler (which arise from moments of the statistical BBGKY hierarchy). One might argue that eikonal waves represent photon gases for which a fluid theory is appropriate.

The book [Kijowski and Tulczyjew, 1979] describes a symplectic structure

$$\omega = dV \wedge dp + dT \wedge dS \tag{16.4}$$

on the the four-dimensional state space of thermodynamics, where (V, p, T, S) are volume, pressure, temperature, and entropy. The authors show that in the case of an ideal gas, the equations of state,

$$\begin{aligned} pV &= RT, \\ pV^{\gamma} &= ke^{S/cv}, \end{aligned} \tag{16.5}$$

define a Lagrangian submanifold with respect to this symplectic structure. If we view this symplectic manifold as a cotangent bundle with base coordinatized by (V, S) or (V, T) or (p, T) or (S, p), this Lagrangian submanifold is the graph of the differential of the internal energy, the Helmholtz free energy, the Gibbs free energy, and the enthalpy, respectively. Unfortunately, the authors do not give any reason for this structure.

We will show here that the principle of maximum entropy as applied to statistical mechanics leads naturally to this symplectic structure. We can see quite easily why the equation of state manifold should be Lagrangian. If we describe a loop of states on this manifold (i.e. a Carnot cycle), then the integral over the surface bounded by this loop of $dV \wedge dp$ gives the work done and the integral of $dT \wedge dS$ gives the heat gained. The first law of thermodynamics (energy conservation) says that these must be equal and opposite, so the symplectic structure which is their

sum must vanish on our manifold, which is thus Lagrangian. It is interesting that each term in the symplectic structure has its own physical interpretation

16.2. Seven Approaches to the Maximum Entropy Formalism

The maximum entropy formalism is a very powerful statistical tool, introduced in the papers reprinted in [Jaynes, 1983], which gives a prescription for finding the "least biased" probability distribution consistent with any known data. If we have a discrete number of possibilities and no data to distinguish them, then symmetry forces us to choose the distribution which makes them equiprobable. If we have some information about the distribution, say its mean value, then we would like a prescription to choose among all possible distributions with that mean value.

Assuming there is such a prescription that always gives the same answer in the same situation and is uniform across number of possibilities, it has been shown to be unique (p. 16 of [Jaynes, 1983]). In fact one must choose that distribution coosistent with any known data which maximizes the entropy defined as the sum over states of

$$-p \log p, \tag{16.6}$$

where p is the probability of a state. It is easy to see that if nothing is known, this gives the equiprobable distribution. The basic requirement in the general case is that if we partition the elementary events into subsets and call membership in these subsets the elementary events of a new distribution, then applying one's prescription should give the same distribution in each situation. This is in some respects a renormalization group idea.

16.2.1. Axiomatic Subjective Approach

The work which introduced information theory [Shannon, 1948], gives an axiomatic characterization of the information entropy which applies equally well to the case at hand. Given k possible outcomes of an experiment, with the probabilities $p_1, \ldots, p_k$, one would like a measure of the uncertainty in the value measured in each trial. Shannon requires of such a measure $H(p_1, \ldots, p_k)$, that it satisfy three axioms. 1) H should be continuous as a function of the p_i's. 2)If all the p_i's are equal, then H should be a monotone increasing function of k (more equally likely outcomes means more uncertainty). 3) If a choice is broken into two choices, then H should be a weighted sum of the individual choices. Shannon gives the example

$$H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2} H(\frac{2}{3}, \frac{1}{3}). \tag{16.7}$$

Here we have three possible outcomes, say A, B, and C, with probabilities of 1/2, 1/3, and 1/6 respectively. We may alternatively view this as two events: A and D, with probabilities 1/2 and 1/2. D represents the occurance of either B or C. The total uncertainty is the uncertainty in the A vs. D choice, plus the uncertainty in choosing B vs. C weighted by a factor of 1/2 (since this choice only arises half the time).

By approximating the probabilities by rational numbers and breaking the elementary events up into a number (the least common multiple of the denominators) of equally likely events, we may reduce the problem to equiprobable distributions. But for them the third property forces H to be a logarithmic function of the number. For arbitrary distributions, we find that $-p \log p$ (or some multiple of it)

is the unique H satisfying the requirements. To be unbiased, we should choose that distribution which has the least information about our experiment that is still consistent with the known data. We should therefore choose that consistent distribution that maximizes this entropy. This approach chooses the distribution on the basis of not assuming information that we do not have and may therefore be considered "subjective". This allows it to be applied in many circumstances where the "objective" view of probability as frequency in large number of trials doesn't have any meaning. For example, we may ask for the best prediction of something on the basis of one observation.

16.2.2. Counting Sequences of Trials

To see another place where the formula for the entropy comes from, we consider as elementary events, sequences of N trials of the basic experiment and let N go to infinity. The law of large numbers says that it is extremely likely for the number of trials with a given value in a sequence divided by N to be the probabilty of that value. Let us therefore consider all sequences with $n_1 = p_1 \cdot N$ entries with the first value, $n_2 = p_2 \cdot N$ entries with the the second value, and so on up to the number of possible measured values k. The least biased choice of $p_1, p_2, \ldots, p_k$, is that which is consistent with the known data and which maximizes the number of allowed measurement sequences. How many such sequences are there for given p_i's? We may lay our required measurements down in $N!$ ways, but permutations of the n_i with the same value don't change the measurement sequence. Thus the number

of sequences is

$$= \frac{N!}{n_1!\, n_2! \ldots n_k!} \,. \tag{16.8}$$

Maximizing this is equivalent to maximizing its logarithm. Because the p_i's are fixed as N gets large, all the n_i's get large as well (if they are not zero). We may therefore asymptotically use the crudest Stirling approximation:

$$\log n! \sim n \log n - n, \tag{16.9}$$

for N and each of the n_i's as N goes to infinity. We thus want to maximize

$$\begin{aligned} &(N \log N - N) - (n_1 \log n_1 - n_1) - \cdots - (n_k \log n_k - n_k) = \\ &= (n_1 + \cdots + n_k) \log N - n_1 \log n_1 - \cdots - n_k \log n_k \\ &= -n_1 \log(\frac{n_1}{N}) - \cdots - n_k \log(\frac{n_k}{N}), \end{aligned} \tag{16.10}$$

since

$$n_1 + \cdots + n_k = N. \tag{16.11}$$

Equivalently we want to maximize

$$\sum_{i=1}^{k} -p_i \log p_i. \tag{16.12}$$

which is the maximum entropy prescription.

16.2.3. Via Steepest Descents in Two Ways

The usual proof of Stirling's formula used above, uses steepest descents on the integral formula for the gamma function.

Because it is perhaps the simplest example of a combinatorial quantity giving rise to nice analytical asymptotics, it is worth examining the classical application of steepest descents to the integral formulation of the gamma function to obtain Stirling's expression for $x!$ as x gets large. It is interesting that the form of the integral is quite similar to those appearing in statistical mechanics.

The gamma function of $x+1$ is the Laplace transform with respect to t of t^x evaluated at 1 (and so the x-fold convolution of the Laplace transform of t). For large x, the expression

$$\begin{aligned} x! &= \Gamma(x+1) \\ &= \int_0^\infty t^x e^{-t} dt \\ &= \int_0^\infty e^{x \log t - t} dt. \end{aligned} \tag{16.13}$$

is of a form ripe for Laplace's method. The exponent is

$$x \log t - t, \tag{16.14}$$

with derivative with respect to t given by

$$\frac{x}{t} - 1, \tag{16.15}$$

and so has its maximum at

$$t = x. \tag{16.16}$$

We expand to second order about this maximum and extend the integration to infinity to get

$$= e^{x \log x - x} \int_{-\infty}^{\infty} e^{-\frac{1}{2x}(t-x)^2} dt. \tag{16.17}$$

This yields the Stirling approximation

$$x! \sim \sqrt{2\pi x}\, x^x e^{-x}. \tag{16.18}$$

We will show here two ways in which to view maximum entropy as coming from a steepest descents argument directly. These will allow us to make connections with eikonal wave theory and path integrals which use stationary phase.

In the first picture, we realize that our system is coupled to the rest of the world. The probability distribution of our system is determined by the state of the rest of the world and all possible such states must be considered in our choice of probability distribution. Thus we think of our desired probability distribution as being an "integral" over all possible distributions consistent with the known data. The distributions must be weighted by the number of external conditions that can produce them. This is the number of ways of rearranging states and is given by the integral over all states of the weight

$$\exp(\sum -p \log p). \tag{16.19}$$

For the systems we are interested in, the thermodynamic limit makes the exponent grow asymptotically (since changing the scale from x to $X = \epsilon x$ makes $dx = \frac{1}{\epsilon} dX$; equivalently, the number of states grows exponentially with the number of particles) and so "steepest descents" tells us that only the maximum entropy distribution con-

tributes. (The integral over all distributions must of course be interpreted asymptotically in terms of approximating sums, though Simon, 1979, discusses rigorous applications of steepest descents on infinite-dimensional spaces).

This is in the spirit of Gibbsian ensembles, but applied to the distributions themselves (i.e. an ensemble of probability distributions). One might imagine many copies of our system and each one has its phase space populated with a swarm of particles whose density is governed by *some* distribution. How many particle swarms correspond to a given distribution? We get multiple systems by exchanging particles but must divide by the number of exchanges between particles of the same probability. Think of chopping phase space into bins. Each probability distribution places a certain number of particles in each bin. The number of distinct ways of obtaining a given distribution is obtained by counting all permutations of the particles and dividing by the number of exchanges which leave the same particles in the bins (and so don't count as a distinct way of obtaining a distribution). Again the exponential of the entropy gives the number of possibilities in the limit as the number of particles becomes infinite and the binning becomes infinitesimal.

The second approach is an empirical one. We say that probability distributions are experimentally determined by measurement sequences and if we know only the distribution, its multiplicity should be the number of distinct measurement sequences that give rise to it. We may make the connection with path integrals in the following way. Consider the space of our observables M crossed with an interval in $\Re$. We can think of parametrized families of measurements as being paths in this space. With appropriate binning (as discussed in the introduction),

each path determines a probability distribution by integrating along the interval. Some paths are consistent with the known data, and we must consider them equally likely; others are not and they have zero probability. The expected distribution is then an integral over those distributions corresponding to the possible paths. We may convert this to an integral over possible *distributions*, if we include a weighting factor equal to the "measure" of paths corresponding to each distribution. But we have seen above that this is just $\exp(\sum -p \log p)$. Again we use "steepest descents" to conclude that the maximum entropy distribution is most likely.

It is perhaps artificial to think of measurements as parameterized by a real parameter (though time might serve this role). We might just as well consider sequences of measurements which asymptotically determine distributions. It is really the distributions that play the role of paths in Feynman's theory in any case. In fact, when one does quantum field theory via path integrals, the integral is over fields and so is quite similar to our integral over distributions.

Let us explicitly write down the formula for the "average" distribution which is analogous to the Feynman path integral. We want to sum over all allowed $\rho(Z)$'s weighted by the factor

$$e^{-S_I(\rho)/\epsilon} \tag{16.20}$$

(the ϵ arises from the scaling discussed earlier in this section). We want to normalize the resulting distribution as well. If we let $D(\rho)$ represent the "measure" on distribution space, and let C be the subset of distributions obeying any imposed

constraints, then

$$\rho_{\text{average}}(Z) \sim \frac{\int_C \rho(Z) e^{-\int \rho(Z)\log\rho(Z)dZ/\epsilon} D(\rho)}{\int_C e^{-\int \rho(Z)\log\rho(Z)dZ/\epsilon} D(\rho)}. \tag{16.21}$$

Integrating over Z, we see that this expression is correctly normalized (since each ρ in C is). Applying steepest descents as $\epsilon \to 0$, we pull out hte ρ with the maximum entropy from the integral in the numerator and the remaining integrals cancel leaving

$$\rho_{\text{average}}(Z) \sim \rho_{\text{max entropy}}(Z). \tag{16.22}$$

We can get the expected value of any functional of ρ by inserting it in place of ρ in the integrand of the numerator. In each case we may use steepest descents to pull it out of the integrand by evaluating it on the maximum entropy distribution. For example, by integrating over the constant energy surfaces each distribution ρ on phase space determines a distribution of energies (i.e. the density of states). By this argument, the average distribution over energies is exactly the one determined by the maximum entropy phase space distribution.

16.2.4. Via Probability in Three Ways

In the very interesting reference: [Tikochinsky, Tishoy, and Levine, 1984], the authors provide three "objective" justifications for the maximum entropy procedure to complement Jaynes' more "subjective" philosophy. Their first technique is to consider the known data to be a sequence of experimental samples and from consistency conditions and the reproducibility of the experiment, they deduce the maximum entropy criterion. This argument is very much like the sequence space

one we gave above, except that there is no need for long sequences in these authors' work.

Their second approach concerns the notion of most stable inference. The idea is that since the data inferred from real samples is likely to be slightly off, one should choose that consistent distribution that is least sensitive to errors in the data. This too leads to the maximum entropy distribution. Intuitively, this distribution is the most spread out that it can be, consistent with the data, and so changes the least as the data varies.

Their last approach uses the notion of sufficient statistics. In later sections we shall need to use Bayes' theorem, which allows one to calculate the probability distribution of a parameter that parameterizes a family of distribution functions, given the actual distribution. A sufficient statistic is a function of some number of sample points which contains all the information that the samples do as far as determining the value of the parameter. If the sample averages of the observed parameters serve as sufficient statistics for the mean value of those parameters, then the probability distribution of those parameters must in fact be the maximum entropy one. Thus if the sample average is all that can usefully be used in determining expectation values, we must have the maximum entropy distribution.

16.3. The Thermodynamic Limit

Our asymptotics will consist of taking the thermodynamic limit. We want the observables to be certain mechanical quantities, like the volume V, the total energy U, the numbers of various particles N_i, the magnetic moment μ, etc. These will all be taken at constant N, the total number of particles, because this will play the role of an asymptotic parameter. If we have n particle species only $n-1$ of the quantities N_i are really independent variables on our space (so pick the first $n-1$ say). We will thus assume that N is precisely known when applying maximum entropy. Let us symbolize the rest of the observables by the variable x, which is a vector in the observable vector space O. We will assume that experimentally only the mean values of the x after many measurement trials are known. We introduce the asymptotic parameter ϵ and let the total particle number scale as: $N = 1/\epsilon$. As N gets large, the boundary effects shrink and so the x really become extensive, and so proportional to N. We therefore introduce the "slow" rescaled (intensive) quantities: $y \equiv \epsilon x$. The behavior of the system expressed in terms of y as $\epsilon \to 0$ will give us the thermodynamic limit.

