ON THE DYNAMICAL PROPAGATION OF SUBVOLUMES AND ON THE GEOMETRY AND VARIATIONAL PRINCIPLES OF NONHOLONOMIC SYSTEMS

by

Jared Michael Maruskin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2008

Doctoral Committee:

Professor Anthony M. Bloch, Co-Chair
Associate Professor Daniel J. Scheeres, Co-Chair
Professor Mario Bonk
Professor Peter S. Smereka

“Da steh ich nun, ich armer Tor,
und bin so klug als wie zuvor!”

Faust I, Johann Wolfgang von Goethe, 1806

In Memoriam

Kenneth Hart Keppen

1967 - 2005

ACKNOWLEDGEMENTS

I would above all like to thank my advisors, Prof. Anthony M. Bloch and Prof. Daniel J. Scheeres, for all of the support and encouragement that they offered me over the last few years. I am especially grateful for the intellectual freedom they afforded me, which contributed greatly to the cultivation of an original and unique research program, culminating in the writing of this thesis. I am indebted to the innumerable conversations I've had with each of them pertaining to the topics contained in these pages, and further to the invaluable insight and expertise they brought to this research. I'd additionally like to thank them for financial support for Summer 2005, Winter 2007, and Winter 2008 semesters, as well as travel support to the SIAM Conference on Dynamical Systems, the SIAM Conference on Control Theory and its Applications, the AAS/AIAA Astrodynamics Specialists Conference, and the Conference on Decision and Control. I would also like to thank Rackham for the two student travel grants they awarded me, and SIAM for the travel grant that they awarded me.

I would also like to thank the other members of the doctoral committee, Prof. Peter Smereka and Prof. Mario Bonk, for the time and energy they spent in reviewing these pages and serving on the committee. I'd further like to acknowledge the contributions and scientific input to this thesis made by our collaborators: Mario Bonk, Fred C. Adams, Kyle T. Alfriend, Dmitry Zenkov, and Jerrold E. Marsden.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

In this chapter we will first present an overview of the journal and conference publications, submitted, accepted, and in progress, which form the backbone of this thesis. Next we will present a comprehensive summary of the thesis and an overview of some of the main results.

1.1 In Press

The thesis you have in front of you is a comprehensive compilation of the research I performed as a Ph.D. Candidate at the University of Michigan during the years 2005-2007 under the advising of Anthony M. Bloch and Daniel J. Scheeres. The research presented in this thesis also benefited from suggestions and scientific input from Mario Bonk, Fred C. Adams, Kyle T. Alfriend, Dmitry Zenkov, and Jerrold E. Marsden. This thesis draws heavily upon the following research papers and conference talks. It is, however, written to provide for a greater contextual understanding of this research and includes some additional discussions not included in the below references.

- **Journal Papers**

  [88] J.M. Maruskin and A.M. Bloch. The Boltzmann-Hamel equations for the optimal control of mechanical systems with nonholonomic constraints. Ac-

cepted for publication in *Systems and Control Letters*, 2008.

[90] J.M. Maruskin, A.M. Bloch, J.E. Marsden, and D. Zenkov. A geometric foundation of quasi-velocities in nonholonomic and vakonomic dynamics. preprint, 2008.

[91] J.M. Maruskin, D.J. Scheeres, F.C. Adams, and A.M. Bloch. The eccentric frame decomposition for central force fields. *Journal of Celestial Mechanics and Dynamical Astronomy*, 100 (1):43-62, 2008.

[93] J.M. Maruskin, D.J. Scheeres, and K.T. Alfriend. Correlation of optical observations of objects in earth orbit. Submitted to *The Journal of Guidance, Control, and Dynamics*, 2007.

[96] J.M. Maruskin, D.J. Scheeres, and A.M. Bloch. Dynamics of symplectic subvolumes. Accepted for publication in *The SIAM Journal of Applied Dynamical Systems*, 2008.

- **Conference Proceedings**

[89] J.M. Maruskin and A.M. Bloch. The Boltzmann-Hamel equations for optimal control. In *The Proceedings of the 46th IEEE Conference on Decision and Control*, pages 554-559 2007.

[92] J.M. Maruskin, D.J. Scheeres, and K.T. Alfriend. Correlation of optical observations of objects in Earth orbit. In *The Proceedings of the Seventh US/Russian Space Surveillance Workshop*, 2007.

[94] J.M. Maruskin, D.J. Scheeres, and A.M. Bloch. Dynamics of symplectic subvolumes. In *The Proceedings of the 46th IEEE Conference on Decision and Control*, pages 5600-5605, 2007.

[95] J.M. Maruskin and D.J. Scheeres. Delaunay variables and the tracking of space debris. In *The Proceedings of the 2007 AAIA/AAS Astrodynamics Specialist Conference*, 2007.

- **Conference Talks**
  - **Talk:** "The Eccentric Frame Method with Application to the Hernquist-Newton Potential," presented at The 38th American Astronomical Society Division of Dynamical Astronomoy Conference, May 7 - 11, 2007, Ann Arbor, MI.
  - **Poster:** "Dynamics of Symplectic SubVolumes," presented at the SIAM Conference on Applications of Dynamical Systems, May 28 - June 1, 2007, Snowbird, Utah.
  - **Talk:** "The Geometry of Quasi-Velocities and the Boltzmann-Hamel equations for Optimal Control," presented at the SIAM Conference on Control and its Applications, June 29-July 1, 2007, San Francisco, CA.
  - **Talk:** "Delaunay Variables and the Tracking of Space Debris," presented at the American Astronautical Society Astrodynamics Specialist Conference, August 19-23, 2007, Mackinaw Island, MI.
  - **Invited Talk (presented by Dan Scheeres):** "Correlation of Optical Observations of Objects in Earth Orbit," presented at the Seventh US/Russian Space Surveillance Workshop, October 29 - November 2, 2007, Monterey, CA.
  - **Talk:** "The Boltzmann-Hamel Equations for Optimal Control," presented at the 46th IEEE Conference on Decision and Control, Deceber 12-14, 2007, New Orleans.

– **Talk:** "Dynamics of Symplectic SubVolumes," presented at the 46th IEEE Conference on Decision and Control, December 12-14, 2007, New Orleans.

1.2 Overview of Thesis

The main focus of this thesis is the understanding and analysis of mechanical systems from a geometric perspective. This is divided into two main themes, the dynamical propagation of subvolumes and the variational principles of nonholonomic mechanics. We will also discuss some modern problems of astrodynamics, especially in Chapters III and IV.

Throughout the rest of this chapter we will introduce each of these main themes and further strive to present a comprehensive overview of the research. §1.3 will be devoted to the introduction of subvolumes in dynamical systems in the context of Space Situational Awareness. In particular we will illustrate the essential ideas behind performing orbit determinations from a subvolume approach. In §1.4 we introduce nonholonomic mechanics and quasi-velocities, and further provide an overview of the key results contained in this thesis.

Chapter II is devoted to the dynamical propagation of subvolumes in Hamiltonian systems, and discusses various symplectic constraints that exist on these subvolumes. We will introduce a set of integral invariants of Wirtinger type that are closely related to the Poincaré-Cartan integral invariants, and discuss the physical significance of these new constraints. We will then present a theorem about the expansion of subvolumes and discuss how this, in general, leads to the local collapse of phase space along solution trajectories. Despite this feature, there is always a preferred symplectic basis that exists for any symplectomorphism that resists this collapse. Finally, we indicate how one might utilize these results for the tracking of asteroids and in

optimal control problems.

Chapter III will be devoted to the progress we have made in the problem of Space Situational Awareness. We present a new technique named Intersection Theory Analysis and show how one can use it to perform correlation and orbit determination between two previously uncorrelated tracks of data made of passing space debris particles from optical observers. We also discuss implications of the internal structure found in the Jacobian and State Transition Matrices, utilizing some of the theory from Chapter II.

The next two chapters are devoted to the geometry and variational principles of nonholonomic mechanics and control. In Chapter V we introduce nonholonomic systems and quasi-velocities, and discuss the precise differential geometry of variational principles in nonholonomic systems. We show the precise nature of why, with the addition of nonholonomic constraints, the theory bifurcates into the two distinct formalisms of dynamical and vakonomic mechanics, each theory in general yielding a different set of solution trajectories. We derive the equations of motion for each case separately. The vakonomic motion is not the physical motion of the system, however the vakonomic formalism has implications for optimal control problems. In Chapter V, we generalize the quasi-velocity techniques, such as Maggi's equations and the Boltzmann-Hamel equations, to the vakonomic setting. We end with a discussion of mechanics on Lie groups, relating the Euler-Poincaré equations to the quasi-velocity setting. Chapter VI will then be spent generalizing the Boltzmann-Hamel equations and the Euler-Poincaré equations to the optimal control setting. We will consider both kinematic and dynamic optimal control problems.

Finally we will conclude with a chapter that fits into the primary main theme, geometry of mechanics, but does not fit into either of the two subthemes. We will

present a new analytical technique for decomposing particle motion in central force fields. The solution of the central force field problem is well known. We present a new analytical tool for visualizing the motion which can also serve as a standard decomposition for describing the dynamics. We will show how a "false" bifurcation, which occurs only when viewing the motion from these special, nonuniformly rotating frames, must occur in the system as you vary the parameters towards a stable circular orbit. The intellectual merit of this decomposition lies in the fact that it allows you to use Floquet theory in analyzing the stability of equatorial motion in axi-symmetric force fields, a technology previously prohibited due to the nonclosure of the orbits.

1.3 Subvolume Propagation and Space Situational Awareness

1.3.1 Background and Introduction

Traditional approaches to studying the dynamics and control of mechanical systems usually focus on individual trajectories and states to determine their dynamic evolution or reachable sets. In reality, however, system states are never precisely known and always exist within some set of finite volume in phase space. While the properties of such distributions are related to the ensemble properties of the individual states, there are additional and fundamental properties and constraints that arise when dealing with finite volume distributions in dynamical systems. The purpose of this research is to better understand these constraints and to apply them to a problem of practical importance.

Specifically, the proposed research will develop a finite volume methodology for the study of dynamics and control problems in which we focus on ensembles of states modeled as compact sets in phase space. This is also a more realistic and robust approach to the study of practical problems, especially in problems where the initial uncertainty distribution is not localized to the neighborhood of any single point.

Our approach takes this view and considers the dynamics and control of compact sets of states in phase space. While such an approach appears, at face value, to only consist of summing over results for individual trajectories, it actually leads to a number of deep results that cannot be discerned when only considering the evolution of individual states.

This research will specifically apply recent progress in the field of symplectic topology to the dynamics and control of conservative mechanical systems. We will incorporate a variety of techniques that exist for a certain group of closed, symplectic subsets of the system, in order to understand the implication they have on the dynamical propagation of uncertainty distributions. Such constraints include Liouville's Theorem, the Integral Invariants of Poincaré-Cartan, and Gromov's Nonsqueezing Theorem (NST) [66, 98], as well as additional results we proposed that apply to lower dimensional sets, including Wirtinger-type integral invariants and volume expansion factors [94]. We have already carried out some research connecting Gromov's NST to practical problems in engineering dynamics and control for linear and nonlinear systems [67, 123]. We intend to apply these results specifically to the problem of Space Situational Awareness (SSA), which seeks to perform orbit determination for the entire class of earth orbiting bodies. The relation between these theoretical results in symplectic topology to SSA will be further elaborated upon in §1.3.2.

1.3.2 Symplectic Constraints

Area Expansion Factors

In [94] we present the idea of studying constraints on the evolution of lower dimensional subvolumes of symplectic spaces. Consider a 2-d surface τ (extendable to $2k$-subvolumes by analogy) that admits a 1-to-1 projection to one of the symplectic planes, see Fig. 1.1. The surface then maps into the future by the Hamiltonian phase

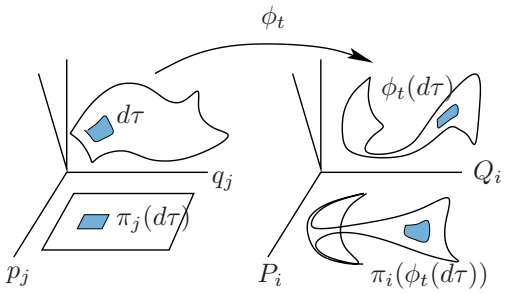

Figure 1.1: Area Expansion Factors, π_i is the i-th symplectic projection map

flow $\phi_t : \langle p, q\rangle \rightarrow \langle P, Q\rangle$. The State Transition Matrix (STM) Φ (the matrix of $d\phi_t$) can be solved by integrating a set of $4n^2$ differential equations to the final time T. As is well known, Liouville's Theorem may be stated in terms of the STM as $\det \Phi = 1$. In [94] we show additional constraints on Φ; in particular, that the sum of the 2×2 symplectic submatrices of any symplectic row or column of Φ must add to unity. This is equivalent to the symplecticity condition $\phi_t^*(\omega) = \omega$, where $\omega = dp \wedge dq$ is the standard symplectic form. Moreover, we characterize all of the area expansion factors and symplectic projection factors quantitatively in terms of Gram determinants of various submatrices of the STM, i.e., all of the shaded area elements shown in Fig. 1.1 are given by the internal structure of Φ and can easily be computed. This has relevance to asteroid tracking [91] and the orbit determination of space debris particles [95, 92]. It also provides a straightforward method of constructing the surface metric of $\phi_t(\tau)$ directly from Φ.

Gromov's Nonsqueezing Theorem (NST)

Gromov's Nonsqueezing Theorem (NST) provides a fundamental limit on full $2n$-dimensional volumes in symplectic spaces [66, 98]. It ascribes to every closed, connected $2n$-dimensional set Ω a nonnegative number known as its *symplectic width*. The symplectic width of Ω is a symplectic invariant. Gromov's NST states, in a practical sense, that the symplectic projection of Ω onto any of the symplectic planes

must have a projected area at least as great as the symplectic width of Ω. This places a fundamental limit on the propagation of orbit uncertainty distributions in Hamiltonian dynamical systems. We have already had some success in exploring the physical implications of this theory on spacecraft uncertainty propagation [123]. The NST acts much like Heisenberg's uncertainty principle in quantum mechanics in determining a fundamental lower limit to which we can know a spacecraft's future location in its coordinate and momentum space when mapped forward in time from an initial covariance distribution.

Poincaré-Cartan and Wirtinger-Type Integral Invariants

The classical integral invariants of Poincaré-Cartan and a closely related integral invariant, that we introduced in [94], provide fundamental constraints on evolution of subvolumes in Hamiltonian systems [67, 123]. Let Ω be a $2k$-dimensional subvolume of the $2n$ dimensional phase space ($k < n$) and ϕ_t the Hamiltonian phase flow. The integral invariants of Poincaré-Cartan states

$$\frac{1}{k!}\int_{\Omega}\omega^k = \frac{1}{k!}\int_{\phi_t(\Omega)}\omega^k,$$

so that the sum of the oriented $2k$-volume projections on each symplectic "$2k$ plane" is conserved. In [94], we introduce the Wirtinger-type integral invariant and prove the following inequality:

$$\frac{1}{k!}\int_{\Omega}|\omega^k| = \frac{1}{k!}\int_{\phi_t(\Omega)}|\omega^k| \leq \mathrm{Vol}_{2k}(\phi_t(\Omega)) \qquad \forall t \in \mathbb{R}, \tag{1.1}$$

so that this integral invariant represents a minimum $2k$-volume that the body Ω may obtain. When Ω is a full $2n$-dimensional volume, the inequality is replaced with an equality (Liouville's Theorem). This inequality therefore acts as a lower dimensional version of Liouville's theorem and places a fundamental constraint on the

evolution of lower dimensional uncertainty distributions in Hamiltonian dynamical systems. In [94] we also relate this constraint to the local collapse of phase space prevalent in chaotic systems. Regardless of the chaoticity or time duration, every canonical transformation has with it, locally, a distinguished symplectic basis which resists collapse [94]. We believe this basis has implications for the identification of the maximal and minimal growth directions and could have application to optimal control law design in the face of uncertainty.

Space Situational Awareness and Intersection Theory Analysis

A main and motivating component of our research will be applying the theory of symplectic subvolumes to developing the technology for efficient tracking, orbit determination, and cataloging of space debris particles in earth orbit, a population of more than 300,000 orbiting bodies. This represents a problem of recent interest to space faring nations, and is referred to as Space Situational Awareness (SSA). The United States Air Force Space Command in particular has installed a network of 25 radar and optical sensors for this task, which make several thousand observations per night. New, high powered telescopes, such as PanSTARRS and the DAPRA SST, are about to come online, that will increase the size of the catalog from 15,000 objects to 150,000-300,000 objects. This compounds the necessity for new and efficient algorithms to aid in the orbit determination process. The feasibility of this approach has been explored in [95, 92].

The basic problem arises due to our inability to make a complete orbit determination from a single radar or optical observation. Such an observation is known as an uncorrelated track (UCT). The objective is to correlate UCT's that belong to the same physical object. An optical sensor can measure the angles and angular rates of an object during a single pass. An admissible region in the range and range rate

plane can then be computed for each observation from certain physical constraints [95, 130]. The standard approach to fitting an orbit between two separate UCT's is to use a least squares approach. The downfall, however, is that a nominal orbit is not known from either UCT individually, which causes the method to be poorly posed and to not converge in general. Our approach is a two step approximation scheme. Step 1 is to treat the angles and angular rates as determined. The uncertainty distribution is then modeled by a two-dimensional manifold in six-dimensional phase space. Step 2 is to incorporate the error in the co-space due to uncertainty in the angle and angular rate measurements. To do this, we will use our results in [94] and known symplectic constraints on phase volumes and subvolumes to develop a theory of handling *thick manifolds*, in particular, 2-d manifolds with a 4-d thickness.

Initial work in Step 1 has already been carried out, and has been presented at [95] and as an invited talk at [92]. Changing our perspective to one concerned with finite volumes instead of individual trajectories has already led to immediate progress in this classical problem. The principal tool that emerged due to our initial study is one we call Intersection Theory Analysis (ITA). We begin by mapping the admissible region associated with a particular UCT from the topocentric spherical or observation coordinates into Delaunay variables by following the schematic below. The Delaunay variables (L, l, G, g, H, h) are symplectic coordinate-momentum pairs (lower case for coordinates); in fact, they are the action-angle variables of the classical 2-Body Problem (2BP) [33, 42, 43, 50].

$$\left\{\begin{array}{c}\text{topocentric}\\ \text{spherical}\end{array}\right\} \rightarrow \left\{\begin{array}{c}\text{geocentric}\\ \text{cartesian}\end{array}\right\} \rightarrow \left\{\begin{array}{c}\text{orbital}\\ \text{elements}\end{array}\right\} \rightarrow \left\{\begin{array}{c}\text{Delaunay}\\ \text{Variables}\end{array}\right\}$$

In this way, the uncertainty distribution is represented by a 2 dimensional surface in 6 dimensional, symplectic Delaunay space. Five of the Delaunay variables are

constants of motion for the unperturbed 2BP. Meanwhile the angle l moves at a linear rate that depends only upon its conjugate momentum L. The symplectic projection of the uncertainty manifold onto the (G, g) and (H, h) planes are therefore static; however, since angles are modulo 2π, the projection onto the (L, l) plane exhibits a "shredding" phenomenon. All of the dynamics are encapsulated in this picture.

To illustrate how an estimate for a nominal orbit can be constructed from 2 UCT's using ITA, consider Fig. 1.2. Discretized admissible regions (2-d uncertainty manifolds) belonging to 2 separate UCT's measured from the same object, observed at different locations 70 hours apart, are mapped concurrently into Delaunay space and projected onto the symplectic planes. Each discrete point particle in the field is referred to as a *virtual debris* (VD) particle. The lightly colored VD field corresponds to the uncertainty manifold of the first measurement, propagated 70 hours into the future. The "shredding" is evident on the (L, l) plane, a signature of its dynamic evolution. The basic, underlying principle behind ITA is the systematic cutting off of VD particles in the nonoverlapping regions by performing simple set intersections in coupled 2-d planes. This reduces the uncertainty distribution to much less than the *apparent* overlap regions in Fig. 1.2, due to the fact that these are only *projections* of individual distributions. In fact, we show the method reduces this large, initial uncertainty distribution to a localized uncertainty *about a single point* [95]. It now makes sense to perform a least squares approximation, to determine the best fit orbit about that single, nominal point.

1.3.3 Future Directions in Space Situational Awareness

Our first objective is to perform further analysis on 2-d symplectic surfaces embedded in higher dimensional spaces, especially in regards to the SSA problem. We would develop the necessary computer algorithms required for the efficient implemen-

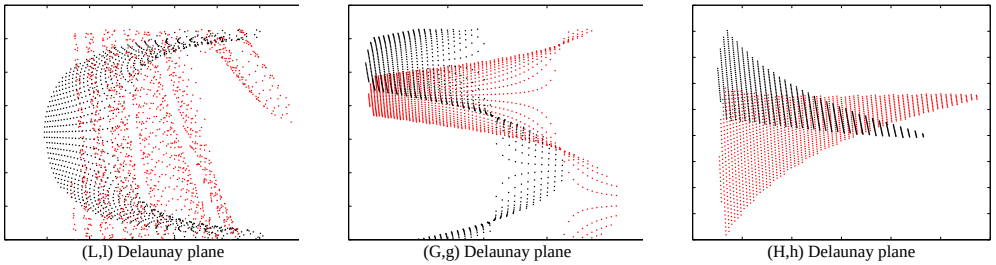

Figure 1.2: Projections of 2 separate uncertainty manifolds onto Delaunay planes

tation of ITA. In addition to establishing this basic computing package, we intend to analyze the metric structure of the Delaunay space. Utilizing the results of [94], a unique metric structure is induced in Delaunay space for the uncertainty manifold corresponding to each optical or radar measurement. What is the implication of these metric structures when multiple uncertainty distributions are mapped concurrently to the *same* Delauanay space with *separate* metric structures? Can these metric structures further aid the efficacy of ITA? We will then apply these results to the analysis of whether or not separate UCT's could belong to the same physical object, and whether or not UCT's could belong to maneuvering bodies.

A further goal is to research and develop the theory of *thick manifolds*. This study is motivated by our desire to understand the effect that local uncertainty in the remaining four co-dimensions has on ITA. Using these results, coupled with the theory on area expansion factors [94], we can construct quantitative probability maps on the surface projections in Delaunay space that will describe the contraction or expansion of the local uncertainty in the 4-d cotangent space of the surface manifold. We will also consider the complementary problem of understanding the constraints on 4-d manifolds with a 2-d thickness. This problem arises in the analysis of uncertainty distributions from radar measurements, where it is the range and range rates that are determined.

Another primary avenue of research will be the investigation of the interdependence that Gromov's NST, probability limits, and subvolume constraints have with each other and their effect on the analysis of uncertainty distributions related to the space debris tracking problem. By changing perspective to one of propagations of subvolumes, we have already made immediate progress on the classical problem of tracking space debris [95]. We wish to, in addition, understand the effect that the symplectic constraints have on uncertainty limits in the space debris problem. However, we will also search for other areas ripe for application, including the tracking of near earth asteroids and understanding the role of uncertainty in control and optimal control problems.

1.4 Nonholonomic Mechanics and Control

In the second portion of this thesis we will consider systems with nonholonomic constraints, and further how these constraints manifest themselves during the construction of variational fields and the use of variational principles. There is a rich geometric distinction that occurs at the differential level in these systems; the cause of much confusion and controversy during the advent of nonholonomic mechanics is attributed to this distinction. This distinction, which forms the basis of understanding variational principles in nonholonomic mechanics, has been discussed and analyzed from different viewpoints, some of the key references for which are Greenwood [56, 58], Hamel [60, 61, 62], Kane [68], Maggi [81], Neimark and Fufaev [105, 106], Papastavridis [113], Rumiantsev [119], Suslov [128, 129], and Voronets [131].

1.4.1 Nonholonomic Constraints

We will be considering systems with n degrees of freedom and m nonholonomic constraints. Throughout our conversation on nonholonomic mechanics we will take

the constraints to be linear, so that they are of the form:

$$\sum_{i=1}^{n} a_i^\sigma(q)\dot{q}^i = 0 \qquad\qquad \sigma = 1, \ldots, m \tag{1.2}$$

During this introduction, we will use letter types (i.e. Greek, roman, etc.) consistent with the summation conventions used throughout Chapter V and VI. However, since these conventions have not yet been introduced, no knowledge of them will be required to follow our present conversation. A more general set of linear constraints is given by:

$$\sum_{i=1}^{n} a_i^\sigma(q,t)\dot{q}^i + a_t^\sigma(q,t) = 0 \qquad\qquad \sigma = 1, \ldots, m \tag{1.3}$$

Nonholonomic constraints of the form (1.3) are called *rheonomic*, whereas constraints of the form (1.2) are called *scleronomic*. Rheonomic systems apply to virtually all practical problems in engineering and mechanics, so limiting our discussion to these types of constraints is not at all restrictive. In this thesis, however, we will focus exclusively on scleronomic constraints. We can do this without any loss in generality; if the system is rheonomic, one may consider its evolution on the extended phase space. By introducing an additional variable $q^0 = t$, $\dot{q}^0 = 1$, the system reduces to a scleronomic one.

The key feature that makes the constraints (1.2) *nonholonomic* as opposed to *holonomic* is their integrability. For an n-dimensional configuration manifold Q with m *holonomic* or *integrable* constraints, there exists an $(n-m)$-dimensional submanifold $S \subset Q$ on which the motion is constrained to. Hence, for the case of holonomic constraints, one can reduce the dimensionality of the system by m degrees of freedom. There exists a set of generalized coordinates $(q^1, \ldots, q^{n-m})$ which completely specify the configuration of this reduced system.

We will consider systems that are *nonholonomic*, so that the constraint coefficients

of (1.2) satisfy the inequality:

$$\frac{\partial a_i^\sigma}{\partial q^j} \neq \frac{\partial a_j^\sigma}{\partial q^i}$$

It is interesting to note that if one considers the underlying manifold to be the tangent bundle, TQ, and makes no distinction between coordinate and velocity, one sees that nonholonomic constraints on Q are holonomic constraints on TQ. Therefore there exists a $(2n-m)$-dimensional submanifold

$$\mathcal{S} \subset TQ$$

on which the system is confined. Note that there are no restrictions on the attainable positions, rather the restriction is on the types of motion that can be realized at any point. One might therefore conclude that a reduced set of coordinates can be erected as generalized coordinates for the submanifold $\mathcal{S}$:

$$(q^1, \ldots, q^n, u^{n-m+1}, \ldots, u^n)$$

This is indeed the case, and the $(n-m)$ generalized coordinates, $u^{n-m+1}, \ldots, u^n$, are referred to as a set of *quasi-velocities* for the system. To obtain the full dynamics, one augments these with m additional quasi-velocities defined by

$$u^\sigma = \sum_{i=1}^{k} a_i^\sigma \dot{q}^i \qquad\qquad \sigma = 1, \ldots, m$$

They are referred to as quasi-velocities, as they are, at least for the first m, not derivatives of any scalar function. In other words, there does <u>not</u> exist a set of $f^\sigma : Q \to \mathbb{R}$, $\sigma = 1, \ldots, m$, such that:

$$u^\sigma = \frac{df^\sigma}{dt} = \sum_{k=1}^{n} \frac{\partial f^\sigma}{\partial q^k} \frac{dq^k}{dt} \qquad\qquad \text{for any } \sigma \in \{1, \ldots, m\}$$

Whether such an f^I exists for the remaining $(n-m)$ quasi-velocities depends on one's choice of quasi-velocities. The dynamics are trivial for the first m quasi-velocities,

u^σ. The solution is $u^\sigma(t) = 0$. There are numerous ways of writing a reduced system that involves only $(2n - m)$ first order equations of motion. We will discuss some of these methods in Chapter V and extend them to optimal control problems later in Chapter VI. Notice that when one uses Lagrange multipliers, one would require a total of $(2n + m)$ first order differential equations of motion. This is due to the fact that, in addition for solving the equation of motion for each coordinate of TQ individually, one must also solve an additional m differential equations for the constraint forces that force the particle to remain on the submanifold $\mathcal{S} \subset TQ$.

1.4.2 Variational Principles

For unconstrained systems or systems with only holonomic (integrable) constraints, the correct dynamical motion satisfies *Hamilton's Principle*, which states that the true trajectory that passes between two arbitrary fixed points minimizes the action

$$I = \int L(q, \dot{q})\ dt \tag{1.4}$$

with respect to all kinematically admissible curves that go between those two fixed points. As it turns out, this principle fails for systems with nonholonomic constraints. Choosing Hamilton's Principle leads to a set of equations known as the vakonomic equations of motion, which are the *incorrect* dynamical equations of motion.

When taking variations of the above action, one obtains

$$\delta I = \int \left(\sum_{i=1}^{n} \frac{\delta L}{\delta q^i} \delta q^i \right)\ dt$$

where $\delta L / \delta q^i$ is the so-called "variational derivative." The n variations δq^i are not free, but rather must satisfy the constraints in some way. In order to enforce Hamilton's Principle, i.e. in order to find the unique *kinematically admissible* curve which minimizes (1.4), the constrainsts one places on $\{\delta q^i\}_{i=1}^n$ must somehow enforce the

condition that the varied path $q^i(t) + \delta q^i(t)$ is kinematically admissible. We will discuss how one achieves this in §5.6. As noted above, this procedure yields the *incorrect* dynamical equations of motion.

On the other hand, one can choose the variations $\{\delta q^i\}_{i=1}^n$ so that they satisfy the *Principle of Virtual Work*, which states that it is the *infinitesimal variations* δq^i themselves, and not the varied curves, which should be kinematically admissible. In other words, the Principle of Virtual Work states the constraints must satisfy the relations:

$$\sum_{i=1}^{n} a_i^\sigma(q)\delta q^i = 0 \qquad\qquad \text{for } \sigma = 1, \ldots, m$$

Choosing variations that satisfy the Principle of Virtual Work leads to the correct dynamical equations of motion. This is not Hamilton's Principle, as the class of curves one minimizes (1.4) with respect to are not kinematically admissible.

Our goal of Chapter V will be to understand the analytical and geometric implications of these constraints on the variations in terms of quasi-velocities. We will discuss the choices that arise when defining variations; each one can lead to the correct dynamical *or* vakonomic equations of motion; and then use this knowledge to extend the quasi-velocity techniques, such as Maggi's equations and the Boltzmann-hamel equation, to vakonomic motion, and then further in Chapter VI to kinematic and dynamic optimal control problems.

CHAPTER II

Symplectic Subvolumes and Expansion Factors

2.1 Introduction

In this chapter we will explore fundamental constraints on the evolution of certain symplectic subvolumes possessed by any Hamiltonian phase space. This research has direct application to optimal control and control of conservative mechanical systems. We relate geometric invariants of symplectic topology to computations that can easily be carried out with the state transition matrix of the flow map. We will show how certain symplectic subvolumes have a minimal obtainable volume; further if the subvolume dimension equals the phase space dimension, this constraint reduces to Liouville's Theorem. Finally we present a preferred basis that, for a given canonical transformation, has certain minimality properties with regards to the local volume expansion of phase space. In Chapter III we will show how a subvolume approach to dynamical systems is applicable to orbit determination and tracking of space debris.

2.1.1 Overview

The traditional approach for studying the dynamics and control of mechanical systems is to focus on individual trajectories and states in order to determine where they will go and where they can be forced to go. In reality, however, system states are never precisely known and can only be determined to exist within some set of

finite volume in the dynamical system's phase space. By treating such systems as a sum of individual trajectories, one loses the geometrical insight and deeper results offered by more wholistic approaches.

In this paper, we will be concerned with understanding fundamental constraints on the evolution of compact $2k$-dimensional symplectic sets that evolve along a nominal trajectory of the system. Different symplectic constraints arise on such sets, including conservation of the signed $2k$-volume projections on the coupled symplectic planes as well as the constraints implied from Gromov's Nonsqueezing Theorem (see Scheeres et al [123] for a discussion of these constraints in relation to orbit uncertainty evolution, and McDuff [98], Hofer and Zehnder [66], and Audin, et al [7] for recent results on symplectic topology). We will further present an additional constraint for a minimal obtainable volume that exists on certain classes of $2k$-dimensional symplectic sets and show how such a constraint leads to the local collapse of phase space along solution curves in Hamiltonian phase space. This collapse of phase space is fundamentally linked to the expansion of symplectic subvolumes. Finally, for any fixed final time, no matter how long into the future or how chaotic the system, we will produce a distinguished orthogonal symplectic basis that resists collapse. The basis may collapse as time evolves, but will return to being orthogonal at the final time. The uncertainty of any $2k$-dimensional differential volume initially parallel to the symplectic planes of this basis, even though it may increase dramatically during the course of its evolution, will always return to its initial uncertainty at the final time.

Since the resulting equations produced by applying Pontryagin's Maximum Principle to optimal control problems are Hamiltonian, the results we discuss here should provide geometric insight to the evolution and control of uncertainty distributions in such systems. This theory provides fundamental limits on dynamical orbits, and

hence if one provides a control it provides limits on the accuracy of the control in the face of uncertainty. It also provides limits on uncertainty propogation in optimal control systems. Moreover, the preferred minimal uncertainty basis we produce should have numerous benefits to the design of fixed finite time optimal control laws where precise state information is unknown. Applications of these results to control systems will be the topic of future research.

2.1.2 Outline

In Section 2, we introduce Hamiltonian systems and the state transition matrix (STM). The term STM is used in Linear Systems Theory to represent the matrix of the differential of the flow, i.e. for a Hamiltonian flow $\phi_t : Q \to Q$, the state transition matrix is the matrix of the linear mapping $d\phi : TQ \to TQ$. We show how classical identities on the Lagrange and Poisson brackets relate to constraints on the STM. Specifically, for any symplectic column of the STM Φ, the sum of the 2×2 symplectic subdeterminants must add up to unity.

In Section 3, we consider surfaces that can be explicitly parameterized by one of the symplectic planes. We derive area expansion factors from the parameterization plane to the surface, its image under the Hamiltonian phase flow, and the symplectic projections of its image. If the state of the system is somewhere on the initial surface (with equal a priori probability), we interpret these various expansion factors as a probability map that leads one to understand where the particle, after applying the Hamiltonian phase flow, is likely to be found. In section 5, we provide an application of this to asteroid tracking.

In Section 4, we present an expansion property of $2k$-volumes which is closely related to Wirtinger's Inequality. We show how this leads to the fact that subvolume expansions in the differential neighborhood of the Hamiltonian flow leads to a

collapsing property in systems which exhibit chaos.

In Section 5 we will discuss how these constraints on subvolume expansions, when considered with Liouville's Theorem, leads to the local collapse of the phase space around nominal trajectories. Interestingly, we will also show, that given a canonical transformation, there exists a preferred basis that resists collapse. In particular, the volume of a $2k$ subvolume chosen to be initially parallel to k of the symplectic planes will return to its initial value at this fixed final time.

2.2 Hamiltonian Systems

In this section we review the classical and modern theory of Hamiltonian systems. We further relate Poisson and Lagrange brackets and symplectic forms to sums of subdeterminants of the State Transition Matrix. For additional background on Hamiltonian systems and symplectic manifolds, see Arnold [6], Bloch [16], Greenwood [56], Marsden and Ratiu [83], and Silva [126].

2.2.1 The Classical Approach

Hamilton's Equations

In an N degree of freedom Hamiltonian System, one has a $2N$-dimensional phase space spanned by N generalized coordinates $\{q_i\}_1^N$ and their N conjugate momenta $\{p_i\}_1^N$. The conjugate pairings of coordinates and momenta, (q_i, p_i) form what are known as symplectic pairs. The dynamical equations of motion are derivable from a Hamiltonian function $H(q,p)$ and Hamilton's equations:

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i} \tag{2.1}$$

Solutions curves of the system (2.1) are called the Hamiltonian phase flow, and are denoted $\phi_t(q,p)$.