In our discussions we will often want to distinguish the mechanical variables y and their thermodynamically conjugate variables. Since the mechanical variables are additive when we couple systems, we will sometimes refer to them as "the extensive variables" (even though they are intensive with respect to the scaling of ϵ) and their thermodynamic conjugates (like temperature and pressure) as "the intensive variables" since these equalize in coupled sytems. This nomenclature is introduced merely to keep from repeating the awkward phrase "and their thermodynamically

conjugate variables".

16.3.1. The Density of States

The density of states available to the system with given x's will be denoted by $\Omega(x)$. So $\Omega(x)dx$ is a density on O whose integral over a region represents the number of microstates represented by that region. If the x's are large, then the number of states of a system is equal to the product of the number of states in each of two subsystems into which it decomposes (since interaction becomes irrelevant asymptotically). Thus

$$\Omega(y) = \Omega(y - \xi)\Omega(\xi), \tag{16.23}$$

as $\epsilon \to 0$. We may find the asymptotic dependence of Ω on ϵ by taking the logarithm:

$$\log\Omega(y) = \log\Omega(y - \xi) + \log\Omega(\xi), \tag{16.24}$$

and taking $\xi = y/2$ to get

$$\log\Omega(y) = 2\log\Omega(y/2), \tag{16.25}$$

and by extending this to first binary fractions:

$$\log\Omega(y) = 2^n\log\Omega(2^{-n}y), \tag{16.26}$$

and then by continuity, to all reals:

$$\log\Omega(y) = \frac{1}{\epsilon}\log\Omega(\epsilon y). \tag{16.27}$$

and we obtain finally

$$\Omega(y) = e^{(\log\Omega(x))/\epsilon}. \tag{16.28}$$

This shows that as $\epsilon \to 0$, the density of states scales as an exponential with a $1/\epsilon$ in the exponent.

16.3.2. The Partition Function

The partition function $Z(X)$ corresponding to the density of states $\Omega(x)$, where $X \in O^*$ is in the dual space (i.e. the space of linear functions) to $x \in O$, is given by the multiple Laplace transform:

$$\begin{aligned} Z(X) &= \int_0^\infty \cdots \int_0^\infty e^{-(x,X)} \Omega(x) d^n x \\ &= \frac{1}{\epsilon} \int_0^\infty \cdots \int_0^\infty e^{(\log\Omega(y) - (y,X))/\epsilon} d^n y. \end{aligned} \tag{16.29}$$

Let us now use steepest descents to get the $\epsilon \to 0$ asymptotic behavior. The exponent is a maximum at that value of y where

$$X = \frac{\partial \log\, \Omega(y)}{\partial y}. \tag{16.30}$$

Let us call this point $y_0(X)$. Then asymptotically we have

$$Z(X) = \frac{1}{\epsilon} \frac{\sqrt{2\pi}}{\sqrt{-\partial^2 \log\Omega / \partial y^2|_{y_0(X)}}} e^{(\log\Omega(y_0(X)) - (y_0(X),X))/\epsilon}. \tag{16.31}$$

So we see that the partition function, like the density of states, also scales as an exponential with a $1/\epsilon$ in the exponent asymptotically. Notice that $y_0(X)$ defines a Legendre transformation from y space to X space generated by the function $\log\Omega(y)$ and that the exponents of Ω and Z are the Legendre transforms of each other.

16.4. Maximum Entropy Applied to Statistical Mechanics

Let us now apply the maximum entropy formalism in this framework. Let us denote the underlying phase space of our system by Γ (this will be of the order of 10^{23} dimensional). We have the space P of probability distributions on Γ and a map $o : \Gamma \to O$ which represents the value of the observables of interest in a given microscopic state (O is a linear space of observables discussed above). We may integrate the O valued function o with respect to each probability distribution to get a map

$$mP \to O, \tag{16.32}$$

giving the mean values of the observables for each probability distribution. We also have the information entropy

$$S : P \to \mathbb{R}, \tag{16.33}$$

which is a positive real valued function on P obtained by integrating $-p \log p$ over Γ for each measure $p \in P$. Our goal is to define a map

$$E : O \to P. \tag{16.34}$$

representing the most likely distribution with the given mean values of o. The image of $y \in O$ lies in $m^{-1}(y) \subset P$. We define it to be the maximum of S restricted to this set. We may then pull back S along E to get the entropy as a function on O.

The constrained extremization required is most easily carried out using Lagrange multipliers X which lie in the dual space O^*. S is a maximum on $m^{-1}(y)$ at $p \in P$ if and only if there exists a Lagrange multiplier $X \in O^*$ such that

$$S = X \circ m \tag{16.35}$$

is a maximum on P at p. We may work on the linear space of unnormalized distributions if we introduce a Lagrange multiplier λ to ensure the maximum is normalized. Thus we obtain the requirement

$$\frac{\delta}{\delta p}\left(-\int_\Gamma p(z)\log p(z)dz - \langle X, m(p)\rangle - \lambda\int_\Gamma p(z)dz\right) = 0. \tag{16.36}$$

Inserting the definition of m and carrying out the functional derivative gives

$$\begin{aligned}
&= \frac{\delta}{\delta p}\left(-\int_\Gamma p(z)\log p(z)dz - \langle X, \int_\Gamma o(z)p(z)dz\rangle - \lambda\int_\Gamma p(z)dz\right)\\
&= \frac{\delta}{\delta p}\left(-\int_\Gamma p(z)\log p(z)dz - \int_\Gamma p(z)\langle X, o(z)\rangle dz - \lambda\int_\Gamma p(z)dz\right)\\
&= \frac{\delta}{\delta p}\int_\Gamma p(z)\ \left(-\log p(z) - \langle X, o(z)\rangle dz - \lambda\right)dz\\
&= -\log p(z) - \langle X, o(z)\rangle dz - 1 - \lambda.
\end{aligned} \tag{16.37}$$

Let us call

$$e^{-1-\lambda} \equiv \frac{1}{Z}. \tag{16.38}$$

We must choose $\lambda(X)$ and therefore $Z(X)$ to ensure that p is a properly normalized probability distribution. Solving for $p(z)$, we find

$$p(z) = \frac{1}{Z(X)}e^{-\langle X, o(z)\rangle}. \tag{16.39}$$

The normalization condition shows us that

$$Z(X) = \int_\Gamma e^{-\langle X, o(z)\rangle}dz \tag{16.40}$$

is the partition function.

Given $y \subset O$, we solve for $X \in O^*$ by requiring that the corresponding distribution give y as its mean value of o. Looking at the expression for Z, we see that this is equivalent to requiring that

$$y = -d(\log Z)|_X. \tag{16.41}$$

Here we are identifying $O \approx O^{**}$. This map $L : O \to O^*$ is the Legendre transform generated by $\log Z$. We may pull back the entropy on P by E to give a function on O. We see that this is

$$S \circ E(y) = \log(Z \circ L(y)) + \langle X, y \rangle. \tag{16.42}$$

And so the surface

$$X(y) = L(y) = dS(y) \tag{16.43}$$

is a Lagrangian submanifold in $O \times O^*$.

Jaynes has given a nice demonstration of the second law of thermodynamics using maximum entropy [Jaynes, 1983]. We need only assume that the measured entropy for a given set of thermodynamic parameters is the entropy of the maximum entropy distribution with mean values given by the measurements (we have just seen that this is equivalent to the Gibbs distribution giving the correct value—the basic assumption in traditional statistical mechanics). We will show that if we start with a canonical distribution corresponding to one set of thermodynamic parameters and push it forward by any canonical transformation of the underlying phase space, then the values of the thermodynamic parameters obtained from the pushed forward distribution correspond to an entropy which is larger than that of the first set. We first recall that the information entropy of the pushed forward distribution is the same as the entropy corresponding to the initial parameters. This is because the integral $\int -p \log p$ doesn't change under volume preserving diffeomorphisms and canonical transformations preserve volume. Next, the entropy corresponding to the new parameters is the information entropy of the maximal entropy distribution with them as mean values (i.e. the Gibbs canonical distribution). Since

this entropy is maximal and the pushed forward distribution is another distribution with the new parameters as mean values, the new entropy is greater than or equal to the information entropy of the pushed forward distribution. But this shows that the new entropy is greater than or equal to the old entropy. Since information entropy measures our ignorance, this interpretation of the second law simply says that if we begin with a known (canonical) distribution, follow it in detail under a canonical transformation, and then forget everything but the mean values of some thermodynamic parameters, we are bound to lose information (or at least not gain it).

16.5. Some Symplectic and Contact Geometry

"As each skylark must display its comb, so every branch of mathematics must finally display symplectisation." p. 74 of Arnold, 1984

In this section we will collect together some of the definitions and results of symplectic and contact geometry and give some motivation for their use in the contexts we have in mind. We have seen that in wave theory we get asymptotic interals over

$$e^{iS/\epsilon} \tag{16.44}$$

where S is the action and that in statistical mechanics we get asymptotic integrals over

$$e^{S/\epsilon} \tag{16.45}$$

where S is the entropy. By using stationary phase or steepest descents, we asymptotically reduce these expressions to ones involving only regions with specified differential dS. When we are studying families of values parameterized by y (eg. the point in space we are observing our wave or the thermodynamic observables), we often obtain S as a function of y and are interested in points where dS has a value equal to a Lagrange multiplier in the dual space of y. The level sets of S also often have physical interest (eg. the wavefront or the isentropic states). Thus we are motivated to study the geometric structures associated with the differentials and level sets of functions.

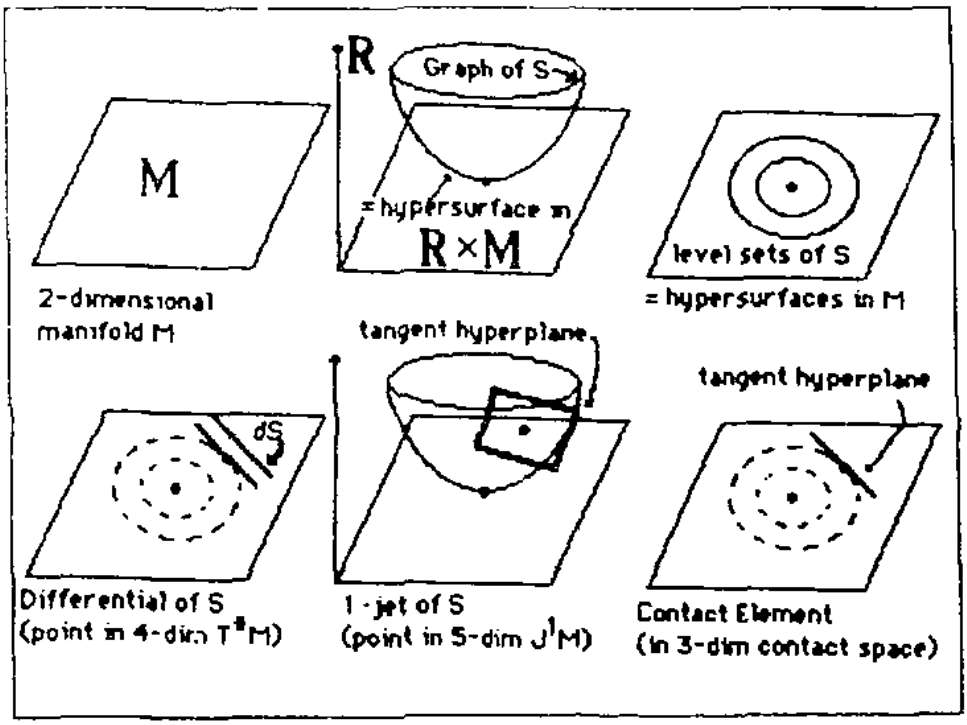

Figure 16.1: Spaces associated with a function on a manifold.

16.5.1. Hypersurfaces Determined by a Function

Every function S on a manifold M of dimension m, determines two natural kinds of *hypersurface* (a hypersurface is a codimension-1 submanifold (i.e. of one dimension less than the ambient space it lies in)). For example, consider the function $S(x, y) = x^2 + y^2$ defined on the 2-dimensional plane coordinatized by x and y. Its *level sets* form a family of hypersurfaces of M parametrized by S (with occasional non-submanifolds that are of measure zero generically, by Sard's theorem). For $S = x^2 + y^2$, the level sets are the circles $x^2 + y^2 =$constant. On the other hand, S's *graph* is a hypersurface in $M \times \Re$. For $S = x^2 + y^2$, the graph is a paraboloid of revolution in (x, y, S) space. The *differential* dS of S is a one-form on M (which

geometrically represents the gradient of S). For $S = x^2 + y^2$, we see that $dS = 2x\,dx + 2y\,dy$. This gives the first order behavior of S near each point of M. The first order behavior of a hypersurface at a point in a manifold is represented by a *hyperplane* in the tangent space of the manifold at that point (i.e. a codimension one subspace of the tangent space). We may thus form the set of all hyperplanes in TM that are tangent to level sets of S and the set of all hyperplanes in $T(M \times \Re)$ that are tangent to the graph of S. The tangent hyperplanes to the level sets are exactly those vectors which dS annihilates. Thus this set of hyperplanes contains all the information that dS does except its length. For $S = x^2 + y^2$, the vectors which are scalar multiples of

$$y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y} \tag{16.46}$$

are annihilated by dS. At each point, this vector spans the tangent space to S's level set. The tangent hyperplanes to the graph of S give all the information of dS, but in addition, the place they are based at tells us the value of S (which isn't known from just dS). The tangent hyperplane to the graph of $S = x^2 + y^2$ at the point (x, y, S), assumed to be away from $x = 0$, $y = 0$, is spanned by the vectors

$$y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y} \tag{16.47}$$

and

$$\frac{1}{2x}\frac{\partial}{\partial x} + \frac{1}{2y}\frac{\partial}{\partial y} + 2\frac{\partial}{\partial S}. \tag{16.48}$$

16.5.1.1. The Underlying Manifold M

We thus get several natural additional spaces of interest when we begin to consider functions on a manifold. Let the coordinates q represent points in M. The space of all q's is M and is of dimension m.

16.5.1.2. The Graph of a Function

The space whose points give both q and the value of a function S at q is $M \times \Re$. This is $(m+1)$-dimensional and is where the graph of S lives.

16.5.1.3. The Cotangent Bundle

The space whose points represent q and the differential $p = dS$ of a function at q, is the *cotangent bundle* T^*M. This is $2m$ dimensional and as we have seen earlier has the *canonical one-form* $p\ dq$ and a natural *symplectic structure* $dq \wedge dp$ defined on it. We defined an m dimensional submanifold of T^*M to be *Lagrangian* if the symplectic form vanishes on it. We have seen in section 7.1.4 that the graph of dS is a Lagrangian submanifold. For waves this represents the local wavevector as a function of position in an eikonal wave (in spacetime the manifold represents the solution to the intial value problem). It is important to represent this in the space of both y's and k's, because even though this surface is smooth, its projection may not be, and our wave can develop multiple branches and caustics. For thermodynamics it gives the intensive variables as a function of the extensive ones (recall the nomenclature convention from 16.3.1). We may think of it as the equation of state.