Matrix Formalism

Let $x = \langle p_1, q_1, p_2, q_2, \ldots p_N, q_N \rangle$ be the phase space position of the system, ordered by symplectic pairs. Define the matrix

$$J_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{2.2}$$

The symplectic matrix J is defined as the $2N \times 2N$ block-diagonal matrix with J_2's down the main diagonal:

$$J = \begin{pmatrix} J_2 & \mathbb{O}_2 & \cdots & \cdots & \mathbb{O}_2 \\ \mathbb{O}_2 & J_2 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & J_2 & \mathbb{O}_2 \\ \mathbb{O}_2 & \cdots & \cdots & \mathbb{O}_2 & J_2 \end{pmatrix}$$

Grouping the coordinates in symplectic pairs will be useful to us later as we will be looking at various "symplectic columns" of the STM. Given the above definitions, Hamilton's equations can be cast into the following matrix form:

$$\dot{x} = \frac{d}{dt}\phi_t(x_0) = J \cdot \frac{\partial H}{\partial x} \tag{2.3}$$

where $x = \phi_t(x_0)$. The right hand side is the so-called *symplectic gradient* of the Hamiltonian function.

Lagrange Brackets

Given a transformation

$$Q_i = Q_i(q, p) \qquad \text{and} \qquad P_i = P_i(q, p)$$

we may introduce the *Lagrange bracket* expression for the two variables (u, v), (which can take on any of the values $q_1, \ldots, q_n, p_1, \ldots, p_n$):

$$[u, v] = \sum_{i=1}^{N} \left(\frac{\partial P_i}{\partial u} \frac{\partial Q_i}{\partial v} - \frac{\partial Q_i}{\partial u} \frac{\partial P_i}{\partial v} \right), \tag{2.4}$$

The exactness conditions required for a canonical transformation can then be cast into the following equivalent conditions:

$$[q_j, q_k] = 0 \qquad [p_j, p_k] = 0 \qquad [p_j, q_k] = \delta_{jk} \tag{2.5}$$

The Hamiltonian phase flow $\phi_t(q, p)$ is a continuous one parameter family of canonical transformations.

Poisson Brackets

Alternatively, one can define the Poisson bracket as:

$$\{u, v\} = \sum_{i=1}^{N} \left(\frac{\partial u}{\partial p_i} \frac{\partial v}{\partial q_i} - \frac{\partial u}{\partial q_i} \frac{\partial v}{\partial p_i} \right) \tag{2.6}$$

where (u, v) can now be any of the variables $Q_1, \ldots, Q_n, P_1, \ldots, P_n$. The sufficient conditions for a canonical transformation can also be written as follows, in terms of the Poisson bracket:

$$\{Q_j, Q_k\} = 0 \qquad \{P_j, P_k\} = 0 \qquad \{P_j, Q_k\} = \delta_{jk}$$

2.2.2 The Geometric Approach

Symplectic Manifolds

A *symplectic structure* on an even-dimensional manifold M is a closed nondegenerate differential two-form ω on M:

$$d\omega = 0 \qquad \text{and} \qquad \forall \xi \neq 0,\ \exists \eta : \omega(\xi, \eta) \neq 0$$

The form ω is called the symplectic form and the pair (M, ω) is called a symplectic manifold.

On any symplectic manifold, there exists a vector space isomorphism between its cotangent and tangent bundles. At $x \in M$, we have

$$I_x : T_x^* M \to T_x M$$

defined by the symplectic form and the following relation. A vector $\xi \in T_x M$ is mapped to the one-form $I_x^{-1}(\xi)$ which acts on a vector $\eta \in T_x M$ as follows: $I_x^{-1}(\xi)(\eta) = \omega(\eta, \xi)$.

Hamiltonian Flows

Let H be a function $H : M \to \mathbb{R}$ which we will call the Hamiltonian. The associated Hamiltonian vector field on M is defined by IdH. The flow generated by the vector field IdH is the Hamilton phase flow ϕ_t. If $M = \mathbb{R}^{2N}$ with the standard symplectic form $\omega_0 = \sum_i dp_i \wedge dq_i$, we recover Hamilton's equations (2.1).

A transformation $\phi : M \to M$ is considered *symplectic* or *canonical* if it preserves the symplectic form, i.e., $\phi^* \omega = \omega$. The Hamiltonian phase flow ϕ_t is a one parameter family of canonical transformations.

Integral Invariants

A differential k-form α is an *integral invariant* of the map ϕ if the integrals of α on any k-chain σ is preserved as follows:

$$\int_{\phi(\sigma)} \alpha = \int_\sigma \alpha$$

The symplectic form ω is an integral invariant of the Hamiltonian flow.

To gain a physical intuition for what the integral invariant $\int \omega$ represents, consider now a closed parametrized surface $\psi(\sigma)$ in $\mathbb{R}^{2n} = (p, q)$, with a parametrization given by $\psi : \sigma \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^{2n}$, $\psi : (u, v) \longrightarrow (q(u, v), p(u, v))$. Then

$$\iint_{\psi(\sigma)} dp \wedge dq = \sum_{i=1}^{n} \iint_\sigma \det \left(\frac{\partial(p_i, q_i)}{\partial(u, v)} \right) \, du \, dv$$

$\iint_\sigma \det\left(\frac{\partial(p_i,q_i)}{\partial(u,v)}\right) dudv$ represents the oriented area of the projection of the surface $\psi(\sigma)$ on the i-th symplectic plane. By considering $\psi(\sigma)$ as an initial surface and applying the Hamiltonian phase flow, mapping the surface to $\phi_t(\psi(\sigma))$; we recognize, as a physical interpretation of the preservation of the symplectic form under canonical mappings, that the sum of the oriented areas of the projections onto the N symplectic planes is preserved.

2.2.3 The State Transition Matrix

Definition

If $\phi : M \to M, \phi(p,q) = (P,Q)$ is a canonical transformation, its differential

$$d\phi : T_{(p,q)}M \to T_{(P,Q)}M$$

is, when represented in matrix form, known as the state transition matrix (STM) Φ, a terminology adopted from Linear Systems Theory. If ϕ_t is the Hamiltonian phase flow and, for a given initial condition x_0, $x(t) = \phi_t(x_0)$, the matrix $\Phi(t)$ maps a variation of the initial state to its corresponding variation along the curve, i.e. $\phi_t(x_0 + \delta x_0) = x(t) + \Phi(t) \cdot \delta x_0 + O\left((\delta x_0)^2\right)$ as $\delta x_0 \to 0$.

Dynamics

If ϕ_t is the Hamiltonian phase flow, we have:

$$\frac{d}{dt}\phi_t(x_0) = J \cdot \frac{\partial H}{\partial x}(x)$$

If we perturb the initial conditions to $x_0 + \delta x_0$, we find:

$$\frac{d}{dt}\phi_t(x_0 + \delta x_0) = J \cdot \frac{\partial H}{\partial x}(x + \delta x)$$

By expanding this in a Taylor Series, one sees:

$$\frac{d}{dt}\Phi = J \cdot \frac{\partial^2 H}{\partial x^2} \cdot \Phi \tag{2.7}$$

This defines a system of $4N^2$ differential equations that can be integrated numerically, simultaneously along with the nominal solution curve $\phi_t(x_0)$.

Relation to Lagrange and Poisson Brackets

We will relate the Lagrange and Poisson Brackets to determinants of various submatrices of the STM. We will arrange the coordinates in a symplectic order, so that $x = \langle p_1, q_1, \ldots, p_N, q_N \rangle$ and $X = \langle P_1, Q_1, \ldots, P_n, Q_N \rangle$. In this fashion, the STM Φ is thought of as

$$\Phi = \frac{\partial X}{\partial x}$$

Define the following subdeterminants:

$$M_{ij} = \det \begin{pmatrix} \frac{\partial P_i}{\partial p_j} & \frac{\partial P_i}{\partial q_j} \\ \frac{\partial Q_i}{\partial p_j} & \frac{\partial Q_i}{\partial q_j} \end{pmatrix} = \frac{\partial P_i}{\partial p_j}\frac{\partial Q_i}{\partial q_j} - \frac{\partial P_i}{\partial q_j}\frac{\partial Q_i}{\partial p_j} \tag{2.8}$$

Hence, M_{ij} is the subdeterminant of the intersection of the ith *symplectic* row of the STM with its jth *symplectic* column.

It is easy to see the Lagrange and Poisson brackets are related to these subdeterminants as follows:

$$[p_j, q_j] = \sum_{i=1}^{N} M_{ij} = \delta_{ij} \tag{2.9}$$

$$(P_i, Q_i) = \sum_{j=1}^{N} M_{ij} = \delta_{ij} \tag{2.10}$$

It is well known that Liouville's Theorem manifests itself as a constraint on the State Transition Matrix by the requirement that its determinant must equal unity. These equations show us that, for a given canonical transformation, additional structure exists as constraints on certain combinations of determinants of 2×2 submatrices. For a fixed symplectic column, the sum of the determinants of the n different 2×2 submatrices (which stack up to form the symplectic column) must add to unity.

Similarly for a fixed symplectic row, the sum of the determinants of the n side-by-side 2×2 submatrices must also add to unity.

Relation to Differential Forms

Consider again a canonical transformation $\phi : M \to N$ that sends

$$\phi(p_1, q^1, \ldots, p_n, q^n) = \langle P_1, Q^1, \ldots, P_n, Q^n \rangle$$

The coordinates of the image of ϕ can be thought of as being functionally dependent upon the coordinates of the domain, so that $P_i = P_i(p_1, q^1, \ldots, p_n, q^n)$ and $Q^i = Q^i(p_1, q^1, \ldots, p_n, q^n)$. Furthermore the differential forms dP_i and dQ^i which form a basis for T^*N can be written in terms of the corresponding basis dp_i and dq^i for T^*M as follows:

$$\begin{aligned} dP_i &= \sum_{k=1}^{n} \left(\frac{\partial P_i}{\partial q^k} dq^k + \frac{\partial P_i}{\partial p_k} dp_k \right) \\ dQ^i &= \sum_{k=1}^{n} \left(\frac{\partial Q^i}{\partial q^k} dq^k + \frac{\partial Q^i}{\partial p_k} dp_k \right) \end{aligned}$$

The entries in the $(2i-1)$-th row of Φ are the components of the differential form dP_i with respect to the basis $\langle dp_1, dq^1, \ldots, dp_n, dq^n \rangle$. Similarly the entries of the $(2i)$-th row of Φ are the components of the differential form dQ^i with respect to the same basis.

We now give an alternate interpretation to the quantities M_{ij} defined in (2.8). By definition of the wedge product we have:

$$dP_i \wedge dQ^i \left(\frac{\partial}{\partial p_j}, \frac{\partial}{\partial q^j} \right) = \det \begin{pmatrix} dP_i\left(\frac{\partial}{\partial p_j}\right) & dP_i\left(\frac{\partial}{\partial q^j}\right) \\ dQ^i\left(\frac{\partial}{\partial p_j}\right) & dQ^i\left(\frac{\partial}{\partial q^j}\right) \end{pmatrix} = \det \begin{pmatrix} \frac{\partial P_i}{\partial p_j} & \frac{\partial P_i}{\partial q^j} \\ \frac{\partial Q^i}{\partial p_j} & \frac{\partial Q^i}{\partial q^j} \end{pmatrix} = M_{ij}$$

However since the transformation is symplectic, for fixed j it follows that:

$$\sum_{i=1}^{n} dP_i \wedge dQ^i \left(\frac{\partial}{\partial p_j}, \frac{\partial}{\partial q^j} \right) = \sum_{i=1}^{n} dp_i \wedge dq^i \left(\frac{\partial}{\partial p_j}, \frac{\partial}{\partial q^j} \right) = \sum_{i=1}^{n} \delta_{ij} = 1$$

This shows the relation (2.9). A similar argument can be constructed to prove (2.10).

2.3 Symplectic Surfaces

We begin the section by defining a few classes of distinguished surfaces that occur in symplectic spaces. We then give an efficient method for computation of area expansion factors and area projection factors in terms of the Gram determinant of certain matrices deriveable from the State Transition Matrix. For a classical exposition of surfaces, see Osserman [112].

2.3.1 Surface Classifications

We will begin by making the following fairly natural definitions.

Definition 1. *A **globally symplectic surface** is a two-dimensional submanifold of the phase space $\mathbb{R}^{2n}$ which admits a 1-to-1 mapping to at least one of the symplectic planes via the projection operator, i.e. it is a surface which can be parameterized in explicit form by one of its symplectic coordinate pairs.*

This characteristic is not an invariant one. It is possible, for example, for a lamina parallel to a symplectic plane to fold under some symplectic map, so that its image under the map is not 1-to-1 with any symplectic plane.

Another surface type we will consider is the following:

Definition 2. *A **parametrically symplectic surface** (or **parasymplectic surface**, for short) is a 2-dimensional submanifold of $\mathbb{R}^{2n}$ that admits a parameterization that is a symplectic one, i.e. one with a parameterization map that is canonical.*

Notice that the restriction of the symplectic form ω to the two-dimensional tangent bundle of any parasymplectic surface is itself a symplectic form on that submanifold.

The parasymplecticity of a surface is an invariant characteristic. Let σ be a lamina on the symplectic plane (u, v) which is the parameterization of the surface $\phi(\sigma) \subset \mathbb{R}^{2n}$,

where ϕ is the parameterization map. Let ϕ be a symplectomorphism, which exists if $\phi(\sigma)$ is a parasymplectic surface. Let $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be an arbitrary symplectomorphism which takes $\psi : \phi(\sigma) \to \psi(\phi(\sigma))$. Then $\psi(\phi(\sigma))$ is parasymplectic, with symplectic parameterization $\psi \circ \phi : \sigma \to \psi(\phi(\sigma))$.

We will consider $\pi_i : \mathbb{R}^{2n} \to \mathbb{R}^2$ the i-th *symplectic* projection operator, so that $\pi_i(\langle p_1, q_1, \ldots, p_N, q_N \rangle) = \langle p_i, q_i \rangle$.

2.3.2 Area Expansion Factors

Notation

We define the $2n \times 2$ matrix Π_κ as:

$$\Pi_\kappa = \begin{bmatrix} \mathbb{O}_2 & \mathbb{O}_2 & \cdots & \mathbb{O}_2 & \mathbb{I}_2 & \mathbb{O}_2 & \cdots & \mathbb{O}_2 \end{bmatrix}^T \tag{2.11}$$

where the $\mathbb{I}_2$ appears in the κ-th symplectic row. For any $2n \times 2n$ matrix A, the product $A \cdot \Pi_\kappa$ is the κ-th symplectic column of A; $\Pi_\kappa^T \cdot A$ is the κ-th symplectic row of A; and $\Pi_\kappa^T \cdot A \cdot \Pi_\lambda$ is the 2×2 intersection of the κ-th symplectic row with the λ-th symplectic column.

In this notation, the STM subdeterminant M_{ij}, defined previously, can be expressed as:

$$M_{ij} = \det(\Pi_i^T \cdot \Phi \cdot \Pi_j)$$

Globally Symplectic Surfaces

We will consider now a surface τ which is *globally symplectic* with respect to the j-th symplectic plane; i.e., the projection map $\pi_j : \tau \subset \mathbb{R}^{2n} \to \pi_j(\tau) \subset \mathbb{R}^2$ is one-to-one. We can parameterize τ by its symplectic shadow on the j-th symplectic plane. Now let the surface τ be mapped into the future by the Hamiltonian flow $\phi_t : \langle p_1, q_1, \ldots, p_N, q_N \rangle \to \langle P_1, Q_1, \ldots, P_N, Q_N \rangle$. We will now consider the projection

of $\phi_t(\tau)$ onto the i-th symplectic plane. For our analysis we will consider a differential area element $d\tau$ of τ (see Fig. 2.1).

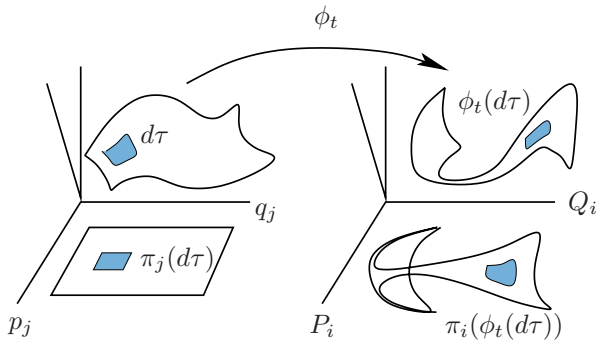

Figure 2.1: Area Expansion Factors

The surface τ is described by the parameterization $u = p_j$ and $v = q_j$ by

$$\langle p_1(u,v), q_1(u,v), \ldots, u, v, \ldots, p_N(u,v), q_N(u,v)\rangle$$

We define the matrix $L = \left[\frac{\partial x_i}{\partial u_j}\right]$ as follows:

$$L = \begin{bmatrix} \frac{\partial p_1}{\partial u} & \frac{\partial p_1}{\partial v} \\ \frac{\partial q_1}{\partial u} & \frac{\partial q_1}{\partial v} \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ \frac{\partial p_N}{\partial u} & \frac{\partial p_N}{\partial v} \\ \frac{\partial q_N}{\partial u} & \frac{\partial q_N}{\partial v} \end{bmatrix}$$

where the j-th symplectic row is equated to $\mathbb{I}_2$. L is the Jacobian matrix of the parameterization map that takes $\pi_j(d\tau) \to d\tau$. The metric for the surface τ in terms of variations in the $u - v$ plane is given by:

$$[g_{ij}] = \begin{bmatrix} \frac{\partial \mathbf{x}}{\partial u} \cdot \frac{\partial \mathbf{x}}{\partial u} & \frac{\partial \mathbf{x}}{\partial u} \cdot \frac{\partial \mathbf{x}}{\partial v} \\ \frac{\partial \mathbf{x}}{\partial v} \cdot \frac{\partial \mathbf{x}}{\partial u} & \frac{\partial \mathbf{x}}{\partial v} \cdot \frac{\partial \mathbf{x}}{\partial v} \end{bmatrix} = L^T \cdot L$$

where $\mathbf{x} = \langle p_1(u,v), \ldots, q_N(u,v) \rangle$, as usual. Hence the metric determinant is:

$$g = \det(g_{ij}) = \mathfrak{G}(L) = \det(L^T \cdot L)$$

where $\mathfrak{G}(L)$ is the Gram determinant of the matrix L, defined by this equation.

The surface area of τ is thus given by:

$$S.A.(\tau) = \iint_{\pi_j(\tau)} \sqrt{g} du dv$$

A simple application of this result gives us the total physical area of the area element $d\tau$ in terms of the area of its projection:

$$\frac{\mathcal{A}(d\tau)}{\mathcal{A}(\pi_j(d\tau))} = \sqrt{\mathfrak{G}(L)}$$

Similarly we have:

$$\frac{\mathcal{A}(\phi_t(d\tau))}{\mathcal{A}(\pi_j(d\tau))} = \sqrt{\mathfrak{G}(\Phi \cdot L)}$$

where Φ is the STM associated with ϕ_t and x_0.