It is important to think of this as a surface in the higher dimensional space, because its projection onto the intensive variables becomes singular at phase transitions.

16.5.1.4. The First Jet Bundle

The space whose points represent q, the differential $p = dS$ of functions, and the value S of functions, is the *first jet bundle* J^1M. This is $(2m + 1)$-dimensional and we saw earlier that this has a natural contact structure on it, defined as the set of tangent hyperplanes annibilated by the one-form: $dS - p\,dq$. Any function S defines an m dimensional submanifold of J^1M by $q \mapsto (q, dS, S)$. An m dimensional submanifold of a contact manifold is called a *Legendre* submanifold if it is tangent to the contact planes at each point. The submanifold of J^1M determined by a function S is Legendre. For waves this means including the value of the phase with the position and wavevector. When we forget about it (by reduction), we get the Lagrangian submanifold above. For thermodynamics, this gives the relation between the entropy and the intensive and extensive mechanical varibles.

16.5.1.5. The Space of Contact Elements

The set of hyperplanes in a linear space of dimension m forms a smooth manifold of dimension $m - 1$ (eg. the set of lines through the origin in a plane may be thought of as a circle). A hyperplane of the tangent space of a manifold at some point is called a *contact element* at that point. The set of all contact elements of M forms a manifold of dimension $2m - 1$ whose points represent a point q of M and a tangent hyperplane there. We have seen that this manifold of contact elements is

itself a contact manifold. A tangent vector to the space of contact elements, which we may think of as representing an infinitesimal motion of the contact element it is based at, is in the contact structure if the velocity of the basepoint of the contact element lies in the contact element. The set of contact elements which are tangent to a level set of S forms a submanifold of the space of contact elements of M that is of dimension m and is in fact a Legendre submanifold. We may think of this as parameterizing the wavefronts, including the direction. When a wavefront begins to cross itself (as at a caustic), it is important to keep the direction of the wavefront as well as the position. The Legendre submanifold is always smooth, even though the wavefront may develop cusps and self intersections.

Similarly, the set of contact elements of $M \times \Re$ is a contact manifold of dimension $2m + 1$. The set of contact elements that are tangent to the graph of S forms an m dimensional Legendre submanifold. This set will be important to our understanding of the Legendre transform.

16.5.2. The Conormal Bundle

If we are given a codimension n submanifold N of M (that is thus of dimension $m - n$), we may think of it as the simultaneous level set of n linearly independent functions (locally). This motivates us to consider the set of all covectors in T^*M based on N, which annihilate the tangent space to N. This is called the *conormal bundle* of N in M. (If M had a metric, then this would be all vectors that are perpendicular to N). This is a Lagrangian submanifold of T^*M. In the limiting case where N is a point of M, the conormal bundle is just the set of covectors based

at that point. If N is the whole of M, then it is the zero section of T^*M. If M were Riemannian, then there is a natural projection of the (co)normal bundle of N into M given by sending (q,p) in the normal bundle to the point a distance $|p|$ along the geodesic in M starting at q in the direction p. This gives the set of rays in M that are traversed by light emitted by N, where the metric represents the (anisotropic) index of refraction. The singularities of the projection from the normal bundle to M represent the caustics. They are the points which lie at the center of curvature of some direction on the surface ([Arnold, 1983] p. 83).

16.5.3. The Wavefront Set

This map is also related to the wave front set of a distribution d on M introduced by Hörmander ([Hörmander, 1983] p. 252). We associate with d a Lagrangian submanifold of T^*M by saying that a covector p is in d's wavefront set if the pushforward of d to $\Re^1$ along any smooth function whose differential is p is still singular (i.e. there exists a smooth function on $\Re^1$ whose integral with respect to the pushforward of d doesn't approach zero as the region of integration vanishes). Thus a point δ-function at q on M has a wavefront set that includes all covectors at q while a δ-function supported on a submanifold N has a wavefront set that includes only the conormal bundle of N. This is of interest because the singularities of the solution of a hyperbolic P.D.E. with singular intial conditions must lie on the projection to M given above of the wavefront set. Thus for the wave equation on a Riemmanian manifold, a δ-function intial condition will lead to singularities on a growing sphere (with respect to the metric) which we recongize as slices of

the light cone. A singular hypersurface moves like a wavefront and propagates only in the direction of the rays. We can understand this close relation between high frequency asymptotics and the evolution of singular distributions by recognizing that the singular aspects are due to the infinitely high frequencies, and a singular distribution can be represented as an integral over the asymptotic parameter of a family of eikonal waves. This can be related to Huygens principle. The fact that singularities move on rays is behind a beautiful discussion on p. xi of [Guillemin and Sternberg, 1977] explaining why the frequency of a bowed violin is the same as that of a strummed one (a priori, the frequencies of driven oscillations should have nothing to do with free oscillations). The explanation is that when the string snaps away from the bow, a singular kink is generated which goes down the string and back to kick the string off the bow again, generating a frequency equal to that of the normal mode corresponding to that periodic ray. The reference gives figures showing the string motion.

16.5.4. The Space of Tangent Contact Elements

We may also consider the set of all contact elements of M which are tangent to N (i.e. which contain N's tangent space). This set is a Legendre submanifold of the space of contact elements of M. This represents the local pieces of the wavefront that will be emitted from N. Even if N is lower dimensional, like a point, the emitted wavefront will be n dimensional (like a sphere about the point).

16.5.5. Legendre Transforms and Linear State Spaces

As we have discussed in section 7.1.1, the scaling we perform to do our asymptotics stretches the underlying manifold M in both the wave case (where we went to a slow space) and in the statistical mechanical case (where we went to rescaled mechanical variables). Asymptotically, any non-trivial manifold structure disappears and we are left with $\Re^m$. In this case the cotangent bundle becomes $\Re^m \times \Re^{m*}$. When the base space is linear, there are more geometrical operations which we may perform. The new freedom is to project not only "vertically" to $\Re^m$, but also "horizontally" to $\Re^{m*}$. Essentially we have decided how to identify all the cotangent spaces at different points of M. We may do this by choosing coordinates on M, which gives such an identification but depends on coordinate choice. As $\epsilon \to 0$, however, all smooth coordinate systems lead to the same asymptotic identification. This asymptotic identification of cotangent spaces is non-uniform in q, but all our operations, like local Fourier transform, always include a window which scales so as to eliminate the non-uniform parts.

16.5.5.1. The Legendre Map

Given any function S on $\Re^m$, its differential takes its values in $\Re^{m*}$. Thus dS is a map from $\Re^m$ to its dual space, which we may call the *Legendre map*. In the case of waves, this maps y space into k space. For thermodynamics, it takes a set of extensive variables into their thermodynamically conjugate intensive variables. We have seen that these are the stationary points for the Fourier and Laplace transforms respectively.

16.5.5.2. The Legendre Transform

When S is strictly convex (its Hessian (second derivative matrix) is positive definite) then this map is one to one. If S is bounded below by some quadratic form, then it is a diffeomorphism. In this case it makes sense to ask for the function $T(p)$ on $\Re^{m*}$ whose corresponding Legendre map is the inverse of the one generated by $S(q)$. One sees that in this case:

$$T(p) \equiv \langle p, q(p) \rangle - S(q(p)), \tag{16.49}$$

where $q(p)$ is the inverse of the first Legendre map, generates the inverse.

Let us show this explicitly in the coordinates q^i where $1 \leq i \leq N$. Let us use L_S to denote the Legendre map $L_S(q) = dS(q) = p$ defined by S. In coordinates this reads

$$p_i = (L_S(q))_i = \frac{\partial S}{\partial q^i}. \tag{16.50}$$

By the condition imposed on S, this map is invertible. We denote the inverse by L_S^{-1} and the function T we defined is then given by

$$T(p) = (L_S^{-1}(p))^i p_i - S(L_S^{-1}(p)). \tag{16.51}$$

We want to show that the Legendre map L_T defined by the function T is actually the inverse of L_S. This Legendre map is expressed in coordinates as

$$\begin{aligned}(L_T(p))^i = &\frac{\partial T}{\partial p_i} \\ &- (L_S^{-1}(p))^i + p_j \frac{\partial (L_S^{-1}(p))^j}{\partial p_i} - \frac{\partial S}{\partial q^j}(L_S^{-1}(p)) \cdot \left(\frac{\partial}{\partial p^i}(L_S^{-1}(p))^j \right).\end{aligned} \tag{16.52}$$

We have used the Leibniz rule to do the derivative of the first term in T and the chain rule to do the second term. Now recognize that

$$\frac{\partial S}{\partial q^j}(L_S^{-1}(p)) = p_j \tag{16.53}$$

to see that the last two terms cancel. We are finally left with

$$L_T = L_S^{-1} \tag{16.54}$$

as desired.

In general we may define:

$$T(p) \equiv \sup_q[\langle p, q\rangle - S(q)] \tag{16.55}$$

to be the Legendre transform of S ([Arnold, 1983] p 19). The previous definition agrees with this one in the situations to which it applies. If S is (strictly) convex then so is T. We shall see that this is important for thermodynamics, since $-S$ must be a convex function of the extensive variables.

16.5.5.3. The Legendre Transform and a Function's Graph

If we are given a number T and a vector $p \in \Re^{m*}$, then the equation

$$T = \langle p, q\rangle - S \tag{16.56}$$

defines a hyperplane in (q, S) space (i.e. $\Re^m \times \Re$). The equation above says that the graph of S in $\Re^m \times \Re$ hits this hyperplane at the point where it has the slope p. The value of T is minus the S intercept of this hyperplane (i.e. the point where

it hits the axis $q = 0$) as is shown in figure (16.2a). We parameterize the space of non-vertical (i.e. they don't contain lines parallel to the $q = 0$ axis) hyperplanes in (q, S) space by (p, T) as above. These are called Plücker coordinates (see for example p. 88 of [Jenner, 1963]). The map that sends points of the graph of S in (q, S) space to the hyperplane tangent to the graph there, goes to $p = \partial S/\partial q$ which is image of q under the Legendre map and T which is the value of the Legendre transform of S at p.

16.5.5.4. Legendre Transforms and Projective Duality

The map which sends points of a hypersurface to the hyperplane tangent to the surface there has been the object of mathematical study for a long time. It is behind the notion of projective duality where, for example, all theorems of geometry in the (projective) plane may have the words "point" and "line" exchanged (eg. two points determine a line, two lines determine a point). To make this work out, one must tack on "directions at infinity" to R^m so that parallel lines really intersect at infinity. This leads to projective geometry, where the m dimensional projective space RP^m is defined as the space of lines through the origin of $\Re^{m+1}$. A line through the origin of $\Re^{m+1*}$ defines a linear form on $\Re^{m+1}$ up to magnitude, which may be identified with the hyperplane through the origin of $\Re^{m+1}$ on which it vanishes. This in turn is made up of lines through the origin, and may be thought of as an arbitrary hyperplane in P^m. Therefore we call the space of hyperplanes in P^m its projective dual space P^{m*}.

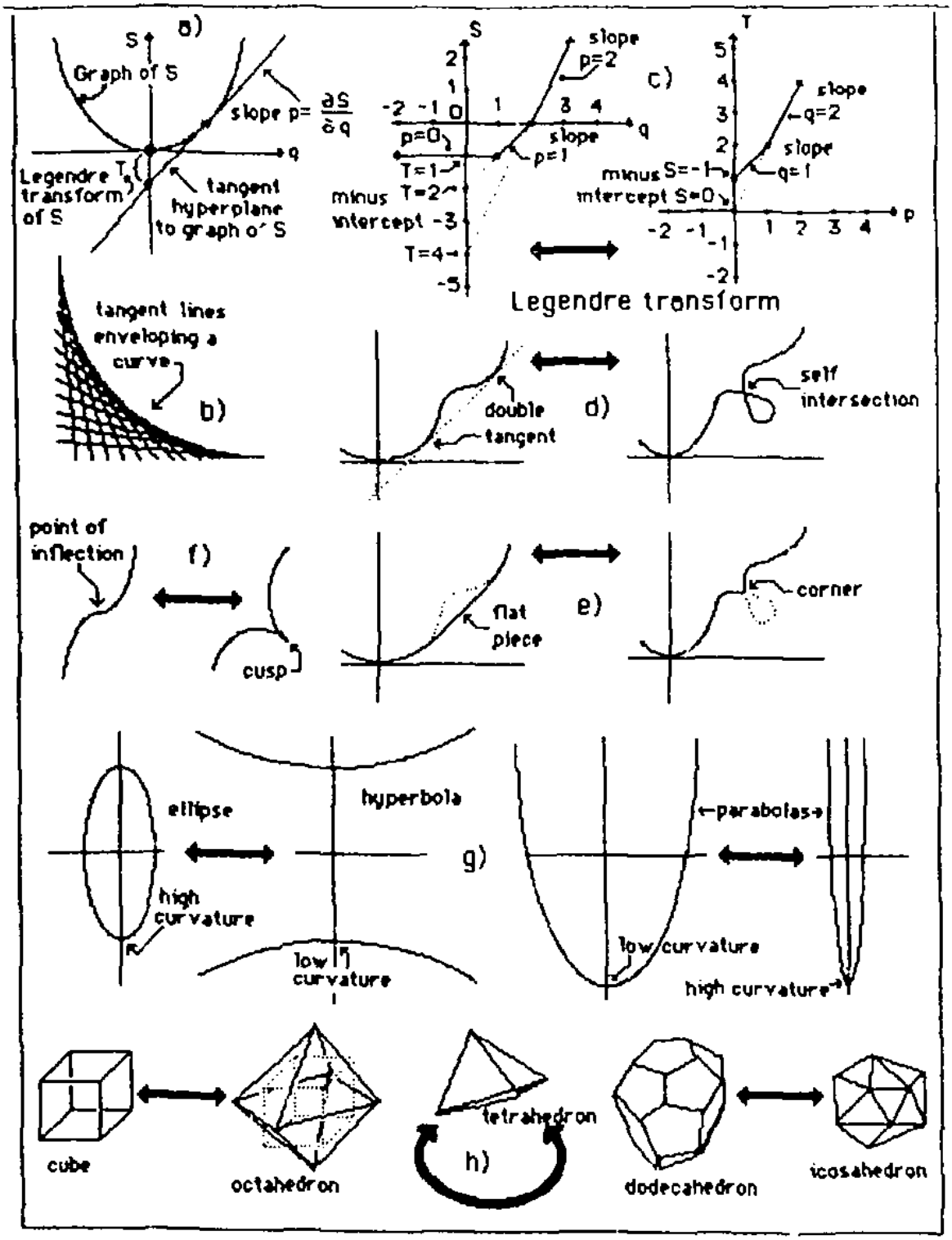

Figure 16.2: Various aspects of Legendre transforms and projective duality: explanations are given in the text.

To each hypersurface in P^m, we may associate the projectively dual hypersurface in P^{m*} defined as all hyperplanes that are tangent to the first hypersurface. The *original hypersurface is the envelope of the planes defined by its dual* (e.g., the tangent lines to a curve as shown in figure (16.2b) themselves form a curve in the space of lines). This relationship is involutive in the sense that the dual of the dual brings you back to the original. The graph of the Legendre transform of a function is the dual of the graph of the function in this sense (p. 20 of [Arnold, 1983]).

Flat places in the surface, where it includes straight line segments and so there is an interval of different points with the same slope, correspond to corners in the dual surface, which has an interval of different slopes at the same point. For example, in figure (16.2c) the graph of S is made of 3 straight segments joined at two corners. The graph of T, its Legendre transform, is made of 2 straight segments (corresponding to the corners in the graph of S) and 3 corners (corresponding to the segments of S). The entire graph of S for $q \leq 1$ has slope 0 and S intercept -1 and therefore corresponds to the single point $p = 0$ and $T = 1$. As we follow S's graph around the corner at $q = 1, S = -1$, the slope goes from 0 to 1 and the intercept from -1 to -2. This single point therefore corresponds to the whole line segment over $0 \leq p \leq 1$ in the graph of T. The line segment corresponding to $1 \leq q \leq 2$ *in* S*'s graph again has a single slope and intercept and corresponds to* the point $p = 1, T = 2$. The corner at $q = 2, S = 0$ gives rise to the line segment over $1 \leq p \leq 2$. Finally the entire line over $2 \leq q$ corresponds to the single point $p = 2, T = 4$. This entire analysis may be applied in reverse to go from $T(p)$ to $S(q)$ showing that Legendre transforms are involutive. For example, the line segment in

T's graph over $1 \leq p < 2$ has slope 2 and intercept 0 and so corresponds to the single point $q = 2, S = 0$ in the graph of S. This is important in thermodynamics where flat places in the graph of the entropy as a function of the extensive variables correspond to phase transitions.

Double tangents (i.e. when a hyperplane is tangent at two points of the surface) correspond to points of self intersection of the dual surface (an example is shown in figure (16.2d)). In thermodynamics, we take the convex hull of the region below the graph of entropy, and so places with double tangents get turned into flat regions as shown in figure (16.2e). The dual surface replaces the intersection of two surfaces by their coming together at a corner and stopping.

Surfaces defined by algebraic equations have duals defined by algebraic equations. In the 2-dimensional plane, a curve with an inflection point (i.e. flat to the second order) has as its dual a cusp (whose edges are tangent to the second order) as shown in figure (16.2f). A conic section in the plane gets taken to a conic section. As shown in figure (16.2g), ellipses go to hyperbolas and parabolas go to parabolas. The duals of polyhedra in 3 dimensions have vertices corresponding to the original faces and faces corresponding to the original vertices (eg. a cube and an octahedron, an icosahedron and a dodecahedron, and a tetrahedron and itself are dual as shown in figure (16.2h)). The graphs of q^a/a and p^b/b are dual when $1/a + 1/b = 1$ and so they are Legendre transforms of each other. A norm $f(q)$ on a linear space may be defined by the unit sphere it defines. There is a natural norm on the dual space given by $g(p) = \max_{f(x)\leq 1}|(p, x)|$. Its unit sphere is the dual of the original one (this exemplifies the relationship between hypersufaces defined by

level sets and by graphs of functions).

16.5.5.5. Legendre Transforms and Uncertainty Relations

Since q and p are dual variables, acting on one by an invertible linear transformation A is equivalent to acting on the other by the inverse A^{-1} of that transformation. This means that the Legendre transform of a function $S(A \cdot q)$ is equal to $T(A^{-1} \cdot p)$ if $T(p)$ is the Legendre transform of $S(q)$. This is the asymptotic formulation of uncertainty principles for the Fourier and Laplace transforms. If $S(q)$ is a quadratic form, then its Legendre transform is also a quadratic form. In fact these are the unique functions for which the value of the Legendre transform is equal to that of the function at the corresponding point. The widths of the forms (and so the volume of the unit spheres they define) are inverses of each other. Since the exponential of a quadratic form is a Gaussian, this says that in thermodynamics, when the probability distribution of an extensive variable is Gaussian (as is commonly the case in fluctuation theory), then the asymptotic distribution over the conjugate intensive variable is also Gaussian with the inverse dispersion. The more precisely you know the temperature, the less precisely you know the energy and vice versa.

16.5.5.6. Legendre Transforms and Jets of Functions

Let us give a final picture of a Legendre transformation (p 366 of Arnold, 1978.) Let us call a map from one contact manifold to another of the same dimension that takes contact planes to contact planes, a *contact transformation*. If we consider the first jet space of M, then the map

$$(q, p, S) \mapsto (p, q, \langle p, q \rangle - S) \tag{16.57}$$

is a contact transformation which takes the graph of dS and S into the graph of its Legendre transform.

16.6. The Origin of the Lagrangian Submanifolds in Physics

In this section we will state the theorem from symplectic geometry which may be viewed as being responsible for the Lagrangian submanifolds in both wave theory and thermodynamics. In both these cases we have reduced the quantities of interest to integrals of asymptotic exponentials over large spaces which we then reduce to a variational principle for the exponent by stationary phase or steepest descents. In the wave case, we obtain the wave at a given point as an integral over all paths of an exponential with the action S of a path in the exponent (see50 [Schulman, 1981]), leading to the principle of least action (or actually stationary action). In the statistical case we obtain the probability of a given set of measurable quantities as an integral over all distributions of an exponential with the entropy S of a distribution in the exponent, leading to the principle of most entropy (maximum entropy).

16.6.1. Constrained Integration and Extremization

In both cases we have an integral over some space, typically defined by some constraints (the end of the path is at the observation point, or the distributions have given mean values for the quantities of interest). If we project this space to a smaller one, we may first integrate over the fibers of the projection, and then over the smaller space. This leads to a variational principle where we first extremize S over the fibers, giving a function S on the smaller space whose extrema represent the contributions we are interested in.

We discuss the physical examples in the next few sections. To see what is going on geometrically, consider the projection from $\mathbb{R}^2$ to $\mathbb{R}$ taking $(x, y) \mapsto x$. If we want

the extremal value of

$$S(x,y) = 15 + (x-1)^2 + (y-2)^2. \tag{16.58}$$

we may first extremize $S(x,y)$ at each x, holding x fixed and letting y vary. The stationary points under this constrained variation satisfy

$$\frac{\partial}{\partial y}S(x,y) = 0 = 2(y-2). \tag{16.59}$$

The surface $y = 2$ is made up of the constrained critical points. S restricted to this surface is

$$S_c(x) = 15 + (x-1)^2, \tag{16.60}$$

which we may think of as living on the projected space. We now extremize over x yielding

$$\frac{d}{dx}S_c(x) = 0 = 2(x-1). \tag{16.61}$$

Thus $x_c = 1$ and the critical value of S is $S_c(1) = 15$.

16.6.2. Paths Constrained on Surfaces

For example as in figure (16.3), we may first sum over paths which go through given points on the surfaces: $P_1, \ldots P_s$ in space before reaching the point of observation. We extremize over paths subject to these constraints and so obtain the action as a function on

$$P_1 \times \cdots \times P_s. \tag{16.62}$$

We may now find the critical points on this space to find the actual paths taken. This finite dimensional integral and variational principle make good mathematical sense. We actually define the "integral over all paths" in terms of finer and finer approximations by such piecewise paths as the asymptotic parameter vanishes. Physically, the rays don't mean anything on a scale smaller than a wavelength, and as we do our scaling, the pieces of path we sum over should get smaller and smaller while including more and more wavelengths.

As we let the number of constraint surfaces on which we specify the point of intersection with a path increase to infinity, we more and more precisely constrain the ray. One can imagine this limiting to the case where giving a point in the surface product space uniquely specifies a path. This is the sense in which the path space can be thought of as an infinite product of interposed surfaces (that foliate space).

16.6.3. The Wavevector as a Kind of Force

We see in this example that the true paths will be those which come into and leave a surface with the same slope, hinting that the dual space of a surface is important. For an extremal ray, when we perturb the point on the surface, the change of S on the incoming part exactly cancels the change of S on the outgoing part to first order. Thus the derivative of the action of a part of the ray with respect to changes in its endpoint acts as a kind of "force". For a valid ray the "forces" on the incoming ray and outgoing ray must balance. We will see the analogy with thermodynamic forces momentarily.

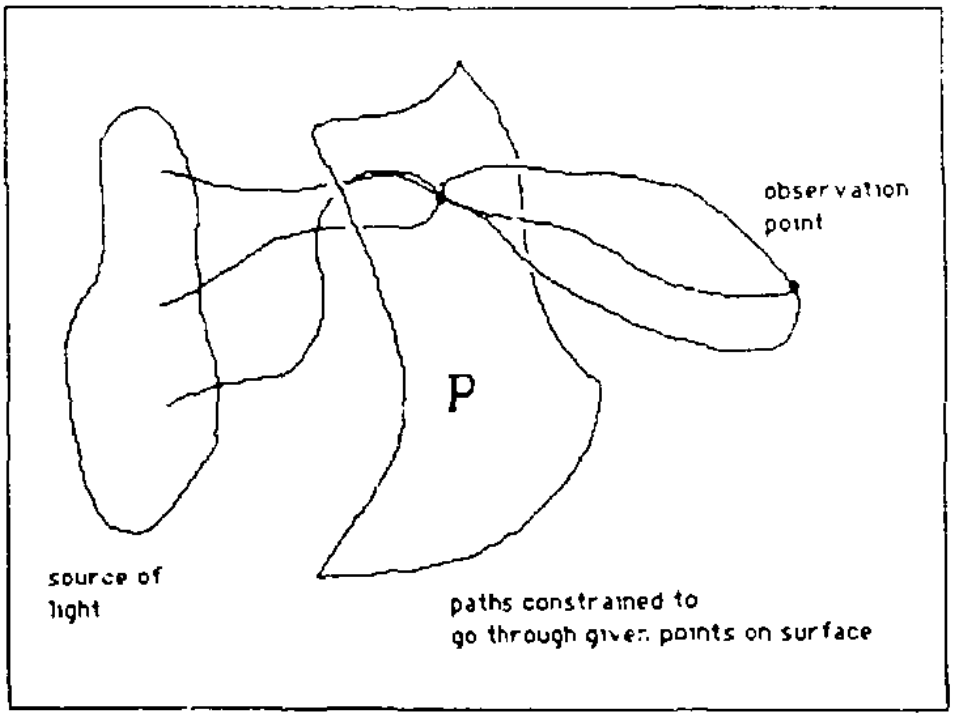

Figure 16.3: Some paths that go through a given point on the surface P.

16.6.4. Distributions Constrained on Subsystems

In the statistical case, we may imagine different pieces of our thermodynamic system to be forced to have given values for their extensive quantities. For example as in figure (16.4), we might have a box with a movable partition which allows the transfer of volume between its two halves and is thermally conducting and so also allows the transfer of energy. We may first do our integral over distributions with a given energy and volume in the left portion. We maximize the entropy subject to this constraint and so obtain an entropy on the finite dimensional space of values of the left side's energy and volume. We extremize this entropy on a finite dimensional space to find the actual equilibrium values of the constrained quantities.

This finite dimensional integral and variational principle makes good mathematical sense. We actually define the "integral over all distributions" in terms of finer and finer partitions of our system as the asymptotic parameter vanishes. Physically, the distributions don't mean anything for too few particles, and so as we do our scaling the distributions we sum over should be constrained to give definite values to smaller and smaller parts of the system while including more and more degrees of freedom. This kind of averaging was discussed in the introduction.

As we let the regions of phase space over which the probability distribution averages are specified become smaller and smaller, we more and more precisely constrain a distribution. One can imagine this limiting to the case where giving a point in the region average product space uniquely specifies a probability distribution. This is the sense in which the space of distributions can be thought of as an infinite product of spaces of averages at points of phase space.

16.6.5. Thermodynamic Forces

We see in this example that the true energy and volume of the left system will be those such that the variation of the left portion's entropy is equal and opposite to the variation of the right portion's entropy to first order, hinting that the dual space of the constrained observables is important. The derivative of the entropy of the left half with respect to the constraint acts like a "force" and the left and right forces must be balanced in equilibrium.

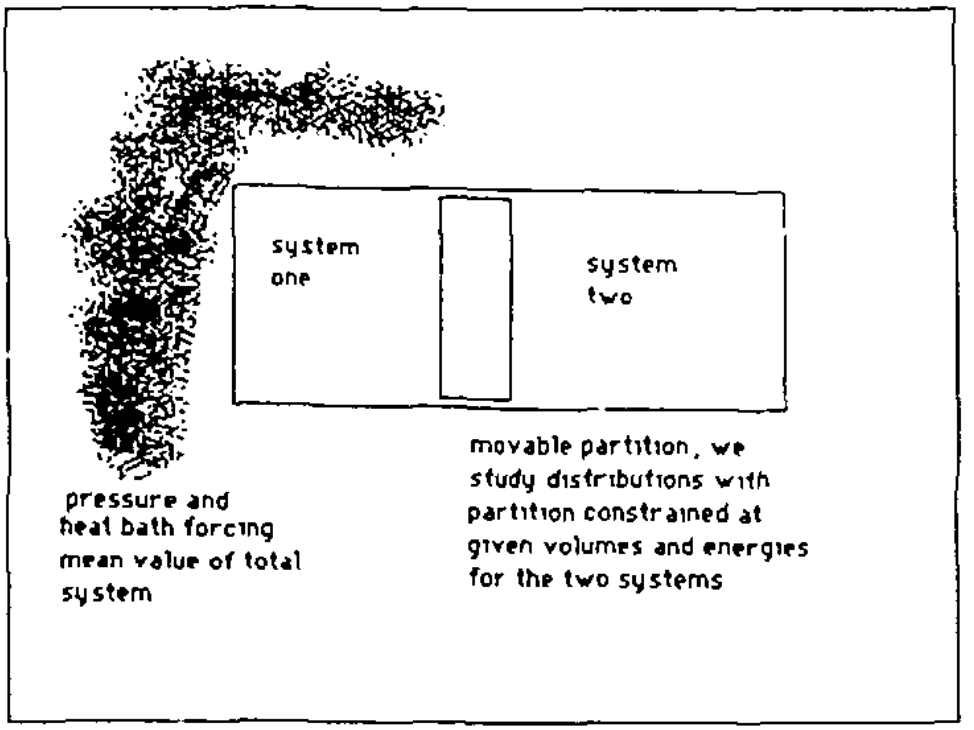

Figure 16.4: A distribution is constrained to give the left region a definite energy and volume.

16.6.6. Lagrange Multipliers and Legendre Maps

This same idea is captured in the notion of Lagrange multiplier, which acts like a system with given "force" instead of given value for any constrained quantity. To maximize a function over a space with an imposed constraint, we may instead maximize over a new system on the whole space with an additional linear piece with given "force" that allows us to a postiori make the critical point satisfy the constraint.