The area of the projection $\pi_i(\phi_t(d\tau))$ is given by the Jacobian:

$$\det\left(\frac{\partial(P_i, Q_i)}{\partial(u,v)}\right) = \sum_{\kappa=1}^{2N} \sum_{\lambda=1}^{2N} \left(\frac{\partial P_i}{\partial x_\kappa}\frac{\partial Q_i}{\partial x_\lambda} - \frac{\partial Q_i}{\partial x_\kappa}\frac{\partial P_i}{\partial x_\lambda}\right) \frac{\partial x_\kappa}{\partial u}\frac{\partial x_\lambda}{\partial v}$$

which can be represented more concisely as:

$$\frac{\mathcal{A}(\pi_i(\phi_t(d\tau)))}{\mathcal{A}(\pi_j(d\tau))} = \det(\Pi_i^T \cdot \Phi \cdot L)$$

Parasymplectic Surfaces

We will now consider the case where τ is parallel to the j-th symplectic plane. In this case, its parameterization map is a symplectic one, and thus it is a parasymplectic surface. All of the above results hold, but the matrix L reduces to the simpler form $L = \Pi_j$, which gives us the following:

$$\mathcal{A}(d\tau) = \mathcal{A}(\pi_j(d\tau))$$

$$\frac{\mathcal{A}(\phi_t(d\tau))}{\mathcal{A}(\pi_j(d\tau))} = \sqrt{\mathfrak{G}(\Phi \cdot \Pi_j)}$$

and, most notably

$$\frac{\mathcal{A}(\pi_i(\phi_t(d\tau)))}{\mathcal{A}(\pi_j(d\tau))} = \det(\Pi_i^T \cdot \Phi \cdot \Pi_j) = M_{ij} \tag{2.12}$$

Preservation of the sum of the oriented symplectic area projections thus gives us the following constraint on the STM:

$$\sum_{i=1}^{N} M_{ij} = [p_j, q_j] = 1 \tag{2.13}$$

Application to Orbit Uncertainty Distributions

Suppose we know that a system can be found anywhere on the surface τ with equal a priori probability. The surface is now mapped into the future by the Hamiltonian flow ϕ_t and we wish to determine where the particle is most likely to be on the $P_i - Q_i$ plane.

We begin by discretizing the $(u = p_j) - (v = q_j)$ plane, each area element with area $\Delta u \Delta v$. Summation will be assumed to be over each district. The probability that the particle is in $d\tau$ is given by:

$$\mathcal{P}(d\tau) = \frac{\mathcal{A}(d\tau)}{\mathcal{A}(\tau)} \approx \frac{\sqrt{\mathfrak{G}(L)}\Delta u \Delta v}{\sum \sqrt{\mathfrak{G}(L)}\Delta u \Delta v} = \frac{\sqrt{\mathfrak{G}(L)}}{\sum \sqrt{\mathfrak{G}(L)}}$$

From (2.12) we see:

$$\mathcal{A}(\pi_i(\phi_t(d\tau))) \approx \det(\Pi_i^T \cdot \Phi \cdot L)\Delta u \Delta v$$

so that the area probability density at $\langle P_i, Q_i \rangle$ is:

$$\sigma \approx \frac{\sqrt{\mathfrak{G}(L)}}{|\det(\Pi_i^T \cdot \Phi \cdot L)| \Delta u \Delta v \sum \sqrt{\mathfrak{G}(L)}}$$

This approach may be helpful in asteroid tracking, where angular and angular rate information is precisely known, but there is initial uncertainty in the $r,\ \dot{r}$ distribution. This problem is treated in Milani, et al. [99].

2.4 Integral Analysis of Symplectic Subvolumes

In this section we will discuss the difference between two fundamental integral invariants defined for an arbitrary $2k$-dimensional subvolume of our $2n$-dimensional phase space, for $k = 1, \ldots, n$. The first is the well-known integral invariant of Poincaré-Cartan. The second integral invariant is closely related, and is tantamount to a global version of Wirtinger's Inequality for lower dimensional subvolumes of phase space.

2.4.1 Signed and Unsigned Integrals of Differential Forms

We will begin our discussion on integral invariants with a brief discussion of the theory of integration of differential forms. Let $\Sigma \subset \mathbb{R}^{2n}$ be a $2k$-dimensional submanifold, parameterized by $\phi : (\sigma \subset \mathbb{R}^{2k}) \to \Sigma$. Let α be a $2k$-form on $\mathbb{R}^{2n}$. Then we define

$$\int_\Sigma \alpha = \int_\sigma \phi^* \alpha$$

For a function $f(x_1, \ldots, x_{2k})$, the pullback of α can be expressed in the following form:

$$\phi^* \alpha = f(x_1, \ldots, x_{2k}) dx^1 \wedge \cdots \wedge dx^{2k}$$

where we take $(x_1, \ldots, x_{2k})$ to be coordinates in $\mathbb{R}^{2k}$. Then the integral of $\phi^*\alpha$ over σ reduces to the ordinary euclidean integral:

$$\int_\Sigma \alpha = \int_\sigma \phi^* \alpha = \int_\sigma f(x_1, \ldots, x_{2k}) dx^1 \cdots dx^{2k}$$

We would like to introduce a further definition as follows. We define the unsigned integral of α over Σ to be:

$$\int_\Sigma |\alpha| = \int_\sigma |f(x_1, \ldots, x_{2k})| dx^1 \cdots dx^{2k}$$

where f has been defined above.

A more rigorous definition of the above integrals must involve a partition of unity, but simplicity has been choosen over rigor so as to illustrate the spirit of the definitions.

2.4.2 The Integral Invariants of Poincaré-Cartan

Consider the standard symplectic form

$$\omega = \sum_{i=1}^{n} p_i \wedge q_i$$

and its k-th exterior product:

$$\frac{1}{k!}\omega^k = \sum_{1 \le i_1 < \cdots < i_k \le n} dp_{i_1} \wedge dq_{i_1} \wedge \cdots \wedge dp_{i_k} \wedge dq_{i_k}$$

Consider a set of $2k$ vectors $(X^1, \ldots, X^{2k})$ in $\mathbb{R}^{2n}$. Then

$$\frac{1}{k!}\omega^k(X^1, \ldots, X^{2k})$$

represents the sum of the oriented $2k$-volume projections of the parallelepiped spanned by $X^1, \ldots, X^{2k}$ on the symplectic "$2k$-planes."

ω^k is known as the integral invariant of Poincaré-Cartan. Given an arbitrary $2k$-dimensional phase volume Ω (in a $2n$ dimensional space) and the Hamiltonian phase flow ϕ_t, we have:

$$\frac{1}{k!}\int_{\Omega} \omega^k = \frac{1}{k!}\int_{\phi_t(\Omega)} \omega^k$$

so that the sum of the oriented $2k$-volume projections on each symplectic "$2k$ plane" is conserved.

2.4.3 The Wirtinger-Type Integral Invariants and Volume

Given a set of vectors $X_1, \ldots, X_{2k} \in \mathbb{R}^{2n}$, Wirtinger's Inequality states that the "2k" volume of the parallelepiped spanned by these vectors is bounded by

$$\frac{1}{k!}|\omega^k(X_1, \ldots, X_{2k})| \le \mathrm{Vol}_{2k}(X_1, \ldots, X_{2k}) \tag{2.14}$$

We make the following two observations. First, it is clear that $|\omega^k|$ is an integral invariant of the Hamiltonian flow ϕ_t, so that, given any $2k$-volume Ω, we have:

$$\frac{1}{k!}\int_{\Omega}|\omega^k| = \frac{1}{k!}\int_{\phi_t(\Omega)}|\omega^k|$$

Moreover,

$$\frac{1}{k!}\int_{\Omega}|\omega^k| \leq \mathrm{Vol}_{2k}(\Omega)$$

This integral invariant is a lower bound on the set of possible $(2k)$ volumes that the body Ω may attain. For the case $k = n$, the volume of Ω is a constant which equals this invariant quantity (Liouville's Theorem).

2.4.4 Parasymplectic $2k$-Volumes

In direct analogy with our discussion of parasymplectic surfaces, we define parasymplectic $2k$-volumes as follows:

Definition 3. *A **parasymplectic** $2k$**-volume**, or parametrically symplectic $2k$-volume, is one that admits a symplectic paramteterization map.*

An example of a parasympletic volume is the following. Take any $2k$-dimensional volume Ω that is parallel to k of the symplectic planes, i.e. a region defined by:

$$\Omega = \langle p_1, q_1, \ldots, p_k, q_k, c_{k+1}, d_{k+1}, \ldots, c_n, d_n\rangle$$

where the variables $p_1, q_1, \ldots, p_k, q_k$ vary over some region of $\mathbb{R}^{2k}$ and c_{k+1}, d_{k+1}, ..., c_n, d_n are constants. Now let ϕ_t be the Hamiltonian phase flow. The $2k$-phase volumes $\phi_t(\Omega) \subset \mathbb{R}^{2n}$ are a one-parameter family of parasymplectic $2k$-volumes. This follows since $\phi_t(\Omega)$ can be parameterized by ϕ_t and the projection operator onto the first k symplectic planes.

Theorem 4. ***[Volume Expansion of Parasymplectic** $2k$**-Volumes]** *Let $\Omega \subset \mathbb{R}^{2k}$ be the parameterization of a volume $\phi(\Omega) \subset \mathbb{R}^{2n}$ in a symplectic phase space*

whose parameterization map ϕ is a symplectic one. Then

$$\mathrm{Vol}_{2k}(\Omega) \leq \mathrm{Vol}_{2k}(\phi(\Omega))$$

Corollary 5. *The $2k$-volume of any parasymplectic $2k$-volume is at least as large as the volume of its symplectic parameterization.*

Proof. To prove our theorem, we only need prove it for a differential volume element. The generalization follows via a simple integration argument.

Let $\phi : \mathbb{R}^{2k} \to \mathbb{R}^{2n}$ be a symplectic parameterization of a surface. Let

$$\left\{ \frac{\partial}{\partial u_1}, \frac{\partial}{\partial v_1}, \ldots, \frac{\partial}{\partial u_k}, \frac{\partial}{\partial v_k} \right\}$$

be an orthonormal basis of $\mathbb{R}^{2k}$ and let

$$X_i = \phi_* \left(\frac{\partial}{\partial u_i} \right) \qquad \text{and} \qquad Y_i = \phi_* \left(\frac{\partial}{\partial v_i} \right)$$

be the push forwards of the basis vectors in the parameterization space. Applying Wirtinger's Inequality (2.14), we have that

$$\begin{aligned} & \frac{1}{k!}\omega^k \left(\phi_* \left(\frac{\partial}{\partial u_1} \right), \cdots, \phi_* \left(\frac{\partial}{\partial v_k} \right) \right) \\ &= \frac{1}{k!}\phi^*(\omega^k) \left(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial v_k} \right) \\ &= \frac{1}{k!}\omega^k \left(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial v_k} \right) \\ &= du_1 \wedge \cdots \wedge dv_k \left(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial v_k} \right) \\ &= 1 \leq \mathrm{Vol}_{2k}(X_1, \ldots Y_k) \end{aligned}$$

Hence any $2k$-dimensional volume measure must be nondecreasing under such a map.

□

2.4.5 The Volume Expansion Factor

In this section we provide a practical approach to determining the volume and the integral invariants of $2k$ subvolumes. We will consider the volume $\Omega \subset \mathbb{R}^{2k}$ to be the parameterization volume of a $2k$-volume in the symplectic space $\mathbb{R}^{2n}$, with parameterization map

$$\phi : (\Omega \subset \mathbb{R}^{2k}) \rightarrow (\phi(\Omega) \subset \mathbb{R}^{2n})$$

The Jacobian of the parameterization is the $2n \times 2k$ matrix given by

$$L = d\phi$$

We will be interested in computing the total volume of Ω, the sum of its oriented symplectic projections (i.e. integral invariant of Poincaré-Cartan), and its minimum obtainable volume (i.e. the integral invariant of the Wirtinger type).

In terms of the parameterization coordinates $\langle u_1, v_1, \ldots, u_k, v_k \rangle \in \mathbb{R}^{2k}$, the following metric is induced on the surface:

$$\mathbf{g} = L^T \cdot L$$

We thus recognize the determinant of the metric $g = \det \mathbf{g}$ as the Gram determinant of the Jacobian matrix L:

$$g = \mathfrak{G}(L) = \det(L^T \cdot L)$$

so that

$$\mathrm{Vol}_{2k}(\phi(\Omega)) = \int_{\Omega} \sqrt{|g|} d\Omega$$

In practical terms, the Gramian of the Jacobian can be identified with the volume expansion factor in the following way:

$$\nu_{2k}(d\Omega; \phi) = \frac{\mathrm{Vol}_{2k}(\phi(d\Omega))}{\mathrm{Vol}_{2k}(d\Omega)} = \sqrt{\mathfrak{G}(L)}$$

where $\nu_{2k}(d\Omega; \phi)$ is the local $2k$-volume expansion factor of $d\Omega$ under the mapping ϕ.

2.5 Differential Analysis of Symplectic Subvolumes

In the previous section we saw how various symplectic constraints manifest themselves in macroscopic subvolumes of the phase space. In this section we will discuss the implication of these constraints on the evolution of the local structure of phase space under the Hamiltonian phase flow along a nominal solution trajectory.

2.5.1 Volume Expansion and the Local Collapse of Phase Space

The setting for this subsection will be the evolution of an infinitesimal neighborhood surrounding a Hamiltonian trajectory through phase space. Consider the Hamiltonian flow:

$$\phi_t : \langle p_1, q_1, \ldots, p_n, q_n \rangle \rightarrow \langle P_1, Q_1, \ldots, P_n, Q_n \rangle$$

Now consider a differential $2n$-"cube" Ω situated at the initial point

$$x = \langle p_1, q_1, \ldots p_n, q_n \rangle,$$

whose faces are parallel with the symplectic planes. Let $\Upsilon \subset \Omega$ be a $2k$-dimensional subset that is parallel with k of the symplectic planes, and let $\Upsilon' \subset \Omega$ be a $2n-2k$ dimensional subset that is parallel with the remaining $n-k$ symplectic planes, such that Ω is a direct sum:

$$\Omega = \Upsilon \oplus \Upsilon'$$

and, therefore

$$\mathrm{Vol}_{2n}(\Omega) = \mathrm{Vol}_{2k}(\Upsilon) \cdot \mathrm{Vol}_{2n-2k}(\Upsilon')$$

The Hamiltonian flow now takes $\phi_t : x \to X$ along with its differential neighborhood. We define

$$\begin{aligned} \overline{\Omega} &= d\phi_t(\Omega) \\ \overline{\Upsilon} &= d\phi_t(\Upsilon) \\ \overline{\Upsilon'} &= d\phi_t(\Upsilon') \end{aligned}$$

We now define the angle β via the relation:

$$\mathrm{Vol}_{2n}(\overline{\Omega}) = \mathrm{Vol}_{2k}(\overline{\Upsilon}) \cdot \mathrm{Vol}_{2n-2k}(\overline{\Upsilon'}) \sin\beta$$

so that β is the angle between the subspaces $\overline{\Upsilon}$ and $\overline{\Upsilon'}$. By Liouville's Theorem, we have:

$$\mathrm{Vol}_{2n}(\overline{\Omega}) = \mathrm{Vol}_{2n}(\Omega)$$

so that:

$$1 = \nu_{2k}(\Upsilon; \phi_t)\nu_{2n-2k}(\Upsilon'; \phi_t) \sin\beta$$

But by Nonsqueezing of parasymplectic $2k$-volumes, we have:

$$\begin{aligned} \nu_{2k}(\Upsilon; \phi_t) &\geq 1 \\ \nu_{2n-2k}(\Upsilon'; \phi_t) &\geq 1 \end{aligned}$$

We conclude that the greater the volume expansion of these lower dimensional differential "slices" Υ and Υ', the greater the inward collapse of their respective subspaces towards each other. In chaos theory, where $\mathrm{Vol}_{2k}(\phi_t(\Upsilon))$ is growing at an exponential rate, we see that β is correspondingly decaying at an exponential rate. Thus chaos (for Hamiltonian systems) necessarily implies the collapse of the phase space along certain directions.

2.5.2 The Symplectic Eigenskeleton

In this section we expose a special basis associated with any linear(ized) symplectomorphism that resists collapse. We will ominiously refer to the symplectomorphism as ϕ_t, keeping the dynamical setting (i.e. ϕ_t is the phase flow of a Hamiltonian system) in mind.

Theorem 6 (The Symplectic Eigenskeleton). *Consider a symplectomorphism $\phi_t : M \to M$ that takes the initial point x_0 to $\phi_t(x_0) = x$. Let $\Phi : T_{x_0}M \to T_xM$ be the State Transition Matrix (STM) of the mapping. Define the **characteristic matrix** of the transformation as $\Psi = \Phi^T \cdot \Phi$ and let $\{\xi_1, \eta_1, \ldots, \xi_N, \eta_N\}$ be the orthonormal eigenbasis of Ψ. Then the following are true:*

1. *There is an interdependency amongst the vectors of Ψ. The eigenvectors occur in pairs, where the $\{\eta_i\}_{i=1}^N$ can be taken to be*

$$\eta_i = J \cdot \xi_i$$

   where the associated eigenvalue of η_i is λ_i^{-1} if λ_i is the eigenvalue associated with ξ_i.

2. *The linear transformation T that takes the standard basis to the eigenbasis of Ψ,*

$$T : \{\hat{p}_1, \hat{q}_1, \ldots, \hat{p}_N, \hat{q}_N\} \to \{\xi_1, \eta_1, \ldots, \xi_N, \eta_N\},$$

   is symplectic. Moreover, the couples $\{\xi_i, \eta_i\}_{i=1}^N$ make symplectic pairs.

3. *The vectors $\{\Phi \cdot \xi_1, \Phi \cdot \eta_1, \ldots, \Phi \cdot \xi_N, \Phi \cdot \eta_N\}$ are orthogonal. Moreover,*

$$||\Phi \cdot \xi_i|| = \sqrt{|\lambda_i|} \qquad \textit{and} \qquad ||\Phi \cdot \eta_i|| = \sqrt{|\lambda_i^{-1}|}$$

4. *If a $2k$-dimensional symplectic subvolume Υ is initially parallel to k of the eigenskelton planes, then the linearized transformation Φ preserves its volume, i.e $\mathrm{Vol}_{2k}(\Upsilon) = \mathrm{Vol}_{2k}(\Phi(\Upsilon))$.*

We call the symplectic eigenbasis of the matrix $\Psi = \Phi^T \cdot \Phi$ the ***symplectic eigenskeleton*** *of the transformation ϕ_t, as it is a property structure of the transformation which resists collapse over a discrete time t.*

Proof. 1. Consider the i-th eigenvector ξ_i of Ψ with eigenvalue λ_i:

$$\Psi \cdot \xi_i = \lambda_i \xi_i$$

Taking the transpose of this equation, right-multiplying by $J \cdot \Psi$, and then recognizing the identity $\Psi^T \cdot J \cdot \Psi = J$, we see that

$$\xi_i^T \cdot J = \lambda_i \xi_i^T \cdot J \cdot \Psi$$

Taking the transpose once more and multiplying by -1 (whilst noting $\Psi^T = \Psi$ and $J^T = -J$) we have

$$\Psi \cdot (J \cdot \xi_i) = \frac{1}{\lambda_i} J \cdot \xi_i$$

Hence, the vectors $\eta_i = J \cdot \xi_i$ are also eigenvectors of Ψ, with eigenvalues λ_i^{-1}.

2. We define the matrices Ξ and N as follows:

$$\Xi = \begin{bmatrix} | & & | \\ \xi_1 & \cdots & \xi_N \\ | & & | \end{bmatrix} \text{ and } \mathrm{N} = \begin{bmatrix} | & & | \\ \eta_1 & \cdots & \eta_N \\ | & & | \end{bmatrix}$$

where $N = J \cdot \Xi$ from Part 1. The transformation matrix T can be represented as:

$$T = \begin{bmatrix} | & | \\ \Xi & N \\ | & | \end{bmatrix}$$

We have temporarily reordered our representation of the basis, so that

$$J = \begin{bmatrix} \mathbb{O}_N & -\mathbb{I}_N \\ \mathbb{I}_N & \mathbb{O}_N \end{bmatrix}$$

Noting again that $N = J \cdot \Xi$, one easily sees:

$$T^T \cdot J \cdot T = \begin{bmatrix} \Xi^T \cdot N & -\Xi^T \cdot \Xi \\ N^T \cdot N & -N^T \cdot \Xi \end{bmatrix}$$

Due to the orthonormality of the eigenvectors (i.e. $\Xi^T \cdot N = N^T \cdot \Xi = \mathbb{O}_N$ and $\Xi^T \cdot \Xi = N^T \cdot N = \mathbb{I}_N$), this expression reduces to:

$$T^T \cdot J \cdot T = J$$

and hence the matrix T is symplectic.

3. Renaming the eigenvectors of Ψ as $\{v_i\}_{i=1}^{2N} = \{\xi_i, \eta_i\}_{i=1}^{N}$, we have from the orthonormality of the eigenbasis of Ψ:

$$v_i \cdot v_j = \delta_{ij}$$

But

$$(\Phi \cdot v_i) \cdot (\Phi \cdot v_j) = v_i^T \cdot \Phi^T \cdot \Phi \cdot v_j = \lambda_j \delta_{ij}$$

Hence the vectors $\{\Phi \cdot v_i\}_{i=1}^{2N}$ are also orthogonal. Moreover:

$$||\Phi \cdot \xi_i|| = \sqrt{\xi_i^T \cdot \Phi^T \cdot \Phi \cdot \xi} = \sqrt{\lambda_i}$$

Similarly

$$||\Phi \cdot \eta_i|| = \frac{1}{\sqrt{\lambda_i}}$$

4. If Υ is a 2-dimensional area element spanned by ξ_1 and η_1, then the area expansion is

$$\frac{\mathrm{Vol}_2(\Phi(\Upsilon))}{\mathrm{Vol}_2(\Upsilon)} = (\Phi \cdot \xi_1) \cdot (\Phi \cdot \eta_1) = \sqrt{\lambda_1} \frac{1}{\sqrt{\lambda_1}} = 1$$

But since this is true of any area element initially parallel to one of the symplectic eigenskeleton planes, and the symplectic eigenskeleton $2k$ volumes are simply direct sums of these area elements, the result follows.

□

2.5.3 Time Evolution of the Symplectic Eigenskeleton

For a fixed initial condition x_0 the Hamiltonian phase flow maps x_0 to $x(t) = \phi_t(x_0)$. As we have seen, the State Transition Matrix $\Phi(t;t_0) = [d\phi_t]$ is determined by the following set of differential equations

$$\frac{d}{dt}\Phi(t;t_0) = J \cdot \frac{\partial^2 H}{\partial x^2} \cdot \Phi$$

In the previous section we defined the characteristic matrix $\Psi = \Phi^T \cdot \Phi$ for a fixed symplectomorphism. By fixing the initial condition $x(0) = x_0$ one can similarly define a characteristic matrix that evolves in time

$$\Psi(t;t_0) = \Phi(t;t_0)^T \cdot \Phi(t;t_0) \tag{2.15}$$

One can easily show that the time-dependent characteristic matrix is governed by the differential equations:

$$\frac{d}{dt}\Psi(t;t_0) = \Phi(t;t_0)^T \cdot \left[J, \frac{\partial^2 H}{\partial x^2}\right] \cdot \Phi(t;t_0)$$

where the bracket $[\cdot,\cdot]$ is the usual matrix commutator bracket. In practice however one does not need to integrate these equations; rather, once the state transition matrix is computed, one can use the algebraic relation (2.15) in determining $\Psi(t;t_0)$. When there is no danger of ambiguity, we will sometimes refer to $\Phi(t;t_0)$ and $\Psi(t;t_0)$ by the shorthand $\Phi(t)$ and $\Psi(t)$, respectively.

From the time-dependent characteristic matrix $\Psi(t)$ one can define a time-dependent eigenskeleton, the set of vectors $\{\xi_i(t), \eta_i(t)\}_{i=1}^n$, which form an orthonormal time-dependent basis of $T_{x_0}Q$, and the time-dependent characteristic eigenvalues $\{\lambda_i(t)\}_{i=1}^n$. Recall $\lambda_i(t)$ is the eigenvalue of $\Psi(t)$ associated with the eigenvector $\xi_i(t)$ and $\lambda_i^{-1}(t)$ is the eigenvalue of $\Psi(t)$ associated with the eigenvector $\eta_i(t) = J \cdot \xi_i(t)$.

2.5.4 Lyapunov Characteristic Exponents

The Lyapunov Characteristic Exponents (LCE's) of a dynamical system play an important role in the study of chaos and ergodic theory. Some of their basic properties were laid out in Oseledec [111]. They are also discussed in the texts of Arnold [5], Brin and Stuck [23] (from a topological viewpoint), and Lichtenberg and Lieberman [80]. The theory and computation of LCE's for Hamiltonian systems is studied in Benettin et al. [10, 11]. LCE's also have a fundamental relation to the Kolmogorov entorpy of the system, see Benettin et al. [12] and Froeschle and Froeschle [51]. LCE's have even been used to study the chaoticity of asteroidal motion, as was done in Froeschle et al. [52].

In this section we will define the Lyapunov Characteristic Exponents of a system and relate them to the eigenvalues of the characteristic matrix Ψ introduced previously. A similar approach is considered in Dieci et al. [45].

Definition 7. *For a fixed initial condition $x(0) = x_0$, tangent vector $v \in T_{x_0}Q$, and Hamiltonian phase flow $\phi_t : Q \to Q$, the* ***finite-time Lyapunov Characteristic Exponent*** *(f.t.LCE) associated with v is:*

$$\chi(v,t) = \frac{1}{t} \ln \frac{||d\phi_t \cdot v||}{||v||}.$$

The ***Lyapunov Characteristic Exponent*** *(LCE) associated with the vector v is then defined as:*

$$\chi(v) = \limsup_{t \to \infty} \chi(v,t)$$

LCE's enjoy a variety of properties. In particular, it is easy to show that $\chi(v)$ is independent of the magnitude of v. Moreover, $\chi(v)$ can only take on $2n$ distinct values, which we'll call σ_i and further order these values such that $\sigma_1 \geq \cdots \geq \sigma_{2n}$. It is further known that for Hamiltonian systems these values occur in positive/negative

pairs and that at least two (σ_n and σ_{n+1}) must vanish. Oseledec [111] further showed the existence of a basis $\hat{e}_1, \ldots, \hat{e}_{2n}$ such that $\chi(e_i) = \sigma_i$. With respect to this basis suppose we have $v = \sum_{i=L}^{2n} x^i \hat{e}_i$ so that x^L is the first non-zero component of v. Then $\chi(v) = \chi(\hat{e}_L) = \sigma_L$.

For the remainder of our discussion of LCE's, we will take the vector v to be a unit vector, so that $||v||$ can be omitted. The vector v may be represented with respect to the time-varying symplectic eigenskeleton as follows:

$$v = x^i(t)\xi_i(t) + y^i(t)\eta_i(t)$$

Notice that the time evolution of the components of v is due to the time-varying nature of the basis. The vector v itself is fixed. We now have:

$$\begin{aligned} ||d\phi_t \cdot v|| &= \sqrt{|v^T \cdot \Phi^T(t) \cdot \Phi(t) \cdot v|} \\ &= \sqrt{(x^i(t)\xi_i^T(t) + y^i(t)\eta_i^T(t)) \cdot \Psi(t) \cdot (x^i(t)\xi_i(t) + y^i(t)\xi_i(t))} \\ &= \sqrt{(x^i(t))^2\lambda_i(t) + (y^i(t))^2\lambda_i^{-1}(t)} \end{aligned}$$

where we have used some of the results of Theorem 6. The finite-time LCE associated with the unit vector v is therefore:

$$\chi(v,t) = \frac{1}{2t} \ln \left(\sum_{i=1}^{n} (x^i(t))^2\lambda_i(t) + (y^i(t))^2\mu_i(t) \right) \tag{2.16}$$

where $\mu_i(t) = \lambda_i(t)^{-1}$ are the remaining n eigenvalues of $\Psi(t)$. Since v is a unit vector, it follows $|x_i(t)| \leq 1$ and $|y_i(t)| \leq 1$. Consider the following $2n$ quantities:

$$\varsigma_i(t) = \frac{1}{2t} \ln \lambda_i(t) \qquad \text{and} \qquad \varsigma_{i+n}(t) = -\varsigma_i(t) \tag{2.17}$$

and their limiting values:

$$\varsigma_i = \limsup_{t\to\infty} \frac{1}{2t} \ln \lambda_i(t) \qquad \text{and} \qquad \varsigma_{i+n} = -\varsigma_i \tag{2.18}$$

Suppose now that the eigenbasis of $\Psi(t)$ has a limit as $t \to \infty$, i.e. there exists vectors ξ_i^∞ and η_i^∞ such that:

$$\xi_i^\infty = \lim_{t\to\infty} \xi(t) \qquad \text{and} \qquad \eta_i^\infty = \lim_{t\to\infty} \eta(t)$$

if this basis exists, then also $x_i(t)$ and $y_i(t)$ will have limiting values $x_i^\infty = \lim x_i(t)$ and $y_i^\infty = \lim y_i(t)$ and therefore, due to (2.16), we have:

$$\chi(v) = \max\{\varsigma_i : x_i \neq 0\} \cup \{\varsigma_{i+n} : y_i \neq 0\}$$

Therefore, if the symplectic eigenskeleton has a limit, the characteristics $\{\varsigma_i\}_{i=1}^{2n}$ are exactly the (out of order) Lyapunov characteristic exponents $\{\sigma_i\}_{i=1}^{2n}$. The symplectic eigenskeleton can be used in computing the LCE's of a Hamiltonian system. However, when integrating (2.7), it is *crucial* to use an integrator that preserves the symplecticity of the State Transition Matrix. We will discuss how to do this in §7.8 during our discussion on Lie groups.

2.5.5 The Henon-Heiles System

The canonical example of a chaotic Hamiltonian system is the Henon-Heiles System, given by the Hamiltonian:

$$H = \frac{1}{2}(u^2 + v^2 + x^2 + y^2) + x^2 y - \frac{y^3}{3} \tag{2.19}$$

Hamilton's equations yield the following dynamical equations of motion:

$$\frac{d}{dt}\begin{bmatrix} u \\ x \\ v \\ y \end{bmatrix} = \begin{bmatrix} -x - 2xy \\ u \\ -y - x^2 + y^2 \\ v \end{bmatrix} \tag{2.20}$$

The equations of motion for the STM are given by (2.7), which work out to be:

$$\frac{d\Phi}{dt} = \begin{pmatrix} 0 & -(1+2y) & 0 & -2x \\ 1 & 0 & 0 & 0 \\ 0 & -2x & 0 & (2y-1) \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \Phi \tag{2.21}$$

The Henon-Heiles system is highly chaotic at the energy $H = 1/6$, see Goldstein [54]. We integrated the system (2.20) with initial conditions $u_0 = 1/\sqrt{3}, x_0 = 0, v_0 = 0, y_0 = 0$ over the time interval $t \in [0, 500]$ and subsequently integrated (2.21) using the kinematic Lie group integrators discussed in §7.8. We will specifically discuss the preservation of the symplecticity of the State Transition Matrix for the Henon-Heiles system in §7.8.3.

Having determined $\Phi(t)$ over the interval $t \in [0, 500]$ we then computed the finite time characteristics defined by (2.17). These are plotted in Fig. 2.2. The positive

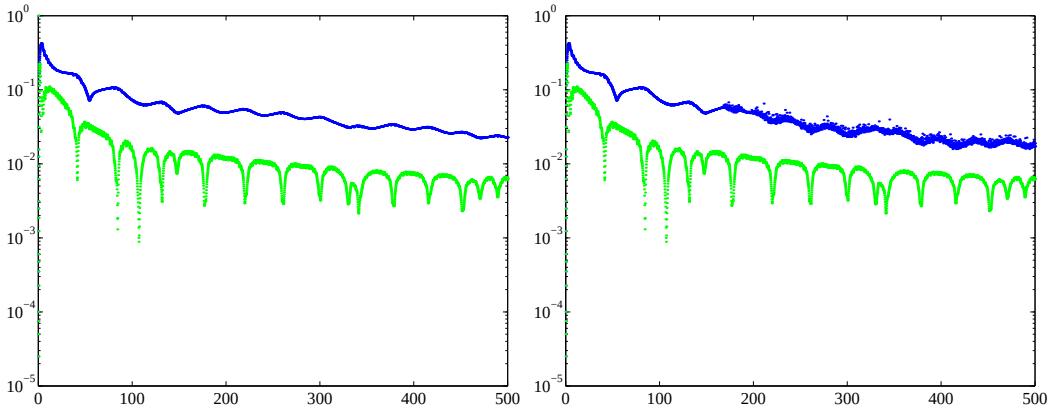

Figure 2.2: the positive $\varsigma_i(t)$'s (left) and the negative of the negative $\varsigma_i(t)$'s (right)

characteristics are on the left and the negative on the right. The eigenvalues of $\Phi(t)$ occur in reciprocal pairs, and the negative characteristics $\varsigma_i(t) < 0$ plotted on the right correspond to the eigenvalues of $\Phi(t)$ that are less than unity. However, this half of the set of eigenvalues of $\Phi(t)$ become exponentially small and so one can see

some numerical noise present in the graphs of the negative $\varsigma_i(t) < 0$ on the right.

Additionally we computed the time-varying symplectic eigenskeleton of the system, the symplectic and orthonormal eigenvectors of $\Psi(t)$. The components of $\xi_1(t)$ and $\xi_2(t)$ are the left graphs of Fig. 2.3 and the components of $\eta_1(t)$ and $\eta_2(t)$ are the right graphs. Each component is plotted with a different color, the u-component of each eigenvector is blue, the x-component is green, the v-component is red, and the y-component is mangenta. It appears that the limiting values ξ_1^∞, ξ_2^∞, η_1^∞, and η_2^∞ exist for this system. These vectors can be thought of formally as the symplectic eigenskeleton of the transformation that maps x_0 from $t = 0$ to $t = \infty$.

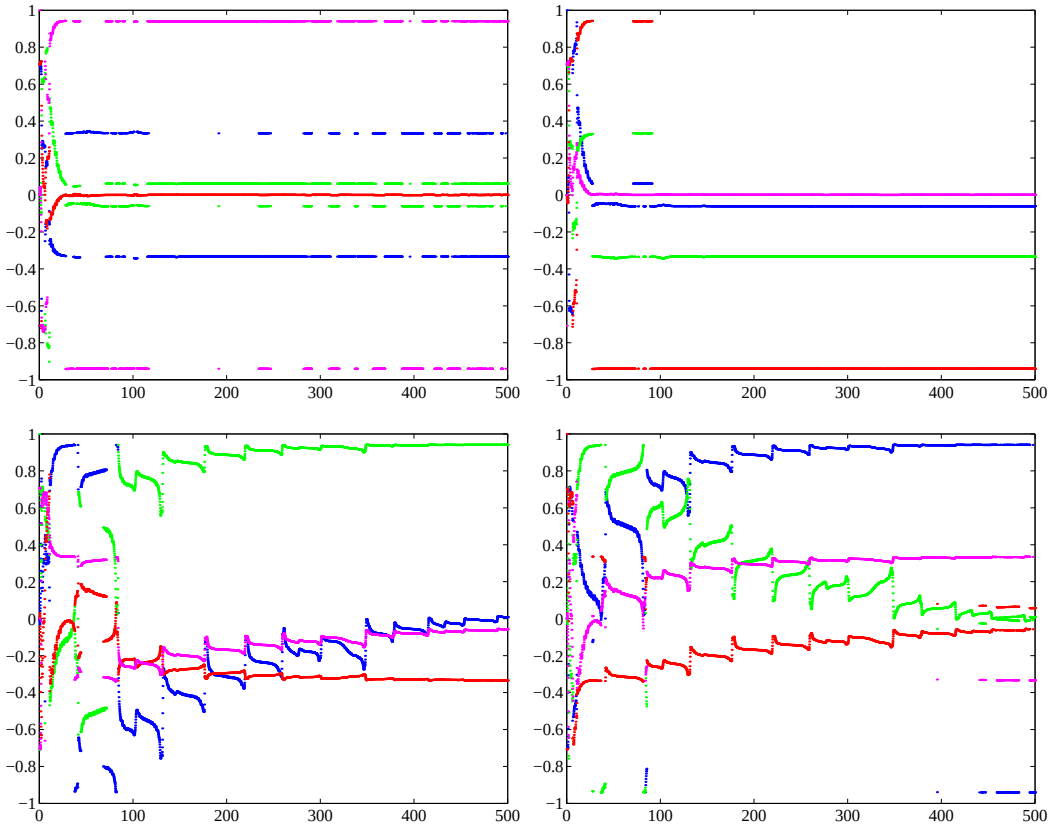

Figure 2.3: components of eigenvectors $\xi_1(t)$ (t.l.), $\eta_1(t)$ (t.r.), $\xi_2(t)$ (b.l.), and $\eta_2(t)$ (b.r.)