For example, if we want to maximize

$$S(x, y) = 2 - x^2 - y^2 \tag{16.63}$$

with the constraint that $y = 1$, we might consider the constrained variational prob-

lem of maximizing

$$S_c(x) = 2 - x^2 - 1 = 1 - x^2 \tag{16.64}$$

over the constraint surface. This yields $x = 0$ and $S = 1$. Alternatively, we may maximize

$$S_L(x, y) = 2 - x^2 - y^2 - \alpha(y - 1) \tag{16.65}$$

over all x and y yielding $x = 0$ and $y = -\alpha/2$. The proper force α to push the *maximum* to $y = 1$ is $\alpha = -2$. This again yields $x = 0$ and $S = 1$. The reason for doing this is that it is often easier to do the unconstrained variations (even with the free parameter α) than to impose the constraint explicitly. We have seen that the relation between the states and the conjugate forces is just the Legendre map generated by S. There is a corresponding function on the dual variables which is the Legendre transform of S. If we think back to our asymptotic integrals, imposing as given $\partial S/\partial y$ in the wave case makes the integral into a Fourier transform with specified wavevector. In the statistical case we get a Laplace transform with specified intensive variables.

16.6.6.1. Constant Force Asymptotic Systems

These systems with given "forces" may often be thought of as asymptotic limits of real systems. Thus for example, a very extended weak spring acts like a constant force (a very strong inextended spring acts like a constant position (wall)), a large system in thermal equilibrium acts like a heat bath with constant temperature and infinite heat capacity, (a small system acts like a thermal insulator, it has zero heat capacity and anything coupled to it has almost constant energy), a large battery

with a large series resistor acts like a constant current source, a large constant current source with a small parallel resistor acts like a constant voltage source, etc.

16.6.7. Lagrangian Submanifolds and Constrained Extremization

Let us now give a theorem of symplectic geometry that describes this kind of situation geometrically. This is given in (Weinstein, 1977) on page 25 and in (Guillemin and Sternberg, 1977) on page 149. Above we have seen that we often want to project a space onto a smaller one while considering the critical points of a function S. The following theorem (16.7) tells us that if we have a projection of M onto N, then those points in T^*N which pullback to points of the graph of dS in T^*M actually form a Lagrangian submanifold of T^*N. These points sit over the critical points of S restricted to each fiber of the projection (i.e. inverse image of a point in N).

16.6.7.1. Parametrizing Lagrangian Submanifolds

This theorem is particularly interesting when there is more than one critical point of S on the fiber over $x \in N$. This means that the corresponding Lagrangian submanifold in T^*N has more than one sheet sitting over x. Thus we may obtain "folded" over Lagrangian submanifolds from perfectly nice ones (i.e. the graph of dS in T^*M). Weinstein shows that this may always be done locally and gives conditions for the global version. This is the key to Maslov's approach to wave

asymptotics. In doing an eikonal study of a linear wave equation, we may represent a wave as a so-called oscillatory integral.

$$u_\epsilon(y) = \int A(y,\alpha)e^{i\theta(y,\alpha)/\epsilon}\,d\alpha. \tag{16.66}$$

If we view our equation as defined on (y,α) space by ignoring α, solutions on the (y,α) space project to solutions on y space (by linearity). We have just seen that even when the Lagrangian submanifold in the cotangent bundle of y space becomes folded over and the asymprotics becomes invalid (i.e. nonuniform at the fold), there is a nice wave on (y,α) space that projects to it. We may do our asymptotics there and project the answer via stationary phase to see that even the folded over Lagrangian submanifold is a representative of the wave.

16.6.7.2. Theorem on Pushing Forward Lagrangian Submanifolds

Let us give the statement of the theorem from [Guillemin and Sternberg, 1977]:

Theorem 16.6. *Let $f : M \to N$ be a smooth map with df of constant rank and let Λ be a Lagrangian submanifold of T^*M. If Λ intersects df^*T^*N transversally, then $df_*\Lambda$ is a Lagrangian submanifold of T^*N.*

16.6.7.3. Application of the Theorem to Waves

This restricts to the case above when A is the graph of the function S. In the case of waves, we take M to be the space of all paths, N to be the 3-dimensional space in which we observe our waves, the projection to be that which sends a path to its endpoint, and the action S of a path to be the function to extremize. The the theorem says that the differentials of the actions of those paths with extremal action for each endpoint form a Lagrangian submanifold in the cotangent space of obervation space N. The fibers of the cotangent bundle are the derivative of action with respect to the observation point and represent the wavevectors at a given point. This cotangent bundle is the wave phase space and the Lagrangian submanifold is the graph of the wavevector at each point for an eikonal wave with wave phase S.

16.6.7.4. Application of the Theorem to Thermodynamics

In the case of statistical mechanics, we take M to be the space of probability distributions, N to be the space of extensive observables that we are studying, the projection to be that which sends a distribution to the mean value of the observables in that distribution, and the entropy S of a distribution to be the function to maximize. The theorem then says that the differentials of the entropies of those distributions with maximal entropy for each mean value of the observables form a Lagrangian submanifold in the cotangent space of the extensive variables. The fibers of the cotangent bundle are the derivative of entropy with respect to the extensive variable and represent the conjugate intensive variables. This cotangent

bundle is the thermodynamic phase space and the Lagrangian submanifold is the graph of the equation of state for an equilibrium system with entropy S.

16.7. Theorem on the Pushforward of Legendre Submanifolds

Let us now extend this theorem to Legendre submanifolds of contact manifolds, because this is a context in which we may understand more of the structure of thermodynamics. Assuming the same transversality condition as in the last theorem (which is generically true), we find that for a projection $M \to N$ and a function S on M, the points in the first jet bundle J^1N of N which pull back to points in J^1M in the graph of S and dS where the derivative of S along the fibers of the projection is zero, together form a Legendre submanifold of J^1N. Let us locally use coordinates (q, a, S, p, α) on M where a parameterizes the fibers of the projection and q are coordinates on N, α and p are the corresponding differentials, and S represents the value of a function. We assume that the coordinates (q, S, p) agree with those of J^1N on the set of pulled back vectors.

The canonical contact structure on J^1M is given by the vectors annihilated by the form

$$dS - p\,dq - \alpha da. \tag{16.67}$$

A contact form on J^1N is given by

$$dS - p\,dq. \tag{16.68}$$

We have seen earlier that the one-jets of S in J^1M form a Legendre submanifold with respect to this contact structure. We are interested in its intersection with the set $\alpha = 0$ (i.e. those points where S's derivative vanishes along the fibers of the projection). From the expression for the contact form, we see that at these points $dS - p\,dq$ vanishes on S's one-jets and so the projected submanifold is contact on N.

We may use the theorem above to see that the projection has the same dimension as N (adding the S direction to both T^*M and T^*N doesn't do anything to the dimension). Thus we may conclude that the projection is a Legendre submanifold of J^1N.

16.7.1. The Contact Structure for Thermodynamics

Let us apply this result to the thermodynamic situation. Here the manifold N is made of the asymptotically scaled extensive thermodynamic variables we are considering. We have been using y for coordinates on this and they represent such quantities as the energy U, the volume V, the numbers of the various species of particles or molecules (not including the total as discussed above) $N_1 \ldots N_n$, the magnetic moment μ, the electric dipole moment Π etc. On the first jet space J^1N, the function variable is the entropy S, and the derivative directions are coordinatized by the thermodynamically conjugate variables X to the y. The conjugate variable to E is the inverse temperature: β, to V is the pressure over the temperature: p/T, to N_i is minus the i'th chemical potential over the temperature: $-\mu_i/T$, to the magnetic moment is minus the magnetic field strength over the temperature: $-H/T$, to the electric dipole moment is minus the electric field over the temperature: $-E/T$, etc.

16.7.1.1. The Contact Form for Jets of Entropy

The contact form is then

$$dS - \frac{1}{T}(dU + p\,dV - \sum_i \mu_i\,dN_i - H \cdot dM - E \cdot dP + \cdots). \tag{16.69}$$

We have seen that our asymptotic theory guarantees that this form vanishes on the equation of state surface in J^1N since it is a Legendre submanifold. We recognize this as the first law of thermodynamics.

16.7.1.2. The $U, V, (1/T), (p/T)$ Symplectic Manifold

For simplicity, from now on we shall consider only $(S, U, V, (1/T), (p/T))$ space. The other coordinates behave in exactly the same way if they are desired in a theory. We have seen that we may project our contact space along the S direction to obtain the *symplectic manifold coordinatized by* $(U, V, (1/T), (p/T))$. The contact form given above goes into the canonical one form on this space (since it is constant on the fibers):

$$\frac{1}{T}(dU + p\,dV). \tag{16.70}$$

The corresponding symplectic structure is:

$$dU \wedge d(\frac{1}{T}) + dV \wedge d(\frac{p}{T}). \tag{16.71}$$

By our general theory, the equation of state surface is a Lagrangian submanifold with respect to this symplectic structure.

16.7.1.3. The (S, V, p, T) Symplectic Manifold

We have seen earlier that a contact form gives the same contact structure when it is multiplied by any nowhere vanishing function. Let us use that freedom to get an equivalent contact form on our contact space by multiplying by $-T$:

$$dU - T\,dS + p\,dV. \tag{16.72}$$

This too vanishes on our Legendre submanifold. Now the form is constant along the U direction and so we may project it to $(S, V, p, -T)$ space, where it becomes the canonical one-form. The corresponding symplectic structure is exactly the one given by [Kijowski and Tulczyjew, 1979] that we listed in section 12.1. We may obtain this same symplectic structure as the canonical cotangent structure by viewing any of the pairs: (V, S), (V, T), (p, T), or (S, p) as the base and the other two variables as the cotangent fibers. Our Lagrangian submanifold is then represented as the graph of four different functions. As we have seen in great detail these are the Legendre transforms of one another, and are known as the internal energy, the Helmholz free energy, the Gibbs free energy, and the enthalpy.

16.7.2. Legendre Transforms and Thermodynamic Potentials

The reason for introducing these extra generating functions for our surface is that it is they that are extremized under different combinations of constraints. We saw that for given extensive variables the system maximizes its entropy. For adiabatic variation of a system, the entropy is an adiabatic constant of the motion (we have seen that this is exactly the same situation as the adiabatic invariance of

the action of almost periodic orbits and leads there to the pseudoforces of reduction in mechanics which are the analog of the thermodynamic forces we are discussing here and the wave action density conservation for waves). The fact that entropy is maximized when energy and volume are fixed is equivalent to the fact that energy is minimized when entropy and volume are held fixed. An analagous situation is that the shapes in three dimensions which minimize their surface area for given volume, are the same as those which maximize their volume for given surface area. Thus by the same argument we used to show entropy was a concave function of the extensive variables, we see that the energy is a convex function of the other extensive variables and the entropy. As an example of a Legendre transform in both S and V, we see that the Gibbs free energy $U - TS + PV$ is minimized for given temperature and pressure.

16.8. Phase Transitions and the Geometry of the Equation of State

In Gibbs, 1873, the function $S(U,V)$ (or equivalently $U(S,V)$) was called the fundamental equation because it contains all the thermodynamic information about a substance. He was the first to recognize ([Israel, 1979] p. xii) that this function contains more information than the usual "equation of state" which is a relation of the form

$$f(p,V,T) = 0. \tag{16.73}$$

For example, in the case of an ideal gas one needs the relation

$$pV^{\gamma} = \text{constant}, \tag{16.74}$$

in addition to the equation of state

$$pV = NkT \tag{16.75}$$

to specify the behavior of the gas.

I will, nonetheless, call the expression of S as a function of the mechanical variables the equation of state, because it really describes the allowed relations between the intensive and extensive variables for a substance. For this example, the surface describing the possible states is a two dimensional surface in (U,V,T,p) space (or equivalently in (S,V,T,p) space). The usual equation of state only says that this surface lies in a three dimensional one given by $f = 0$ and requires another constraint to obtain complete information.

16.8.1. ... ...tics and Phase Transitions

We have seen that in the eikonal wave situation, places where the Lagrangian submanifold corresponding to an eikonal wave does not project nicely onto y space correspond to caustics of the wave field. These structures correspond to higher order derivatives vanishing at critical parameters in our stationary phase. In the statistical mechanical context, the situation is simpler because only maxima contribute to the state as opposed to arbitrary critical points. In Rene Thom's catastrophe theory such a condition is called the Maxwell condition. The places where the thermodynamic Lagrangian submanifold does not project nicely onto the intensive variables correspond to first order phase transitions.

16.8.2. Convexity and First Order Phase Transitions

For definiteness, let us use the extensive variables U and V to descibe the ideas of this section, though any set y would do as well. [Gibbs, 1873b] considers the form of the entropy as a function of U and V. He showed that S is a *concave* function of these variables. This means that for any $t \in [0,1]$, and $(U_a, V_a), (U_b, V_b)$ in the domain of interest, we have the inequality

$$S((1-t)U_a + tU_b, (1-t)V_a + tV_b) \geq (1-t)S(U_a, V_a) + tS(U_b, V_b) \qquad (16.76)$$

If we think of the graph of S as a two dimensional surface in (U, V, S) space, this just means that the graph of S does not fall below a line segment joining any two points on it. Equivalently, the region below this graph is convex (and so one says $-S$ is a convex function).

Gibbs' argument runs as follows. Assume we had a point (U_c, V_c, S_c) on the graph of S which lay below a line segment joining the two allowable states: (U_a, V_a, S_a) and (U_b, V_b, S_b). As the system wants to maximize its entropy as much as possible, instead of going into a homogeneous phase with U_c, V_c, it will split into two phases, one of U_a, V_a and one with U_b, V_b in such a way to have the total be U_c, V_c, and yet get greater entropy than S_c. In fact the system will try to do this in the way that gives the maximum total entropy. The combination of phases with the highest entropy will lie on the convex hull of the region below S. (The convex hull of a region is the smallest convex region containing it. It contains at least all points of all line segments whose ends lie in the original region.) Thus the actual entropy function will be concave. "Flat parts" of its graph (where a tangent plane contains more than a point) correspond to states which are linear combinations of the states corresponding to the extreme points which are at the boundary of the flat regions (and represent pure phases of the substance).

Notice that if the graph of the entropy contains a straight line segment, then the corresponding derivative along that direction is constant. Thus all points in a flat region have the same values for the intensive variables corresponding to the flat directions. If we choose an underlying smooth entropy function arbitrarily, it is non-generic for it to contain any straight line segments (though one would have to verify that this is true of entropies that arise from physically possible statistical mechanical situations). Therefore all the phase transition type behavior comes from taking the convex hull, and we may classify the possibilities.