2.6 Tracking a Hyperbolic Flyby in the Earth-Moon System

As an application to the theory of area expansion factors and projections, we will consider the case of the hyperbolic flyby of an asteroid in the earth-moon system, where the asteroid's initial trajectory is not coherent with the plane of the lunar orbit. Let $\mu_\oplus$ and $\mu_{\leftmoon}$ be the gravitational parameters of the earth and moon, respectively, and a the radius of the moon's (circular) orbit. The angular velocity of the moon is thus taken to be:

$$\omega = \sqrt{\frac{\mu_\oplus}{a^3}}$$

so that the moon's position is given by:

$$\mathbf{r}_{\leftmoon} = \langle a\cos(\omega t), a\sin(\omega t), 0\rangle$$

Consider the position of the asteroid to be given in spherical coordinates as:

$$\mathbf{r}_* = \langle r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta\rangle$$

We now define the function $\psi(r,\theta,\phi,t)$, for ease of notation, as:

$$\psi = |\mathbf{r}_* - \mathbf{r}_{\leftmoon}| = \sqrt{r^2 + a^2 - 2ra\sin\theta\cos\Omega}$$

where we have defined the selenecentric azimuth as:

$$\Omega = \phi - \omega t$$

Definining the conjugate momenta as:

$$p_r = \dot{r}, \qquad p_\theta = r^2\dot{\theta}, \qquad p_\phi = r^2\sin^2\theta\dot{\phi}$$

the Hamiltonian for this system is

$$H = \frac{1}{2}\left\{p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2\sin^2\theta}\right\} - \frac{\mu_\oplus}{r} - \frac{\mu_{\leftmoon}}{\psi(r,\theta,\phi,t)}$$

Ordering the phase space as:

$$\mathbf{x} = \langle p_r, r, p_\theta, \theta, p_\phi, \phi \rangle$$

the dynamical equations of motion for the state $\mathbf{x}$ and the state transition matrix Φ are given by:

$$\dot{\mathbf{x}} = J \cdot \frac{\partial H}{\partial \mathbf{x}} \qquad \text{and} \qquad \dot{\Phi} = J \cdot \frac{\partial^2 H}{\partial \mathbf{x}^2} \cdot \Phi$$

Explicitly, the equations of motion for the hyperbolic flyby of the asteroid in the earth-moon system are given by the following:

$$\dot{r} = p_r \qquad \dot{\theta} = \frac{p_\theta}{r^2} \qquad \dot{\phi} = \frac{p_\phi}{r^2 \sin^2\theta}$$

$$\dot{p}_r = \frac{p_\theta^2}{r^3} + \frac{p_\phi^2}{r^3 \sin^2\theta} - \frac{\mu_\oplus}{r^2} - \frac{\mu_\leftmoon \left(r - a\sin\theta\cos\Omega\right)}{\phi^{3/2}}$$

$$\dot{p}_\theta = \frac{p_\phi^2 \cos\theta}{r^2 \sin^3\theta} + \frac{\mu_\leftmoon \, ra\cos\theta\cos\Omega}{\psi^{3/2}}$$

$$\dot{p}_\phi = \frac{\mu_\leftmoon \, ra\sin\theta\sin\Omega}{\psi^{3/2}}$$

The coefficient matrix in the STM equations of motion is given by $J \cdot H_{xx} =$

$$\begin{pmatrix} 0 & -H_{rr} & -H_{rp_\theta} & -H_{r\theta} & -H_{rp_\phi} & -H_{r\phi} \\ H_{p_r p_r} & 0 & 0 & 0 & 0 & 0 \\ 0 & -H_{r\theta} & 0 & -H_{\theta\theta} & -H_{\theta p_\phi} & -H_{\theta\phi} \\ 0 & H_{rp_\theta} & H_{p_\theta p_\theta} & 0 & 0 & 0 \\ 0 & -H_{r\phi} & 0 & -H_{\theta\phi} & 0 & -H_{\phi\phi} \\ 0 & H_{rp_\phi} & 0 & H_{\theta p_\phi} & H_{p_\phi p_\phi} & 0 \end{pmatrix}$$

with

$$H_{p_r p_r} = 1 \qquad H_{p_\theta p_\theta} = r^{-2} \qquad H_{p_\phi p_\phi} = r^{-2}\mathrm{s}^{-2}\theta$$

$$H_{rr} = \frac{3p_\theta^2}{r^4} + \frac{3p_\phi^2}{r^4 \mathrm{s}^2\theta} - \frac{2\mu_\oplus}{r^3} - \frac{3\mu_\leftmoon \left(r - a\mathrm{s}\theta\mathrm{c}\Omega\right)^2}{\psi^{5/2}} + \frac{\mu_\leftmoon}{\psi^{3/2}}$$

$$H_{rp_\theta} = \frac{-2p_\theta}{r^3} \qquad H_{rp_\phi} = \frac{-2p_\phi}{r^3 \mathrm{s}^2\theta}$$

$$H_{r\theta} = \frac{2p_\phi^2 \mathrm{c}\theta}{r^3 \mathrm{s}^3\theta} + \frac{3\mu_{\leftmoon}\,(r - a\mathrm{s}\theta\mathrm{c}\Omega) r a \mathrm{c}\theta\mathrm{c}\Omega}{\psi^{5/2}} - \frac{\mu_{\leftmoon}\, a\mathrm{c}\theta\mathrm{c}\Omega}{\psi^{3/2}}$$

$$H_{r\phi} = \frac{3\mu_{\leftmoon}\,(r - a\mathrm{s}\theta\mathrm{c}\Omega) r a \mathrm{s}\theta\mathrm{s}\Omega}{\psi^{5/2}} - \frac{\mu_{\leftmoon}\, a\mathrm{s}\theta\mathrm{s}\Omega}{\psi^{3/2}}$$

$$H_{\theta\theta} = \frac{3p_\phi^2 \mathrm{c}^2\theta}{r^2 \mathrm{s}^4\theta} + \frac{p_\phi^2}{r^2 \mathrm{s}^2\theta} - \frac{3\mu_{\leftmoon}\, r^2 a^2 \mathrm{c}^2\theta\mathrm{c}^2\Omega}{\psi^{5/2}} + \frac{\mu_{\leftmoon}\, r a \mathrm{s}\theta\mathrm{c}\Omega}{\psi^{3/2}}$$

$$H_{\theta p_\phi} = \frac{-2p_\phi \mathrm{c}\theta}{r^2 \mathrm{s}^3\theta}$$

$$H_{\theta\phi} = -\frac{3\mu_{\leftmoon}\, r^2 a^2 \mathrm{c}\theta\mathrm{s}\theta\mathrm{c}\Omega\mathrm{s}\Omega}{\psi^{5/2}} - \frac{\mu_{\leftmoon}\, r a \mathrm{c}\theta\mathrm{s}\Omega}{\psi^{3/2}}$$

where we have abbreviated sin and cos with s and c, respectively. We now have a coupled system of 42 first order differential equations which can be integrated numerically.

For a set of initial conditions we take $\langle p_r, r\rangle_0 \in [-1.1, -0.9] \times [1.8, 2.2]$, and $\theta = -2,\ p_\theta = 1,\ \phi = 0, p_\phi = 0$. We normalize $\mu_\oplus = 1,\ \mu_{\leftmoon} = 0.1111,\ a = 0.5109$ and integrate the initial uncertainty distribution to a time $t_f = 3$. Since we take p_θ and p_ϕ as fixed, our initial uncertainty distribution represents a parasymplectic surface. The questions now becomes, where in the sky should we look to find the asteroid?

We discretize the uncertainty domain into a 20×20 grid of virtual asteroids (VAs). We then integrate the trajectory and STM for each one. Suppose Φ is the STM at time 3. Then the area expansion from the initial (p_{ri}, r_i) plane to the projection of the surface onto the final (θ_f, ϕ_f) plane is given by:

$$\frac{\partial(\theta_f, \phi_f)}{\partial(p_{ri}, r_i)} = \Phi_{\theta p_r}\Phi_{\phi r} - \Phi_{\theta r}\Phi_{\phi p_r}$$

This expansion projection factor is now plotted (in absolute value) against the initial uncertainty distribution and the projection of the final surface on the (θ_f, ϕ_f) plane, see Fig. 2.4.

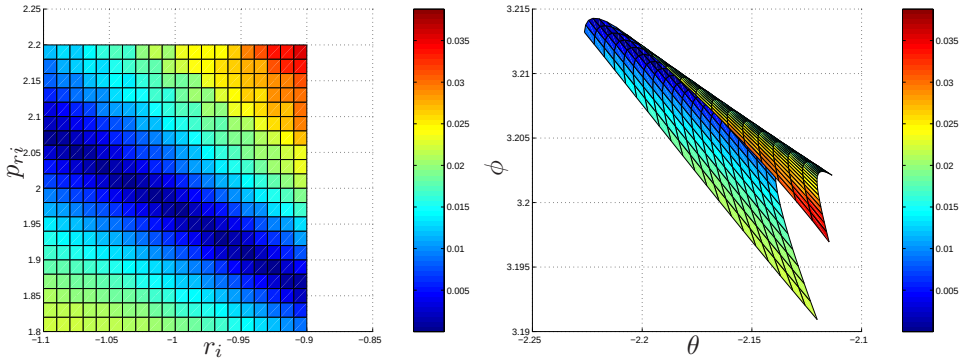

Figure 2.4: $\dfrac{\partial(\theta_f, \phi_f)}{\partial(p_{ri}, r_i)}$ plotted against the Initial and Final distributions

We assume that it is equally probable for the asteroid to be anywhere on the initial (p_{ri}, r_i) uncertainty plane. The state transition matrix gives us much information on the local structure of the uncertainty distribution at a later time. In particular, we see the large blue strip in the middle of Fig. 2.4-L occupies approximately 1/3 the total initial uncertainty area. Therefore the probability of the asteroid being in this region is approximately 1/3. Straight down the center of the dark blue strip the area expansion projection factor is zero, representing the "fold over" line. Since the area expansion is low here, the blue area occupies relatively little area in the future observation $(\theta - \phi)$ plane. Little area and high probability means that this is a good place to start searching for the asteroid, over the yellow and the red regions where the area expansion is much greater and therefore the probability is spread out over a greater space.

We can also make plots of the projection of the mapped surface onto the three symplectic planes, which is done in Fig. 2.5. These are simply the plots of the 2×2 symplectic subdeterminants of the $\langle p_r, r \rangle$ column of the STM. For every grid point, the sum of the three expansion factors equals one.

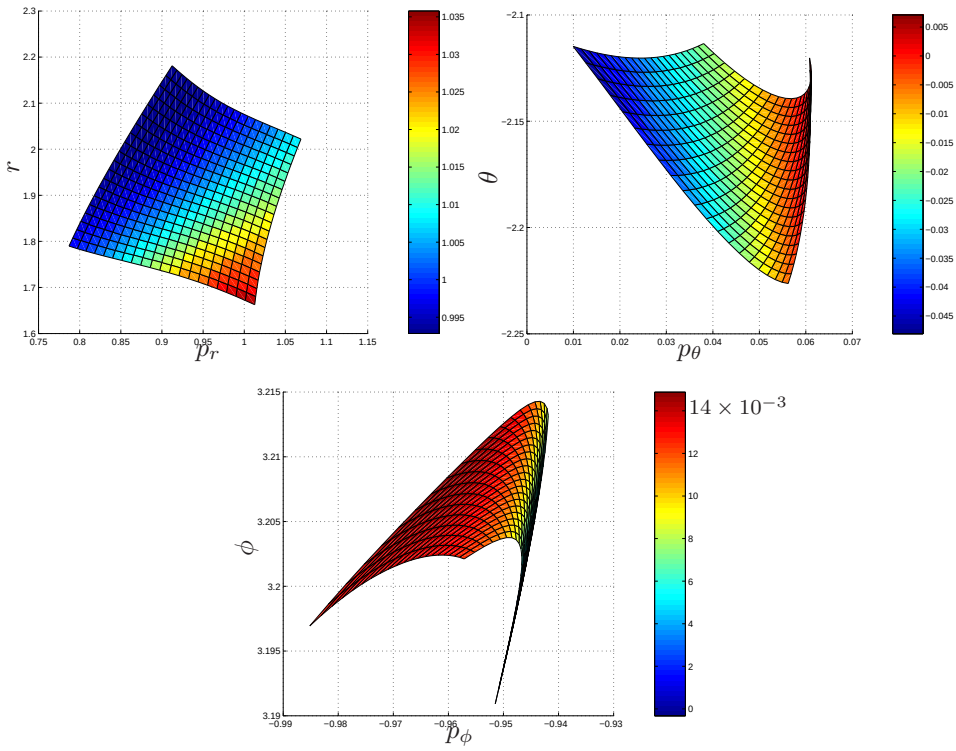

Figure 2.5: Projection of mapped surface onto symplectic planes

2.7 Applications to Control

We view this paper as a theoretical paper which studies some of the fundamental constraints in the propagation of volumes rather than trajectories in dynamical and control systems. This idea has already been advocated in viability theory and in some robust control design papers, see e.g. Mayne [97]. In the final subsection below we present some future directions for using the theory presented here in the context of control. The examples below, though they do not utilize the full breadth of the theoretical developments presented in the paper, were chosen to illustrate some key ideas regarding propagation of surfaces and uncertainties in the control theory setting.

2.7.1 The Kinematic Heisenberg System

The Heisenberg System is a classical underactuated kinematic control problem with nonholonomic constraints, see Bloch [16], Brockett [24]. The configuration manifold for the system is given by $Q = \mathbb{R}^3$, with coordinates $q = \langle x, y, z \rangle$. Motion is constrained by the relation $\dot{z} = y\dot{x} - x\dot{y}$. Supposing we have controls $u,\ v$ over the x and y velocities, the kinematic control system can be written:

$$\begin{aligned} \dot{x} &= u \\ \dot{y} &= v \\ \dot{z} &= yu - xv \end{aligned} \tag{2.22}$$

Suppose the initial state of the system is given to be within the two-dimensional uncertainty distribution $\sigma(0) = \{\langle x, y, z \rangle \in \mathbb{R}^3 : x \in [-1, 1], y \in [-1, 1], z = 0\}$, and we wish to determine an open loop control law that maneuvers the particle to the point $\langle 0, 0, 1 \rangle$ during the time interval $t \in [0, 1]$ in some optimal sense.

Let $\sigma(t)$ be the time evolution of $\sigma(0)$ to time t. We can parameterize the surface $\sigma(t)$ by the initial data $(X, Y) \in [-1, 1] \times [-1, 1]$, so that, at time t, the surface is given parametrically by $\langle x(X, Y; t), y(X, Y; t), z(X, Y; t) \rangle$. The distance from an arbitrary point on the final surface $\sigma(1)$ to the target point is

$$\sqrt{x(X, Y; 1)^2 + y(X, Y; 1)^2 + (1 - z(X, Y; 1))^2}$$

The dynamics (2.22) depend upon the choice of control $\langle u(t), v(t) \rangle$. We thus pose the following control problem:

Problem: Choose $\langle u(t), v(t) \rangle$ so as to minimize:

$$\iint_{\sigma(1)} \sqrt{x^2 + y^2 + (1 - z)^2} d\mathcal{A} \tag{2.23}$$

Solution: First we need to compute the determinant of the surface metric. The State Transition Matrix (STM) dynamics are given by:

$$\dot{\Phi} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -v & u & 0 \end{bmatrix} \cdot \Phi, \qquad \Phi(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which can immediately be integrated to yield:

$$\Phi(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -(y(t) - y(0)) & (x(t) - x(0)) & 1 \end{bmatrix}$$

The metric determinant of the surface at time t is given by the Gram determinant of the first two columns of $\Phi(t)$:

$$g(X, Y; t) = 1 + (x(t) - X)^2 + (y(t) - Y)^2 \tag{2.24}$$

Thus (2.23) is equivalent to:

$$\int_{-1}^{1} \int_{-1}^{1} \sqrt{x^2 + y^2 + (1 - z)^2} \sqrt{g(X, Y; 1)} \; dX \; dY$$

where $\langle x, y, z \rangle = \langle x(X, Y; 1), y(X, Y; 1), z(X, Y; 1) \rangle$. Without loss of generality, let us instead minimize:

$$f = \int_{-1}^{1} \int_{-1}^{1} \left[x(1)^2 + y(1)^2 + (1 - z(1))^2 \right] g(1) \; dX dY \tag{2.25}$$

Define now:

$$\mu = \int_0^t u(\tau) \; d\tau, \; \nu = \int_0^t v(\tau) \; d\tau, \; \alpha = \int_0^t (\nu u - \mu v) \; d\tau$$

so that the solution to (2.22) can be expressed as:

$$\begin{aligned} x(t) &= X + \mu(t) \\ y(t) &= Y + \nu(t) \\ z(t) &= Y\mu(t) - X\nu(t) + \alpha(t) \end{aligned}$$

This exposes the dependence of $x,\ y$, and z on the initial conditions X and Y. Inserting into the surface metric (2.24), one can explicitly integrate (2.25) to find:

$$f = \frac{4}{3}\left(1+\mu^2+\nu^2\right)\left(4\mu^2+4\nu^2+3\alpha^2-6\alpha+5\right)$$

This function has a global minimum at $\mu = 0,\ \nu = 0,\ \alpha = 1$. Any control law that satisfies:

$$\int_0^1 u(t)dt = 0,\ \int_0^1 v(t)dt = 0,\ \int_0^1 (\nu(t)u(t) - \mu(t)v(t))dt = 1$$

will leave the final uncertainty distribution as close to the target point as possible, in the above sense. Notice that a physical interpretation of the vector quantity $\langle \mu(t), \nu(t), \alpha(t)\rangle$ is that it is the position vector of the point on the surface that was initially at $\langle 0,0,0\rangle$. Thus, any control law that leaves the centroid of the surface at the target point will automatically minimize (2.23). One such trajectory, given in Bloch [16] using $\dot{y}(0) = 0$, is:

$$\begin{aligned}
\mu(t) &= \frac{1}{\sqrt{2\pi}}\sin(2\pi t) \\
\nu(t) &= \frac{1}{\sqrt{2\pi}}(1-\cos(2\pi t)) \\
\alpha(t) &= t(1-\sin(2\pi t))
\end{aligned}$$

The uncertainty surface at various time snapshots for the control law $u(t) = \dot{\mu}(t)$, $v(t) = \dot{\nu}(t)$ is given in Fig. 2.6.

2.7.2 The Falling, Rolling Disc

Consider the falling rolling disc of radius $r = 1$, Fig. 2.7, whose configuration is described by the contact point (x, y) and the Classical Euler angles (ϕ, θ, ψ). Suppose we have direct control over the body-axis angular velocities $u = \dot{\phi}\sin\theta,\ v = \dot{\theta},\ w = \dot{\phi}\cos\theta + \dot{\psi}$, and suppose the system is subject to nonholonomic constraints $\dot{x} + \dot{\psi}\cos\phi = 0$ and $\dot{y} + \dot{\psi}\sin\phi = 0$. The dynamics is given by the system:

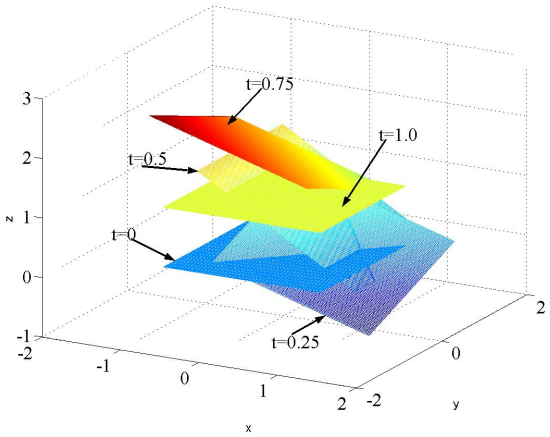

Figure 2.6: Optimal Uncertainty Maneuver

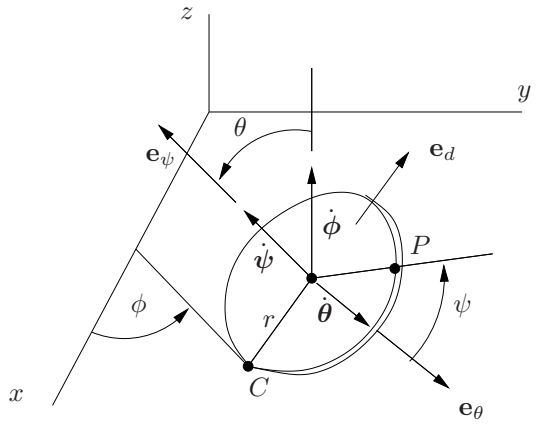

Figure 2.7: Euler Angles of the Falling Rolling Disc

$$\begin{aligned}
\dot{x} &= u\cot\theta\cos\phi - w\cos\phi \\
\dot{y} &= u\cot\theta\sin\phi - w\sin\phi \\
\dot{\phi} &= u\csc\theta \\
\dot{\theta} &= v \\
\dot{\psi} &= -u\cot\theta + w
\end{aligned}$$

Using the notation $\dot{q} = f(q, u)$, the coefficient matrix in the STM dynamics equation is:

$$\frac{\partial f}{\partial q} = \begin{bmatrix} 0 & 0 & -(u\cot\theta - w)\sin\phi & u\csc^2\theta\cos\phi & 0 \\ 0 & 0 & (u\cot\theta - w)\cos\phi & u\csc^2\theta\sin\phi & 0 \\ 0 & 0 & 0 & -u\cot\theta\csc\theta & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -u\csc^2\theta & 0 \end{bmatrix}$$

Since the STM $\Phi(t)$ is initially the identity, we find:

$$\Phi(t) = \begin{bmatrix} 1 & 0 & A & C & 0 \\ 0 & 1 & B & D & 0 \\ 0 & 0 & 1 & E & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & F & 1 \end{bmatrix}$$

where

$$\begin{aligned} A &= \int_0^t -(u\cot\theta - w)\sin\phi \; dt \\ B &= \int_0^t (u\cot\theta - w)\cos\phi \; dt \\ C &= \int_0^t \left(-(u\cot\theta - w)\sin\phi E + u\csc^2\theta\cos\phi\right) \; dt \\ D &= \int_0^t \left((u\cot\theta - w)\cos\phi E + u\csc^2\theta\sin\phi\right) \; dt \\ E &= \int_0^t -u\cot\theta\csc\theta \; dt \\ F &= \int_0^t -\csc^2\theta \; dt \end{aligned}$$

Supposing there is initial uncertainty in the θ and ϕ components, it is the third and fourth columns of the STM that will be crucial in determining the uncertainty evolution. Suppose further our desire is that the projection of the final uncertainty onto the

x, y planes has zero area; i.e. at worst there is a one-dimensional uncertainty in the contact point position. Then we wish at time $t = 1$, that $A(1)D(1) - B(1)C(1) = 0$. To achieve this, one may use any control law with that satisfies the relation

$$u \cot \theta - w = 0$$

Such a control law will leave $A(t) \equiv 0,\ B(t) \equiv 0$, so that the uncertainty projection onto the $x - y$ plane has zero area for all time.

2.7.3 Future Directions in Control

As the same symplectic constraints apply to the evolution of an optimal control system's states and co-states, these results also have an implication for the stability and robustness of an optimal feedback control law. This aspect of the study can be reduced to two fundamental approaches, the implication of initial value distributions on the subsequent evolution of a trajectory in the neighborhood of the true optimal trajectory, and how the symplectic invariants manifest themselves in the solutions of two-point boundary value problems.

First, how do uncertainties in the initial state or in the initial application of the control map to the target conditions? As is well known, by definition an explicit optimal feedback control law is asymptotically stable when restricted to the state variables. However, as the necessary conditions from which the feedback control law can arise form a Hamiltonian system, this implies that the co-states are unstable and should diverge. This becomes an issue if the state is not perfectly determined or if the control function is not exact but only lies in a neighborhood of the true optimal control, and should lead to instabilities arising in the state variables of the system. These relationships can be studied using integral invariants and symplectic capacities to determine the robustness of the specific optimal control laws by studying how the

phase volume surrounding them maps under the necessary conditions. Of special interest will be the identification of the maximum and minimum uncertainty growth directions.

Second, given an optimal control feedback law (i.e., given the solution to the Hamilton-Jacobi-Bellman equation), how do simultaneous uncertainties in both the initial state and target state affect the distribution of the adjoints, and what structure may lie within these distributions that arise from the Hamiltonian formulation of the necessary conditions? Applying the Hamilton Principle Function approach, which provides an explicit solution to the two-point boundary value problem and which is directly, analytically related to the optimal control, we implicitly define an initial set of optimal controls that will lead to a proscribed final region in the neighborhood of the nominal target state. This defines for us an open set of controls, within which lie optimal trajectories that all achieve the final state to within some desired, and proscribed, accuracy. Depending on the size and the distribution of the initial uncertainties and the tolerable final uncertainties, we can identify a symplectic width which should provide explicit ranges in the set of initial controls that will lead to a guaranteed, optimal arrival in the vicinity of the final state. Such a development can provide additional insight into the robustness of optimal controls and how gracefully they will degrade when we allow for finite miss distances for the target state.

There is also a clear identification between distributions in phase space and probabilistic interpretations of the state of a system. Thus, our research also has a direct bearing on predicted uncertainties in a dynamical system after being mapped in time, and will define for us an absolute minimum region within which the uncertainty of the system can be isolated.

2.8 Conclusion

We showed how the expansion of subvolumes in the local neighborhood of a nominal trajectory leads to the local collapse of the supporting phase space. Moreover, we produced a preferred basis, the symplectic eigenskeleton, which resists collapse and returns uncertainty distributions that are initially parallel to the basis to their minimal uncertainty state at a fixed final time.

CHAPTER III

Delaunay Variables and the Tracking of Space Debris

In Chapter II we introduced the idea of propagating subvolumes in dynamical systems and discussed constraints that exist on the evolution of these subvolumes when the equations of motion are Hamiltonian. The primary application of this idea that we shall consider lies in the field of Space Situational Awareness (SSA), and is related to our discussion of probability mappings of uncertainty distributions for the hyperbolic flyby of asteroids, discussed in §2.6. Simply put, the problem of SSA is to carry out orbit determinations for particles of space debris trapped in earth orbit. For a single optical track, angle and angle rates are known, whereas the range and range rate data is largely undetermined. Based on a single track, the system is known to lie in a subvolume of the higher dimensional phase space. In this chapter we will review a new approach for performing these orbit determinations that is based on the theory of the dynamical propagation of subvolumes.

3.1 Introduction

A sequence of optical measurements of an Earth orbiting object over one track has sufficient information to determine the angles and angular rates with some degree of precision, but cannot measure the range or range-rate. Despite the lack of complete state information, constraints on range and range-rate can be determined by apply-

ing physical constraints. Applying these constraints to an observation constrains an object's state to lie within a two-dimensional submanifold of phase space. Such a region can be mapped into orbital element space and propagated in time. As the regions in question are two-dimensional in nature it is possible to model them with high precision without excessive computational burden. A second observation of a space object can similarly be mapped into a similar submanifold of orbit element space and intersected with a previous observation mapped to the same epoch. If the object is the same, this intersection process yields a non-zero set which may be unique, depending on observational geometries. If the object is different the intersection is null in general. Addition of uncertainty in the angle and angle-rate measurements yields finite regions of intersection, sufficient to localize an initial estimate for a connecting orbit if the two mapped observation manifolds have regions of non-zero intersection. If the submanifolds are mapped into a Hamiltonian canonical set of elements, such as the Delaunay or Hamiltonian elements, the projection of this submanifold into the conjugate pairs of coordinates and momenta must sum to a constant, due to the integral invariants of Poincarè-Cartan. This provides additional structure to these regions as this integral invariance is conserved when mapping in time and thus the area of these projections remain constant.

3.2 Introduction

A problem of recent interest to space faring nations is the tracking, orbit determination, and cataloging of all pieces of artificial space debris particles in low, medium, and high Earth orbits, a population of more than 300,000 particles. The United States Air Force Space Command has installed a network of 25 radar and optical sensors for this task, which make about 80,000 observations daily. For more

background on observation of Space Debris, see Rossi [117, 116].

Using optical measurements for particles in medium to high Earth orbits, the angles and angular rates of the passing particles, as seen from an Earth-based telescope, can be measured to high precision, however the range and range-rates are largely undetermined. The set of angles and angular rates of such a measurement is called an *attributable vector.* Recent work has been done in outlining a precise mathematical description of the *admissible region* of the range, range-rate plane given an attributable vector observed by radar or optical measurements, Tommei et al. [130]. The admissible region is a two-dimensional surface that lives in the six-dimensional phase space surrounding the Earth. This surface consists of all points in phase space where the true particle can possibly be found. For this reason we will sometimes refer to it as the uncertainty surface. This uncertainty surface is then discretized by a number of points called *Virtual Debris particles*, or VD particles. Each VD particle is an approximation to a possible orbit for the observed particle of space debris. Viewed as a whole, the set of VD particles forms a *virtual debris field*, or VD field, which approximates the macroscopic uncertainty distribution associated with a given attributable vector. For optical measurements, the admissible region on the range range-rate plane is only restricted by the fact that the debris particle should be gravitationally bound to the Earth (negative geocentric energy) and that it lie within some region (2 and 20 Earth radii) of the observer (Tommei et al. [130]). Our preliminary numerical analysis of a randomaly chosen attributable vector and its corresponding VD field showed that before several hours had passed, about half of the VD's will have crashed into the Earth. Motivated by these observations, we present in this paper a tighter restriction on the uncertainty region of the range range-rate plane. In particular, in addition to demanding the distance between the debris par-

ticle and observer lie between 2 and 20 Earth radii *at the moment of observation*, we place additional restrictions on the periapsis and apoapsis of the orbit. In particular we require the orbit's periapsis to be greater than 1 Earth radius plus 200 km, and the apoapsis to be less than 25 Earth radii. This places additional constraints on the admissible region and reduces its size considerably.

In §3.4, we introduce Delaunay variables and discuss the transformation between the observation space (topocentric polar coordinates) and Delaunay space, and the corresponding Jacobian matrix of the transformation. Delaunay variables were introduced by Delaunay [42, 43], and can be derived several ways. One can derive them by solving the Hamilton-Jacobi equations (Born [22]) or by using Lagrange brackets (Brouwer & Clemence [25]; Abraham & Marsden [1]). A modern geometric derivation is given in Chang & Marsden [33]. A nice aspect of Delaunay variables is that they can be written in symplectic (canonical) coordinate-conjugate momentum pairs. The corresponding equations of motion, even for the perturbed problem, can be written in the form of Hamilton's equations. Therefore the integral invariants of Poincaré-Cartan apply, and the sum of the signed area projections onto the three symplectic planes must be conserved. The Delaunay variables are also the action-angle variables of the two body problem; the angles are the coordinates and the actions are the momenta. For the Kepler problem, all Delaunay variables except a single angle variable are conserved. Even though the two-dimensional uncertainty distribution in geocentric cartesion wraps around the Earth in phase space rapidly, the projection of the same uncertainty distribution on two of the three Delaunay planes is static. The angle variable of the third plane is the mean anomaly, so all the VD particles march at different rates (which depend only on the mean anomaly's conjugate momentum) along this direction. Since the angle variables are modulo 2π, the surface begins to

wrap around and becomes more and more "shredded" as time progresses. Even in the averaged perturbed problem, only the angle variables change in time, so that the same shredding can be seen in each of the symplectic planes.

In §3.6 - §3.8 we will discuss an algorithm for correlating two observations of the same debris particle. We will first consider the case of having two zenith observations at our disposal. We will then add some fuzz to the uncertainty distribution by considering near-zenith observations. Finally we discuss the case when the J_2 term is added to Earth's potential energy.

3.3 Admissible Region

In this section we shall review the *admissible region* for a space debris particle observed by a ground based optical sensor, as presented in Tommei et al. [130], and offer an additional physical constraint that will further limit the size of this region in the range range-rate plane.

3.3.1 Attributable Vectors

Let P be the geocentric position of a space debris particle and P_O the geocentric position of the optical observer. Let the position of the debris particle with respect to the observer be denoted $P_D = \rho\hat{R}$, where $\hat{R}$ is a unit vector pointing from the observer to the particle. This gives us:

$$P = P_O + P_D$$

Let

$$(\rho, \alpha, \delta) \in \mathbb{R}^+ \times [-\pi, \pi) \times (-\pi/2, \pi/2)$$

be the spherical polar coordinates defining P_D. Typically one can choose the J2000 coordinate systems so that α is the right ascension and δ is the declination.

Definition 8. *An optical attributable vector is a vector*

$$A = (\alpha, \delta, \dot{\alpha}, \dot{\delta}) \in [-\pi, \pi) \times (-\pi/2, \pi/2) \times \mathbb{R}^2$$

observed at time t.

The optical attributable vector is precisely the set of coordinates that can be measured from the observer's frame *on the Earth's surface* at time t. Note that unless the debris particle is seen directly overhead, what we call a zenith measurement, this will translate into uncertainty in the polar angles of the *geocentric* frame. Since the coordinate transformation to the inertial geocentric frame depends on the position of the Earth, additional information must be stored along with the attributable vector. The full set of data that should be tabulated with each observation is:

$$x = (A, t, L) \in \mathbb{R}^5 \times \mathbb{N}$$

where $A = (\alpha, \delta, \dot{\alpha}, \dot{\delta})$ is the attributable vector, t is the time of observation, and L is observatory which made the observation (each observatory can be given integer-valued names). A function can then be defined as follows:

$$\psi : (t, L) \to (h, \Theta, \Phi)$$

where h is the altitude of the observatory (which we ignore in the current discussion), and (Θ, Φ) is the inertial angular location of observatory L at time t. The observation data x and inertial orientation function of the observatory ψ can then be unraveled to form the actual useable information for the coordinate transformations:

$$\mathfrak{X} = (A, t, h, \Theta, \Phi) \in \mathbb{R}^8$$

3.3.2 Measuring Attributable Vectors

One of the primary technical difficulties in Space Situational Awareness (SSA) is the inability to accurately estimate the full dynamical state of an object based on

a single track of data. For example, optical search and survey telescopes will only observe objects for very short durations (minutes at most) [124]. This track of data contains information on the angular location of the satellite, but in general does not have sufficient information to allow for an accurate orbit determination of the object. For meaningful orbit characterization, the object must be observed again during a later pass. The problem, of course, is to discern which uncorrelated tracks are the same object.

Our approach to this problem is two-fold. First, as described in this section, is to extract the maximum amount of usable information from a single track of a space object and to also bound the possible domain of the unmeasured state components. Second, described in the next section, is to use this information to correlate one observation track with another from a previous observation in order to detect whether the two objects are the same. In the following we will focus specifically on optical observations of space debris, although the theory and analysis will also apply to radar observations of space debris. This focus is done solely for simplicity of presentation.

The challenge is to extract the maximum amount of usable information from a single track and also derive meaningful constraints on the remaining uncertainty aspects of the state. Instead of determining an overall covariance matrix for the object's state, based on a single track, that contains all 6 dimensions, our approach is to use the track observations to isolate those components of the object's state that can be constrained. For an optical track consisting of several angular observations over a time span of minutes these are the angular location and angular rate of the object at a specific epoch, generally chosen to lie within the track. The idea is to use the multiple angular measurements to develop an improved estimate of the angular location of the object and the angular rate of the object, and then use these

measurements to constrain the unmeasured states of the object. This approach was recently proposed in [130] by Tommei et. al and is further extended in our research. While Tommei et al. discuss that a single track of optical data can determine the angle and angle rates of the object at a specific epoch, they did not outline how this determination could be made.