If we have only one extensive quantity, say V (as in an isothermal Van der

Waals gas), then there is only one type of structure we can get, which is depicted in figure (16.5). We concavify the entropy by adding a line segment which is tangent at the points where it touches the original graph of S. The endpoints of this line segement correspond to the liquid and gas phases (for instance). Physically we think of a puddle of fluid in the bottom of a volume that we are expanding at constant temperature. As we increase the volume the fluid evaporates at constant pressure (the vapor pressure) until it is all gas. Thus, as we move along the segment from one to the other, the proportion changes from all of one to all of the other. The pressure is the slope, so the whole change takes place at constant pressure.

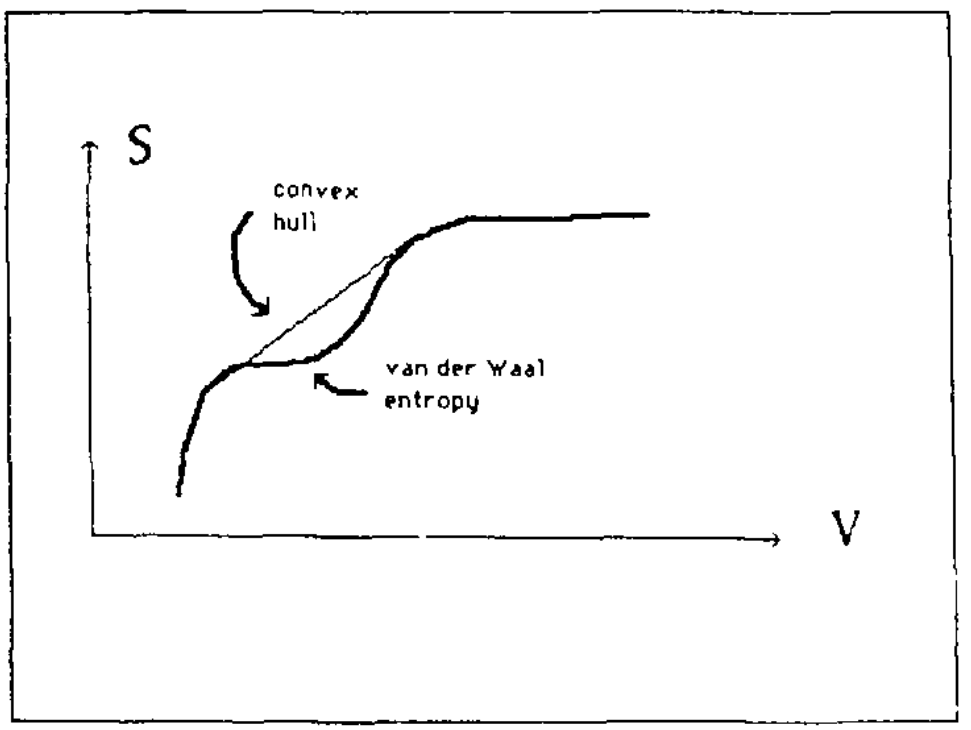

Figure 16.5: Isothermal entropy as a function of volume for the Van der Waals gas at the gas-liquid phase transition.

When we consider 2 extensive variables, say U and V, we find several more

possibilities. If we have three "mountain" peaks as in figure (16.6a), then the convex hull contains a triangular flat region. The corners of this triangle represent pure phases (like solid, liquid, and gas) and the interior points represent linear combinations of these pure states. Since there is a 2-dimensional flat spot, there are two intensive variables that are constant, which are here the pressure and the temperature. This is then a triple point of the substance. The edges of the triangle bound two dimensional ruled surfaces which contain only 1-dimensional line segments (since it is not generic for the original terrain to contain line segments). (A ruled surface may be thought of as a curve in the space of lines.) These represent first order phase transitions between two phases as discussed above. The width of the lines can get shorter as we move along the surface and go to zero as the square root of the parameter labeling the line segments. This disappearance is called a critical point. The more usual picture of these phenomena is given in the intensive space of T and p as in (16.6o). We perform the Legendre transform to get to these variables and as we have seen, points in the graph will correspond to tangent planes of the original graph. If the original is convex, then so is its Legendre transform, but if the original has flat spots, then the transform can have discontinuous first derivatives (i.e. corners). Since the first derivatives are discontinuous in this picture, it is called a first order phase transition. The ruled surfaces correspond to edges with a sharp corner, the flat triple point corresponds to three cornered edges coming together as in a tetrahedron vertex, and the critical point is where an edge smooths out.

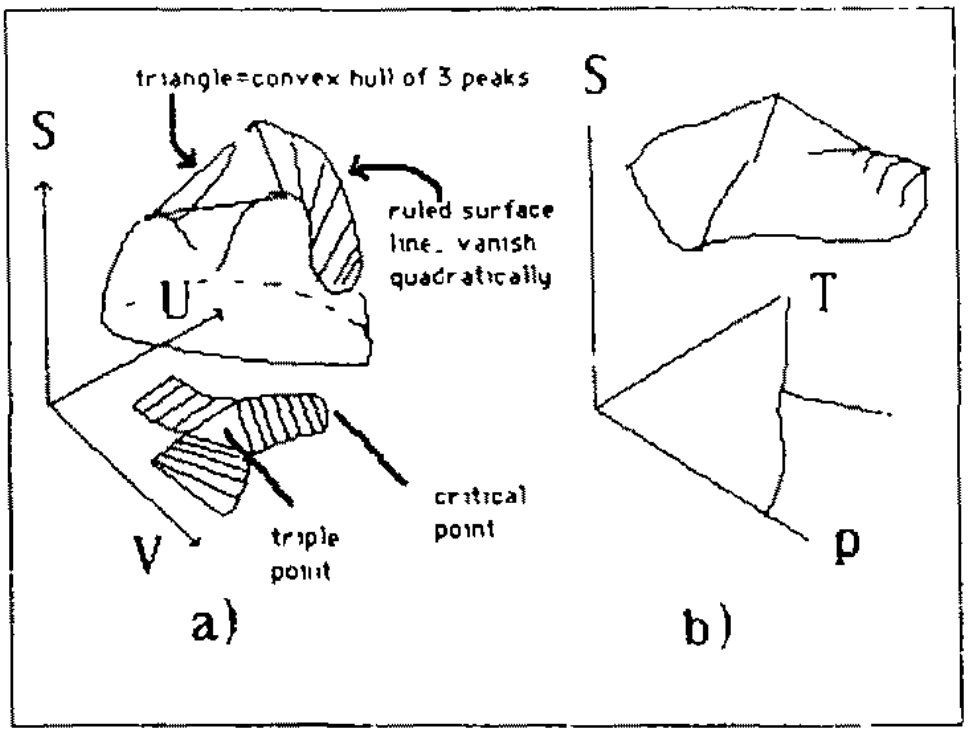

Figure 16.6: a) The entropy as a function of U and V. A flat triangle represents the coexistence of the three phases represented by the corners of the triangle. The ruled surfaces eminating from the edges represent phase transitions of two states. The parabolic end of the ruled surface represents a critical point. b) The entropy on T and p space giving the more usual picture as the Legendre transform of a).

16.8.3. A Generalization of Maxwell's Equal Area Rule

If we are given a manifold M, then we have seen in section 2.1 that J^1M, the space of 1-jets of functions on M, has a natural contact structure. We saw in section 2.4.5 that T^*M, the cotangent bundle, has a natural symplectic structure ω which is minus the differential of the canonical one-form θ. There is a natural projection from J^1M to T^*M, which sends the one-jet of a function at $x \in M$ to its differential there. Legendre submanifolds in J^1M project to Lagrangian submanifolds in T^*M

Any loop in J^1M which lies in a Legendre submanifold therefore projects to a loop in T^*M with zero action (i.e. the integral of θ around the loop vanishes). We may generalize this in

Lemma 16.7. *Any piecewise smooth loop in J^1M whose tangent vector at each point lies in the contact plane at that point projects to a loop with zero action in T^*M.*

Proof. In local coordinates on J^1M, the contact planes are given by the tangent vectors annihilated by the one-form $du - p\,dx$ (where x are coordinates on M, u is the value of the function whose jet the point in J^1M represents, and p its derivative). The integral of this one-form around our loop therefore vanishes (since the loop is tangent to the contact planes). The canonical one form on T^*M pulls back to $p\,dx$ on J^1M. The integral of the canonical one-form is thus equal to the integral of du on each local piece. But u is a well defined function globally on the loop. Therefore the integral of du and therefore of $\theta = pdx$ around the loop is zero. Q.E.D.

Let us now use this lemma to generalize Maxwell's "equal area" rule for first order phase transitions. Let us be given some smooth function S of the variables x_i which are linear coordinates on the linear state space M. This represents the "entropy" as a function of the extensive thermodynamic variables, but without regard for the thermodynamic stability of the state it represents. We have seen that the entropy of the real state of the system, as a function of y will be the smallest concave function S_c that is everywhere greater than or equal to S. Equivalently, the graph of S_c is the boundary of the convex hull of the region below the graph

of S. Consider any piecewise smooth loop in $M \times \mathbb{R}$, whose points are either in the graph of S or the graph of S_ϵ. If we are at a point contained in both graphs, then the differentials dS and dS_ϵ are also equal. (If they weren't equal then their tangent hyperplanes would intersect transversally there and S_ϵ wouldn't be everywhere greater than or equal to S.) We may therefore lift our loop to both J^1M and T^*M by sending each point to the jet or differential of the function whose graph it lies in. Now the tangent vector to the original curve at each point is also tangent to the graph of the function whose jet we use to lift. Thus the tangent vector to our curve in J^1M at each point lies in the corresponding contact plane. By the lemma above, the loop in T^*M has zero action

Furthermore, if S_ϵ is strictly greater than S, then its graph must contain a straight line segment (otherwise we could lower it and still keep it concave). The derivative along this line segment is therefore constant. The Lagrangian submanifold dS_ϵ in T^*M will therefore have a singular projection onto the thermodynamically conjugate (cotangent fiber) variables along this direction. These many states with the same value for the conjugate variables represent different combinations of amounts of the various phases that can coexist with that value. The graph of dS is a Lagrangian submanifold which agrees with dS_ϵ except in this singular region. If we describe a loop consisting of the singular line in the graph of dS_ϵ from one end to the other and then back to the beginning of the line inside the graph of dS, we have just seen that the symplectic area enclosed by the loop is zero.

But this generalizes the usual Maxwell equal area rule. This rule concerns the situation where we hold the temperature fixed and consider the isothermal equation

of state in the (p, V) plane. For the Van der Waals gas, where

$$(V - b)(P + \frac{a}{V^2}) = RT \tag{16.77}$$

these isothermal curves look like figure (16.7). Maxwell showed that the area between the two curves lying above the phase transition line is equal to the area between them below it. Our construction generalizes this to arbitrary loops in the thermodynamic phase space, which need not be isothermal.

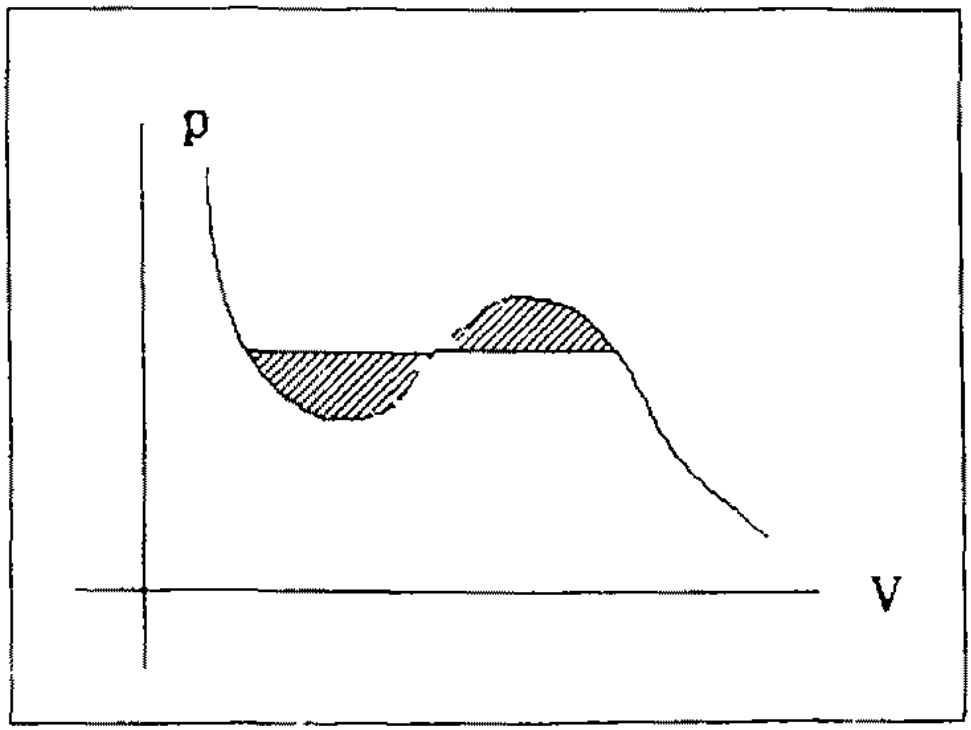

Figure 16.7: The Maxwell equal area construction for the Van der Waals equation of state.

16.9. Relations Between Symplectic Thermodynamics and Mechanics

It is time now to bring together the fascinating structures that we have seen arise asymptotically out of waves and out of statistical mechanics. We have seen many tantalizing clues that these theories have much in parallel and would like to make these structures explicit. That there should be a connection between these theories and some of the parallels were first suggested to me by Robert Littlejohn

16.9.1. A) Eikonal Waves and Stationary Phase

In wave theory we deal with waves in the *eikonal* limit where we study the properties of waves represented by many wavelengths. We introduce asymptotics which stretch the scale length and work with quantities defined in terms of the slow space $y = \epsilon x$. The method of stationary phase lets us asymptotically express quantities that a priori depend on the whole wave in terms of values only near a stationary phase point.

16.9.1. B) Thermodynamic Limit and Steepest Descents

In statistical mechanics we deal with statistics in the *thermodynamic* limit where we study the properties of the statistics of mechanical systems represented by many degrees of freedom. We introduce asymptotics to stretch the scale of the extensive observables and work with quantities defined in terms of the rescaled values. $y = \epsilon x$. The method of steepest descents lets us asymptotically express quantities that depend a *priori* on the entire probability distribution, in terms of values only at a maximum

16.9.2. A) Waves and the Feynman Path Integral

We may view wave mechanics in terms of the Feynman path integral. The value of the wave at a given point in the observation space (typically a point in three space, but more general things may occur) is expressed as an integral, over all paths to the point of interest, of the exponential of i times the action. The action S on a given path is the integral of the Lagrangian along that path. Asymptotically, the exponent scales as $1/\epsilon$. We apply stationary phase to see that only those paths with extremal actions can contribute asymptotically.

16.9.2. B) Probability and the Maximum Entropy Formalism

We may view statistical mechanics in terms of an integral over all observation paths (or equivalently all weighted probability distributions). The value of the probability density at a given point in the observation space (sometimes the three dimensional space of energy U, volume V, and number N, but often more general) is expressed as an integral (the average) of the exponential of the entropy over all probability distributions consistent with the observation point of interest. The entropy S of a given distribution is the integral of $-p \log p$ over that distribution. Asymptotically the exponent scales as $1/\epsilon$. We apply steepest descents to see that only that distribution with maximum entropy can contribute asymptotically.

16.9.3. A) Wave Path Integrals over a Subspace

Often in wave mechanics we don't do the entire path integral at once. An example of a commmon problem is that of finding a light wave at a point in space given its value on some initial surface (or volume, surface, line or point, or many of them may be emitting waves relevant to the region of interest; in the limit we may have a continuum of source types). We may do this by integrating over all paths, but we often like to first integrate over all paths between our point and a given point of the surface, and then integrate the resulting values over the surface. Only the extremal path between the two points will contribute and we may introduce an action defined on the surface, relative to the observation point, which is just the action of the extremal path to that point. The remaining part of the integral to obtain the desired wave value is a finite dimensional integral over the initial surface.

16.9.3. B) Probability Distribution Averages over a Subspace

Often in statistical mechanics we don't do the entire probability integral at once. An example of a common problem is that of finding the probability distribution on the space of two thermodynamic systems in contact with one another (they may exchange any or all of the extensive quantities and there may be many such coupled systems; in the limit we may have a continuum of systems). We may do this by integrating over all distributions consistent with the constraint, but we often like to first integrate over all consistent distributions on the product space for which the first space's thermodynamic quantities have given values and then integrate the resulting distributions over these values. For each set of values on the

first space, only the maximal entropy consistent distributions on the full space will contribute and we may introduce an entropy defined on the space of thermodynamic *observables on the first space, relative to the constraint conditions, which is just* the maximal entropy over all consistent distributions with the given values for the first space. The remaining part of the integral is an integral over the observables of the first system.

16.9.4. A) Lagrange Multipliers and Canonical Conjugacy

For the wave system we may also decide not to impose the constraint that the end of the integrated paths has to end at the point we are interested in. To make the stationary points of this unconstrained problem obey the constraints, we introduce Lagrange multipliers k that are in the dual space to the relaxed constraints. In the wave case, this asymptotically becomes the dual space of the tangent space at the point of observation (i.e. the cotangent space). We say that the variables y and k are canonically conjugate. Instead of the exponential just being of iS/ϵ, it is of $i(S - \langle k, y \rangle)/\epsilon$. Here S is a function of both the initial and final endpoint of the paths, and the integral is over both. We use stationary phase and force the result to apply to the point of interest by choosing k so that the differential of the exponential vanishes at the desired point. This gives: $k = dS$, where this S is a function only of the observation point (the initial point integral already having been done). Thus we see that k is really the wavevector and the eikonal wave is naturally associated with the Lagrangian submanifold defined by dS in the cotangent bundle of y space.

16.9.4. B) Lagrange Multipliers and Thermodynamic Conjugacy

For the statistical system we may also decide not to impose the constraint that the total *mean values of the thermodynamic quantities* are the ones we are interested in. To make the maximum entropy states of this unconstrained problem obey the constraints, we introduce Lagrange multipliers X that are in the dual space to the relaxed constraints. In the statistics case, this asymptotically becomes the dual space of the tangent space of the space of extensive quantities at the point of observation (i.e. the cotangent space). We say that the variables y and X are **thermodynamically conjugate**. Instead of the exponential just being of S/ϵ, it is of $(S - \langle X, y \rangle)$. Here S is a function of both the thermodynamic quantities of the first system and of the total system, and the integral is over both. We use steepest descents and force the result to have the total mean values of interest by choosing X so that the differential of the exponential vanishes at the desired point. This gives $X = dS$, where this S is only a function of the total *mean values* (the integral over the values of the first system already having been done). Thus we see that X is really *the set of conjugate* thermodynamic variables and the overall equation of state is naturally asociated with the Lagrangian submanifold defined by dS in the cotangent bundle of y space.

16.9.5. A) Fourier Transforms and Legendre Transforms

In the wave case, the effect of utilizing the Lagrange multipliers k was to introduce an extra integration over $e^{i(k \cdot y - \omega t)}$. We *recognize* this as the Fourier transform. In general, a wave and its Fourier transform are very different and there is no way

to define our wave as a function of both k and x. We have seen that asymptotically we may introduce the local Fourier transform as a function of k and y by introducing a window function that approaches a δ-function on the slow scale and approaches a constant on the fast wave scale. We find that appropriately scaled Gaussians represent asymptotic states with a definite y and a definite k. There is an absolute uncertainty principle which prevents us from finding such states in x and k. An eikonal wave has a local Fourier transform that is supported on the Lagrangian submanifold $k = dS(y)$. The Fourier transform of an eikonal wave is another eikonal wave whose phase function is the Legendre transform of the original phase function.

16.9.5. B) Laplace Transforms and Legendre Transforms

In the statistics case, the effect of utilizing the Lagrange multipliers X was to introduce an extra integration over $e^{(X,y)/\epsilon}$. We recognize this as the Laplace transform. In general, a distribution and its Laplace transform are very different and there is no way to define a probability distribution on both X and y. We have seen that asymptotically we may introduce the local Laplace transform as a function of X and y by introducing a window function that is an intermediate scale exponential. We find that appropriately scaled Gaussians represent asymptotic distributions with a definite X and a definite y. There is an absolute uncertainty principle that says that asymptotically the dispersion tensor in X and the dispersion tensor in x are inverses. A thermodynamic equation of state has a local Laplace transform that is supported on the Lagrangian submanifold $X = dS(y)$. The Laplace transform of a

thermodynamic probability distribution is another one whose entropy function is the Legendre transform of the original entropy. The partition function is the Laplace transform of the density of states and asymptotically they are both exponentials of quantities over ϵ

Chapter 17: Glossary of Terms

Chapter 18: Bibliography

[Abraham and Marsden, 1978]

R. Abraham and J. E. Marsden, *Foundations of Mechanics, 2nd edition*, Benjamin/Cummings, Reading, Mass., 1978.

[Abraham, Marsden, and Ratiu, 1983]

R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1983.

[Arnold, 1965]

V. I. Arnold, "Sur les propriétées topologiques des applications globalement canoniques de la méchanique classique". *C. R. Acad. Sci. Paris*, Paris, 1965.

[Arnold, 1966]

V. Arnold, "Sur la geometrie differentielle des groupes de Lie de dimension infinie et ses applications a l'hydrodynamique des fluids parfaits", *Ann. Inst. Fourier Grenoble* **16(1)** (1966) 319-361.

[Arnold and Avez, 1968]

V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, W. A. Benjamin, Inc., New York, 1968.

[Arnold, 1978]

V. I. Arnold, *Mathematical Methods of Classical Mechanics*,Graduate Texts in Math., vol. 60, Springer-Verlag, Berlin and New York, 1978

[Arnold, 1983]

V. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer-Verlag, New York, 1983.

[Arnold, 1984]

V. I. Arnold, *Catastrophe Theory*, Springer-Verlag, Berlin, 984.

[Banyaga, 1977]

A. Banyaga, "Formes volumes sur les variétés à bord", *Enseign. Math* **20**, (1977), p. 127-131.

[Birkhoff, 1913]

G. D. Birkhoff, "Proof of Poincaré's geometric theorem", *Trans. Amer. Math. Soc.* **14**, (1913), p. 14-22.

[Born and Wolf, 1970]

M. Born and E. Wolf, *Principles of Optics*, fourth edition, Permagmon Press, Oxford, 1970.

[Brush, 1983]

S. G. Brush, *Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager*, Princeton University Press, Princeton, New Jersey, 1983.

[Burke, 1980]

W. L. Burke, *Spacetime, Geometry, Cosmology*, University Science Books, Mill Valley, California, 1980

[Chaperon, 1983]

M. Chaperon, "A theorem on Lagrangian Intersections", preprint, École Polytechnique, Palaiseau (1983).

[Cary, 1981]

J. R. Cary, "Lie Transform Perturbation Theory for Hamiltonian Systems", *Physics Reports* **79** (1981), 131.

[Chapman and Cowling, 1958]

S. Chapman and T. Cowling, *The Mathematical Theory of Non-uniform Gases*, Cambridge University Press, Cambridge, 1958.

[Chernoff and Marsden, 1974]

P. R. Chernoff and J. E. Marsden, *Properties of Infinite Dimensional Hamiltonian Systems*, Lecture Notes in Math., vol. 425, Springer-Verlag, New York (1974).

[Conley and Zehnder, 1983]

C. C. Conley and E. Zehnder, "The Birkhoff-Lewis fixed point theorem and a conjecture of V. I. Arnold", *Inventiones Math.* 73 (1983), 33-49.

[Cornfeld, Fomin, and Sinai, 1982]

I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory*, Springer-Verlag New York 1982.

[Courant, 1966]

E. D. Courant, "Conservation of Phase Space in Hamiltonian Systems and Particle Beams", in *Perspectives in Modern Physics. Essays in Honor of Hans Bethe*, ed. R. Marshak, Wiley Interscience Publishers, New York, 1966.

[Crawford and Omohundro, 1984]

J. D. Crawford and S. M. Omohundro, "On the Global Structure of Period Doubling Flows", *Physica* 13D **1 & 2**, (1984), p. 161-180.

[de Bruijn, 1981]

N. G. de Bruijn, *Asymptotic Methods in Analysis*, Dover Publications Inc., New York, 1981.

[Dewar, 1972a]

R. L. Dewar, "A Lagrangian Theory for Nonlinear Wave Packets in a Collisionless Plasma", *J. Plasma Physics* **7**, (1972), pp. 267-284.

[Dewar, 1972b]

R. L. Dewar, "A Lagrangian Derivation of the Action Conservation Theorem for Density Waves", *The Astrophysical Journal* **174**, (1972), pp. 301-307.

[Dewar, 1977]

R. L. Dewar, "Energy-Momentum Tensors for Dispersive Electromagnetic Waves", *Aust. J. Phys.* **30**, (1977), pp. 533-575.

[Dewar, 1983]

R. L. Dewar, "Optimal Oscillation-Center Transformations", preprint 1983.

[Dirac, 1958]

P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, 1958.

[Dirac, 1977]

P. A. M. Dirac, *History of Twentieth Century Physics*, Proc. International School of Physics 'Enrico Fermi', Course 57, Academic Press, New York and London, 1977.

[Dubrovin and Novikov, 1983]

B. A. Dubrovin and S. P. Novikov, "Hamiltonian formalism of one-dimensional systems of the hydrodynamic type and the Bogolyubov-Whitham averaging method", *Dokl. Akad. Nauk SSSR* **270**, no. 4, p. 781-785, 1983. English translation: *Soviet Math. Dokl.* **27**, no. 3, p. 665-669, 1983.

[Faddeev and Zakharov, 1971]

L. Faddeev and V. E. Zakharov, "Korteweg-de Vries as a Completely Integrable Hamiltonian System", *Funct. Anal. Appl.* 5 (1971) 280.

[Fefferman, 1983]

C. L. Fefferman, "The Uncertainty Principle", *Bulletin of the American Mathematical Society* **9** (1983), 129–206.

[Feynman, 1948]

R. P. Feynman, "Space-time approach to non-relativistic quantum mechanics", *Reviews of Modern Physics* **20** (1948) p.267.

[Feynman, 1972]

R. P. Feynman, *Statistical Mechanics: A Set of Lectures*, W. A. Benjamin, inc., Reading, Mass., 1972.

[Feynman and Hibbs, 1965]

R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill Book Company, New York, 1965.

[Freed and Uhlenbeck, 1984]

D. Freed and K. Uhlenbeck, *Instantons and Four-Manifolds*, Mathematical Sciences Research Institute Publication 1, Springer-Verlag, New York, 1984.

[Gibbons, Holm and Kupershmidt, 1982]

J. Gibbons, D. D. Holm, and B. Kupershmidt, "Gauge-Invariant Poisson Brackets for Chromohydrodynamics", *Physics Letters* **90A** (1982), 281–283.

[Gibbs, 1873]

J. W. Gibbs, "Graphical Methods in the Thermodynamics of Fluids", *Transactions of the Connecticut Academy* **II**(1873), p. 309-342. Coll. Works I, 1-32.

[Gibbs, 1873b]

J. W. Gibbs, "A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces" *Transactions of the Connecticut Academy II*(1873), p. 382-404. Coll. Works 33-54

[Gilmore, 1981]

R. Gilmore, *Catastrophe Theory for Scientists and Engineers*, John Wiley and Sons, Inc., New York, 1981.

[Grebogi and Littlejohn, 1984]

C. Grebogi and R. G. Littlejohn, "Relativistic Ponderomotive Hamiltonian", preprint, 1984.

[Grebogi, Ott, and Yorke, 1984]

C. Grebogi, E. Ott, and J. Yorke, "Strange Attractors Without Chaos", poster session, American Physical Society Division of Plasma Physics meeting, Boston, 1984.

[Guckenheimer and Holmes, 1983]

J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983.

[Guillemin and Sternberg, 1977]

V. Guillemin and S. Sternberg, *Geometric Asymptotics*, American Mathematical Society, Providence, Rhode Island, 1977.

[Hermann, 1968]

R. Hermann, *Differential Geometry and the Calculus of Variations*, Academic Press, New York, 1968.

[Guillemin and Sternberg, 1984]

V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge, 1984.

[Hirsch, 1984]

M. W. Hirsch, "The Dynamical Systems Approach to Differential Equations", *Bull. of Am. Math. Soc.* **11** pps. 1-64.

[Holm and Kupershmidt, 1984a]

D. Holm and B. Kupershmidt, "Poisson Brackets and Clebsch Representations for Magnetohydrodynamics, Multifluid Plasmas, and Elasticity", *Physica* **D**.

[Holm and Kupershmidt, 1984b]

D. D. Holm and B. A. Kupershmidt, "Poisson Structures of Superfluids", submitted to *Physics Letters* **A**.

[Holm, Marsden, Ratiu, and Weinstein, 1984]

D. D. Holm, J. E. Marsden, T. Ratiu, and A. Weinstein, "A Priori Estimates for Nonlinear Stability of Fluids and Plasmas", to be published in *Physics Reports*, (1984).

[Holmes and Whitley, 1984]

P. Holmes and D. Whitley, "Bifurcations of One and Two-Dimensional Maps", Preprint 1984.

[Iagolnitzer, 1975]

D Iagolnitzer "Microlocal essential support of a distribution and decomposition theorems: an introduction", in *Hyperfunctions and theoretical physics*. Springer Lecture Notes in Math. 449. 1975. pps. 121-132

[Hörmander, 1983]

L. Hörmander. *The Analysis of Linear Partial Differential Operators I*, Springer-Verlag. Berlin, 1983.

[Israel, 1979]

R. Israel, *Convexity in the Theory of Lattice Gases*. Princeton University Press, Princeton, New Jersey. 1979.

[Jaynes, 1983]

E.T. Jaynes, *Papers on probability, statistics and statistical physics*, ed. by R. D. Rosenkrantz, D. Reidel Publishing Company, Boston. 1983.

[Jenner, 1963]

W. E. Jenner, *Rudiments of Algebraic Geometry*. Oxford University Press, Inc., New York, 1963.

[Kac, 1959]

M. Kac, *Probability and Related Topics in Physical Sciences*. Lectures in Applied Mathematics. Volume 1A. American Mathematical Society. Providence. Rhode Island. 1959.

[Kaufman, Cary, and Pereira, 1979]

A. N. Kaufman, J. R. Cary, and N. R. Pereira, "Universal Formula for Quasi-static Density Perturbation by a Magnetoplasma Wave", *Phys. Fluids* **22**(4), 1979, p. 790.

[Kaufman, 1979]

A. N. Kaufman, "Regular and Stochastic Particle Motion in Plasma Dynamics", in *The Proceedings of the International Workshop on Intrinsic Stochasticity in Plasmas*, Cargese, Corsica, France, 1979.

[Kaufman, 1982]

A. N. Kaufman, "Elementary Derivation of Poisson Structures for Fluid Dynamics and Electrodynamics", *Phys. Fluids* **25**(11), 1982, p. 1993.

[Kaufman and Holm, 1984]

A. N. Kaufman and D. D. Holm, "The Lie-Transformed Vlasov Action Principle: Relativistically Covariant Wave Propagation and Self-Consistent Ponderomotive Effects", *Physics Letters* **105A** #6, (1984) p277-279.

[Kaufman and Boghosian, 1984]

A. N. Kaufman and B. M. Boghosian, "Lie-transform Derivation of the Gyrokinetic Hamiltonian System", in *Fluids and Plasmas: Geometry and Dynamics*, J. Marsden Ed., American Mathematical Society, Providence, RI, 1984.

[Kaufman, 1984]

A. N. Kaufman, "A Lie-transformed Action Principle for Classical Plasma Dy-

namics", in *The Proceedings of the XIII International Colloquium on Group Theoretical Methods in Physics*, College Park, MD, 1984.

[Keller, 1958]

J. B. Keller, *Ann. Physics* **4** (1958) p. 180-188.

[Kevorkian and Cole, 1981]

J. Kevorkian and J. D. Cole, *Perturbation Methods in Applied Mathematics*, Springer-Verlag, New York, 1981.

[Khinchin, 1949]

A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publications, Inc., New York, 1949.

[Kijowski and Tulczyjew, 1979]

J. Kijowski and W. Tulczyjew, *A symplectic framework for field theories*, Lecture Notes in Physics 107, Springer, New York, 1979.

[Klein, 1928]

F. Klein, *Vorlesungen ub. die Entwicklung der Mathematik im 19 Jahrhundert*, Springer-Verlag, Berlin, 1928, translation *Development of Mathematics in the 19th Century* by M. Ackerman, Math Sci Press, Brookline, Massachusetts, 1979.

[Kruskal, 1962]

M. Kruskal, "*Asymptotic theory of Hamiltonian and other systems with all solutions nearly periodic*", *J. Math. Phys.* **3**, (1962) 806-828

[Kubo et al., 1965]

R. Kubo. H. Ichimura, T. Usui, N. Hashitsume, *Statistical Mechanics*. North-Holland, Amsterdam, 1965.

[Lagrange, 1808]

J. L. Lagrange, "Mémoire sur la théorie des variations de éléments des plan-'etes", *Mémoires de la classe des sciences mathématiques et physiques de l'instiut de France* (1808), pp. 1–72.

[Landau and Lifshitz, 1960–1981]

L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics, Volumes 1–10*, Pergamon Press Ltd., New York (1960–1981).

[Littlejohn, 1979]

R. G. Littlejohn, "A guiding center Hamiltonian· A new approach", *J. Math. Phys.* **20**(12) (1979) p2445-2458.

[Littlejohn, 1981]

R. G. Littlejohn, "Hamiltonian Formulation of Guiding Center Motion", *Phys. Fluids.* **24** (1981) p1730-1749.

[Littlejohn, 1983]

R. G. Littlejohn, "Variational Principles of Guiding Centre Motion", *J. Plasma Physics* **29** (1983) 111.

[Littlejohn, 1983]

R. G. Littlejohn, "Geometry and Guiding Center Motion", in *Fluids and Plasmas, Geometry and Dynamics*, J. Marsden Ed., American Mathematical Society, Providence, RI, 1983.

[Mackey, 1978]

G. W. Mackey, *Unitary Group Representations in Physics, Probability, and Number Theory*, Benjamin/Cummings, Reading, Massachusetts, 1978.

[Marsden, 1981]

J. E. Marsden, *Lectures on Geometric Methods in Mathematical Physics*, CBMS-NSF Regional Conference Series 37, Philadelphia, PA:SIAM, 1981.

[Marsden and Hughes, 1983]

J. E. Marsden and T. J. R. Hughes, *Mathematical Foundations of Elasticity*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[Marsden, Ratiu, and Weinstein, 1984]

J. E. Marsden, T. Ratiu, and A. Weinstein, "Semi-direct Products and Reduction in Mechanics", *Transactions of the American Mathematical Society* **281** (1984), pp. 147-177.

[Marsden and Weinstein, 1974]

J. Marsden and A. Weinstein, "Reduction of symplectic manifolds with symmetry", *Rep. Math. Phys.* **5** (1974) 121-130.

[Marsden and Weinstein, 1982]

J. Marsden and A. Weinstein, "The Hamiltonian Structure of the Maxwell-Vlasov Equations", *Physica* **4D** (1982), 394–406.

[Marsden and Weinstein, 1983]

J. E. Marsden and A. Weinstein, "Coadjoint Orbits, Vortices, and Clebsch Variables for Incompressible Fluids", *Physica* **7D** (1983), 305–323.

[Maslov, 1965]

V. P. Maslov, *Theory of Perturbations and Asymptotic Methods*, Izdat. Moskov. Gos. Univ., Moscow 1965; French transl., Dunod, Paris, 1972.

[Maslov, 1976]

V. P. Maslov, *Operational Methods*, Mir Publishers, Moscow, 1976.

[Misner, Thorne, and Wheeler, 1973]

C. Misner, K. Thorne, and J. Wheeler, *Gravitation*, Freeman, San Francisco, 1973

[Morrison, 1980]

P. J. Morrison, "The Maxwell-Vlasov Equations as a Continuous Hamiltonian System", *Phys. Lett.* **80A** (1980), 393–396.

[Morrison and Greene, 1980]

P. J. Morrison and J. M. Greene, "Noncanonical Hamiltonian Density Formulation of Hydrodynamics and Ideal Magnetohydrodynamics", *Phys. Rev. Letters* **45** (1980), 790–794.

[Nayfeh, 1973]

A. H. Nayfeh, Perturbation Methods, John Wiley and Sons, Inc., New York, 1973.

[Noether, 1918]

E. Noether, "Invariante Variationsprobleme", *Nachrichten kgl. Gesell. Wissenschaft. Göttingen, Math. Phys. Kl.* **2** (1918), pp. 235-257

[Northrop, 1963]

T. G. Northrop, *The Adiabatic Motion of Charged Particles*, John Wiley and Sons, Inc., New York, 1963.

[Olver, 1984]

P. Olver, "Hamiltonian perturbation theory and water waves", in J. Marsden, ed., *Fluids and Plasmas: Geometry and Dynamics*, American Mathematical Society series in Contemporary Mathematics, Volume 28,1984, p. 231.

[Omohundro, 1984a]

S. M. Omohundro, "Hamiltonian Structures in Perturbation Theory", to be submitted to the *Journal of Mathematical Physics* (1984).

[Omohundro, 1984b]

S. M. Omohundro, "Geometric Hamiltonian Structures and Perturbation Theory" *in Lecture Notes in Physics volume on Local and Global Methods in Nonlinear Dynamics*, ed. B. Cawley, Springer-Verlag, New York (1984).

[Omohundro, 1984c]

S. M. Omohundro, "A Hamiltonian Approach to Wave Modulation", Poster presented at April 1984 Sherwood Plasma Theory Meeting, Lake Tahoe, in preparation for publication.

[Omohundro, 1984d]

S. M. Omohundro, "Modelling Cellular Automata with Partial Differential Equations", *Physica* **10D** (1984) 128-134.

[Omohundro, 1984e]

S. M. Omohundro, "Book Review of *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors* by C. Sparrow", *Mathematical Biosciences* **68** (1984) p. 303-304.

[Ott, 1979]

E. Ott, "Goodness of Ergodic Adiabatic Invariants", *Physical Review Letters* **42** (1979) p.1628-1631.

[Palais, 1961]

R. S. Palais, "Equivalence of nearby differentiable actions of a compact group", *Bull. Amer. Math. Soc.* **67**, p. 362-364, 1961

Pauli, 1933]

W. Pauli, *General Principles of Quantum Mechanics*, (1933) Reprinted in English Translation by Springer-Verlag, Berlin, 1981.

[Poincaré, 1892]

H. Poincaré, *Les méthodes nouvelles de la mécanique céleste, 1,2,3.* Gauthier-Villars, Paris (1892); Dover, New York, 1957.

[Poincaré, 1912]

H. Poincaré, "Sur un théorème de géométrie", *Rend. Circ. Math. Palermo* **33**(1912), 375–407.

[Poston and Stewart, 1978]

T. Poston and I. Stewart, *Catastrophe Theory and its Applications*, Pitman, Boston, 1978.

[Schulman, 1981]

L. Schulman, *Techniques and Applications of Path Integration*, John Wiley and Sons, Inc., New York, 1981.

[Shannon, 1948]

C. E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal* **27**, (1948),p. 379.

[Simon, 1979]

B. Simon, *Functional Integration and Quantum Physics*, Academic Press, New York, 1979.

[Smale, 1967]

S. Smale, "Differential Dynamical Systems", *Bulletin of the American Mathematical Society* **73**, (1967), pp. 747–817, reprinted in: S. Smale, *The Mathematics*

of Time: Essays on Dynamical Systems, Economic Processes, and Related Topics, Springer-Verlag, New York, 1980.

[Smith and Kaufman, 1975]

G. R. Smith and A. N. Kaufman, "Stochastic Acceleration by a Single Wave in a Magnetic Field", *Physical Review Letters* **34** #26 (1975) 1613-1616.

[Smith and Kaufman, 1978]

G. R. Smith and A. N. Kaufman, "Stochastic acceleration by an obliquely propagating wave– An example of overlapping resonances", *Phys. Fluids* **21**(12), 1978, p. 2230.

[Souriau, 1970a]

J-M. Souriau, *Structure des Syst'emes Dynamiques*, Dunod, Paris, 1970.

[Souriau, 1970b]

J-M. Souriau, "Mécanique Statistique, Groupes de Lie et Cosmologie", in *Colloques Internationaux C.N.R.S.*, No. 237- Géométrie symplectique et physique mathématique p. 59.

[Spencer and Kaufman, 1982]

R. G. Spencer and A. N. Kaufman, "Hamiltonian Structure of Two-Fluid Plasma Dynamics", *Phys. Rev.* **A 25** (1982) 2437–2439.

[Spivak, 1979]

M. Spivak, *A Comprehensive Introduction to Differential Geometry, Volume One, Second Edition*, Publish or Perish, Inc., Berkeley, 1979.

[Stoer and Burlirsh, 1980]

J. Stoer and R. Burlirsh, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1980.

[Taylor, 1981]

M. E. Taylor, *Pseudodifferential Operators*, Princeton University Press, Princeton, New Jersey, 1981.

[Thirring, 1978]

W. Thirring, *A Course in Mathematical Physics, I: Classical Dynamics*, Springer-Verlag, New York, 1978.

[Tikochinsky, Tishby, and Levine, 1984]

Y. Tikochinsky, N. Z. Tishby, and R. D. Levine, "Alternative Approach to Maximum-entropy Inference", *Physical Review A* **30**, (1984), pp. 2638–2644.

[Tisza, 1966]

L. Tisza, *Generalized Thermodynamics*, The M.I.T. Press, Cambridge, Massachusetts, 1966.

[van Kampen, 1983]

N. van Kampen, "Constraints", preprint, March 1983.

[Wax, 1954]

N. Wax, editor, *Selected Papers on Noise and Stochastic Processes*, Dover Publications, New York, 1954.

[Weinstein, 1971]

A. Weinstein, "Symplectic Manifolds and their lagrangian submanifolds", *Advances in Math.* **6** (1971), 329–346.

[Weinstein, 1977]

A. Weinstein, *Lectures on Symplectic Manifolds*, Regional Conferences in Math. **29**, American Math. Society (1977).

[Weinstein, 1978]

A. Weinstein, "The Order and Symbol of a Distribution", *Transactions of the American Mathematical Society*, **241** (1978) p.1-54.

[Weinstein, 1981]

A. Weinstein,"Symplectic Geometry", *Bulletin of the American Mathematical Society* **5** (1981), 1–13.

[Weinstein, 1983a]

A. Weinstein, "The Local Structure of Poisson Manifolds", *J. Differential Geometry* **18** (1983) 523–557.

[Weinstein, 1983b]

A. Weinstein, "Sophus Lie and symplectic geometry", *Expo. Math.* **1**(1983) 95–96.

[Weinstein, 1983c]

A. Weinstein, "C^0 perturbation theorems for symplectic fixed points and lagrangian intersections", Preprint, Univ. of California, Berkeley (1983).

[Weinstein, 1984a]

A. Weinstein, "Critical Point Theory, Symplectic Geometry, and Hamiltonian Systems", Preprint, Univ. of California, Berkeley (1984).

[Weinstein, 1984b]

A. Weinstein, notes by S. M. Omohundro, "Equations of Plasma Physics", p. 359 in *Seminar on Nonlinear Partial Differential Equations*, edited by S.'S. Chern, Mathematical Sciences Research Institute Publication 2, Springer-Verlag, 1984.

[Whitham, 1974]

G. Whitham, *Linear and Nonlinear Waves*, John Wiley & Sons, Inc., New York, 1974.

[Wilson and Kogut, 1974]

K. G. Wilson and J. Kogut, "The Renormalization Group and the ϵ Expansion", *Physics Reports* **12** No. 2 (1974) p. 75-200.

[Wolfram, 1984]

S. Wolfram, "Undecidability and Intractability in Theoretical Physics", Preprint 1984.

[Yaffe, 1982]

L. G. Yaffe, "Large N Limits as Classical Mec'.anics", *Reviews of Modern Physics* **54** No. 2 (1982) p. 407-435.