If, during an observation period, a space object passes through the field of view of an optical telescope, the optical telescope can take several measurements of that object, equivalent to a series of angles at specific times, the total time span being a number of minutes. We assume that these angles can be identified with each other to produce a single track of observable data for the space object. There is obviously more information in this pass of data than just a single fix of the angular location of the space object relative to the observer, however there is not enough information to provide an accurate orbit. To capture this additional information, we can estimate the space object's angular location, angular rate and angular acceleration at a fixed epoch, chosen within the tracking pass. The additional information content goes into reducing the uncertainty of the angle, angle-rate and angle-acceleration measurement at epoch. This approach recognizes that there is little information in one track related to the object range and range-rate, and concentrates on fixing the angles, angle-rates, and angle-accelerations to a higher level of precision. A similar approach has been taken to estimate the information content of a single pass of Doppler data for an interplanetary spacecraft [63].

The proposed approach is to model the kinematics of the angles, however it should be noted that the acceleration of either angle is a function of the angles, angle rates, range, and range rate through the equations of motion. Thus, technically, once the angular accelerations and the angle rates are measured for both angles we have an

algebraic relationship:

$$\ddot{\alpha} = f_\alpha(\alpha, \dot{\alpha}, \delta, \dot{\delta}, \rho, \dot{\rho}) \tag{3.1}$$

$$\ddot{\delta} = f_\delta(\alpha, \dot{\alpha}, \delta, \dot{\delta}, \rho, \dot{\rho}) \tag{3.2}$$

from which the range and range rate can be solved for. Practically, however, there is a large uncertainty in these accelerations which translates to a large uncertainty in the estimated values of ρ and $\dot{\rho}$. From a different perspective, the range and range rate are not being directly measured and can at best be constrained with this approach. Due to these uncertainties we have devised an algorithm which does not rely on estimates of range and range rate to perform initial orbit determination.

In the following we provide an example of this approach for estimating a single angle and its associated rates. This can easily be expanded to a full estimate of both angles. At the heart of the approach is to estimate the kinematics of the angular motion of the object, in the Earth fixed frame, during the tracking pass. For simplicity in this example we assume the angular motion can be modeled kinematically as:

$$\alpha(t) = \alpha_o + \dot{\alpha}_o(t - t_o) + \frac{1}{2}\ddot{\alpha}_o(t - t_o)^2 \tag{3.3}$$

For a general application of this approach, higher order derivatives can be added and estimated and biases in the angular motion can also be added. The biases can only be estimated if tracks are compared between different objects. The problem then reduces to estimating the angular position, angular rate and angular acceleration of the object at time t_o. For simplicity in this example we assume 1-dimensional angular motion and a sequence of N equally spaced observations with uncorrelated error statistics centered on the epoch t_o and covering a time span T. This defines an estimation problem for the angle at epoch, α_o, the angular rate at epoch, $\dot{\alpha}_o$, and the

angular acceleration at epoch, $\ddot{\alpha}_o$. There is sufficient information to estimate these provided there are at least 3 angle measurements in the track.

For simplicity we assume that the individual angle measurements are uncorrelated with each other. Then the least-squares estimation problem is to minimize the cost function

$$J = \frac{1}{2\sigma_\alpha^2} \sum_{i=1}^{N} \left(\alpha(t_i) - \alpha_i\right)^2 \tag{3.4}$$

where σ_α is the measurement uncertainty, α_i are the actual measurements and the $\alpha(t_i)$ are the predicted measurements. The quantities to estimate are $\alpha_o, \dot{\alpha}_o, \ddot{\alpha}_o$. Forming the necessary equations, and using the fact that $\alpha(t)$ is linear in the quantities to estimate, we form the normal equations:

$$\Lambda \begin{bmatrix} \alpha_o \\ \dot{\alpha}_o \\ \ddot{\alpha}_o \end{bmatrix} = d \tag{3.5}$$

$$\Lambda = \sum_{i=1}^{N} \frac{1}{\sigma_\alpha^2} \begin{bmatrix} 1 & t_i - t_o & \frac{1}{2}(t_i - t_o)^2 \\ t_i - t_o & (t_i - t_o)^2 & \frac{1}{2}(t_i - t_o)^3 \\ \frac{1}{2}(t_i - t_o)^2 & \frac{1}{2}(t_i - t_o)^3 & \frac{1}{4}(t_i - t_o)^4 \end{bmatrix} \tag{3.6}$$

$$d = \sum_{i=1}^{N} \frac{\alpha_i}{\sigma_\alpha^2} \begin{bmatrix} 1 \\ t_i - t_o \\ \frac{1}{2}(t_i - t_o)^2 \end{bmatrix} \tag{3.7}$$

To characterize the information content in a single pass of data, we can explicitly compute the Information Matrix Λ. To do this we assume the measurements are taken at equal times $t_i = t_o + \frac{T}{2n} i$ where $i = -n, -(n-1), \ldots, -1, 0, 1, \ldots, n$, forming a total of $2n + 1$ measurements over the time-span T. Due to this assumption, the

odd terms will all sum to zero:

$$\Lambda = (2n+1)\begin{bmatrix} 1 & 0 & \frac{n(n+1)}{6}\left(\frac{T}{2n}\right)^2 \\ 0 & \frac{n(n+1)}{3}\left(\frac{T}{2n}\right)^2 & 0 \\ \frac{n(n+1)}{6}\left(\frac{T}{2n}\right)^2 & 0 & \frac{n(n+1)(3n^2+3n-1)}{60}\left(\frac{T}{2n}\right)^4 \end{bmatrix} \tag{3.8}$$

The inverse of this is the covariance matrix, $P = \Lambda^{-1}$, and has information on the accuracy to which the angle quantities are measured. To simplify the computation, assume that $n \gg 1$ (often not a valid assumption) and invert the matrix to find:

$$P = \frac{\sigma_\alpha^2}{(2n+1)}\begin{bmatrix} \frac{9}{4} & 0 & -\frac{15}{2}\left(\frac{2}{T}\right)^2 \\ 0 & 3\left(\frac{2}{T}\right)^2 & 0 \\ -\frac{15}{2}\left(\frac{2}{T}\right)^2 & 0 & 45\left(\frac{2}{T}\right)^4 \end{bmatrix} \tag{3.9}$$

With the approximate determinations of:

$$\sigma_{\alpha_o} \sim \frac{3\sigma_\alpha}{2\sqrt{N}} \tag{3.10}$$

$$\sigma_{\dot{\alpha}_o} \sim \frac{2\sqrt{3}\sigma_\alpha}{T\sqrt{N}} \tag{3.11}$$

$$\sigma_{\ddot{\alpha}_o} \sim \frac{12\sqrt{5}\sigma_\alpha}{T^2\sqrt{N}} \tag{3.12}$$

and a correlation between the angle and angular acceleration uncertainties. Now consider some published tracking data characteristics for the MODEST space surveillance telescope which tracks GEO objects [124]. A usual pass lasts for 5 minutes during which they take $N = 8$ observations. Several published reports indicate optical sensors have 1 arcsecond or 2.8×10^{-4} deg observation uncertainties, which we use for σ_α. Putting these numbers together, this implies that such a track of observations contains information on the angular location and angular rate of the object at epoch t_o with errors on the order of $\sigma_{\alpha_o} \sim 1.5 \times 10^{-4\circ}$, $\sigma_{\dot{\alpha}_o} \sim 1 \times 10^{-6\circ}/\text{s}$ and $\sigma_{\ddot{\alpha}_o} \sim 3 \times 10^{-8\circ}/\text{s}^2$.

This approach combines the information from a track of observations spread out in time and transforms it into a precise estimate of the partial state of the object at a specific epoch. This is a more convenient form in which to transform the information from the track and makes it easier to discuss constraints on the unmeasured components of the space object's state. For an optical observation these unmeasured components are the object's range and range-rate at the epoch t_o and define the attributable vector A.

3.3.3 The Admissible Region

The specific geocentric energy of the particle is

$$E = \frac{1}{2}||\dot{P}||^2 - \frac{\mu}{||P||}$$

where $\mu = GM_\oplus$ is the gravitational parameter of the Earth, $G = 19.91\ \mathrm{R}_\oplus^3\mathrm{M}_\oplus^{-1}\mathrm{hr}^{-1}$, $1\mathrm{R}_\oplus = 1$ earth radius, and $1\mathrm{M}_\oplus = 1$ earth mass.

Given an optical attributable vector A, its corresponding *admissible region* is the set of points on the $(\rho, \dot{\rho})$ plane that have not been ruled out by physical considerations. We impose the following physical constraints on the possible positions of the particle in the topocentric range/range-rate $(\rho, \dot{\rho})$ plane:

- $\mathcal{C}_1 = \{(\rho, \dot{\rho}) : E < 0\}$
- $\mathcal{C}_2 = \{(\rho, \dot{\rho}) : 2 < \rho < 20\}$
- $\mathcal{C}_3 = \{(\rho, \dot{\rho}) : 1.03 < r_p\}$
- $\mathcal{C}_4 = \{(\rho, \dot{\rho}) : r_a < 25\}$

where r_p and r_a are the periapsis and apoapsis (geocentric) radii of the orbit, respectively; and where distance is measured in units of Earth-radii. $\mathcal{C}_1$ and $\mathcal{C}_2$ are the constraints as presented in Tommei et al. [130] and $\mathcal{C}_3$ and $\mathcal{C}_4$ are two additional

physical constraints we place on the admissible region by constraining the periapsis and apoapsis (geocentric) radii of the orbit to always lie within some range. These latter constraints place a restriction on the possible eccentricities of the orbit, which rule out impact orbits and orbits with an extremely high apoapsis. A periapsis of 1.03 corresponds to a periapsis radius at about 200 km above the surface of the Earth. The admissible region is then defined as a subset of the topocentric range/range-rate plane by the condition:

$$\mathcal{C} = \bigcap_{i=1}^{4} \mathcal{C}_i \tag{3.13}$$

In order to compute the periapsis and apoapsis radii of the orbit, one must transfer to geocentric coordinates. In coordinates, we have:

$$\hat{R} = \langle \cos\alpha\cos\delta, \sin\alpha\cos\delta, \sin\delta \rangle, \qquad \hat{R}_\alpha := \frac{\partial \hat{R}}{\partial \alpha}, \qquad \hat{R}_\delta := \frac{\partial \hat{R}}{\partial \delta}$$

Then:

$$P = P_O + \rho\hat{R} \tag{3.14}$$

$$\dot{P} = \dot{P}_O + \dot{\rho}\hat{R} + \rho\dot{\alpha}\hat{R}_\alpha + \rho\dot{\delta}\hat{R}_\delta \tag{3.15}$$

To illustrate the improvement the additional constraint $\mathcal{C}_3 \cap \mathcal{C}_4$, let us consider the following example attributable vector. Suppose the optical observer's position in standard coordinates is polar angle $\Theta = \pi/3$ (measured as the polar angle from the north pole) and azimuthal angle $\Phi = 0$ (measured from inertial x-axis), and the observer makes the following zenith observation $A = (0, \pi/6, 0.1, 0.03)$. Fig. 3.1 shows the resulting admissible region. The outlined region is the admissible region presented in Tommei et al. [130], i.e. $\mathcal{C}_{\text{tom}} = \mathcal{C}_1 \cap \mathcal{C}_2$. The inside dotted region is the (discretized) admissible region as presented here, i.e. $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4$. As one can see, these additional constraints significantly reduces the area of the admissible region that one must consider in making the orbit determination.

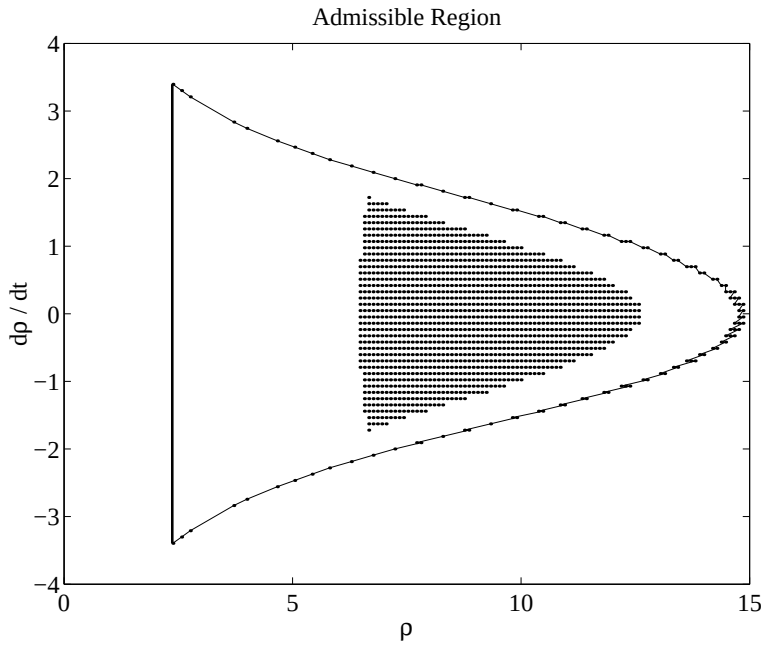

Figure 3.1: Admissible Region for attributable vector $A = (0, \pi/6, 0.1, 0.03)$, zenith measurement

3.4 Delaunay Variables

In this section, we will define the set of canonical Delaunay variables that we will use and then indicate how one would compute the corresponding Jacobian matrix of the transformation.

3.4.1 Transformation to Delaunay Variables

One can transfer the uncertainty surface (admissible region) into geocentric cartesian coordinates and then let each point of this surface evolve as a Keplerian orbit. If one does so, one sees that the surface spreads out fairly quickly. As an alternative, we will transfer the surface into Delaunay variables. This is done in several steps. We first transfer the topocentric spherical observation coordinates into geocentric cartesian coordinates using (3.14)-(3.15):

$$T_1 : \langle \rho, \dot{\rho}, \mathfrak{X} \rangle \rightarrow \langle x, y, z, \dot{x}, \dot{y}, \dot{z} \rangle$$

This transformation depends on the full set of recorded data $\mathfrak{X} = (A, t, h, \Theta, \Phi)$, as it depends on the location of the observer at that time. As usual, $A = \langle \alpha, \delta, \dot{\alpha}, \dot{\delta} \rangle$ is the admissible vector associated with a given track of data. Next, we transfer the cartesian coordinates into a set of orbital elements, using standard relations (see Danby [40], Roy [118], Crassidis & Junkins [37], Montenbruck & Gill [100], etc.).

$$T_2 : \langle x, y, z, \dot{x}, \dot{y}, \dot{z} \rangle \to \langle a, e, i, \omega, \Omega, M \rangle$$

where a is the semi-major axis, e is the eccentricity, i is the inclination, ω is the argument of periapsis, Ω is the longitude of the ascending node, and M is the mean anomaly. Lastly we transfer the orbital elements into Delaunay variables.

$$T_3 : \langle a, e, i, \omega, \Omega, M \rangle \to \langle L, l, G, g, H, h \rangle$$

where the Delaunay variables are defined as in Ferraz-Mello [50]:

$$\begin{aligned} l &= M, & L &= \sqrt{\mu a}, \\ g &= \omega, & G &= L\sqrt{1-e^2}, \\ h &= \Omega, & H &= G\cos i, \end{aligned} \tag{3.16}$$

so that the total transformation from the observation space to Delaunay space at the initial time t_0 is given by the composition:

$$T(t_0; t_0) = T_3 \circ T_2 \circ T_1 \tag{3.17}$$

The reason for this seemingly superfluous notation, i.e. $T(t_0; t_0)$, will be made clear by the end of the section. For the case of zero-eccentricity or zero-inclination orbits, the Delaunay variables become singular, and one could instead choose Poincaré nonsingular canonical variables:

$$T_3' : \langle a, e, i, \omega, \Omega, M \rangle \to \langle \mathfrak{l}, \mathfrak{L}, \mathfrak{g}, \mathfrak{G}, \mathfrak{h}, \mathfrak{H} \rangle$$

which are defined by the relations:

$$\begin{aligned} \mathfrak{l} &= M + \omega + \Omega, & \mathfrak{L} &= L = \sqrt{\mu a} \\ \mathfrak{g} &= \omega + \Omega & \mathfrak{G} &= G - L = L(\sqrt{1-e^2} - 1) \\ \mathfrak{h} &= \Omega & \mathfrak{H} &= H - G = G(\cos i - 1) \end{aligned}$$

Both the Delaunay variables and the Poincaré nonsingular variables are symplectic sets of coordinate-momenta pairs. The transformations $T_3 \circ T_2$ and $T_3' \circ T_2$ are therefore symplectomorphisms. In this work we will restrict our attention to cases where singularities are not present and one can use Delaunay variables.

We choose Delaunay variables because, like the orbital elements, five of them are constants of motion for the unperturbed Kepler problem. Additionally, unlike the orbital elements, there is a natural pairing of the Delaunay variables into coordinate-momenta symplectic pairs, i.e. the Delaunay variables $l,\ g$, and h are angle variables, to be modded by 2π, and $L,\ G$, and H are action variables, or the conjugate momenta. In terms of the Delaunay variables, the equations of motion of the system reduce to Hamilton's equations:

$$\begin{aligned} \frac{dl}{dt} &= \frac{\partial \mathcal{F}}{\partial L} & \frac{dL}{dt} &= -\frac{\partial \mathcal{F}}{\partial l} \\ \frac{dg}{dt} &= \frac{\partial \mathcal{F}}{\partial G} & \frac{dG}{dt} &= -\frac{\partial \mathcal{F}}{\partial g} \\ \frac{dh}{dt} &= \frac{\partial \mathcal{F}}{\partial H} & \frac{dH}{dt} &= -\frac{\partial \mathcal{F}}{\partial h} \end{aligned} \tag{3.18}$$

where

$$\mathcal{F} = -\frac{\mu^2}{2L^2} + \mathcal{R}(L, l, G, g, H, h) \tag{3.19}$$

and where $\mathcal{R}(L, l, G, g, H, h)$ is the disturbing force expressed in terms of Delaunay variables. For the unperturbed Kepler problem, the equations of motion work out

as:

$$\frac{dl}{dt} = \frac{\mu^2}{L^3}, \qquad \frac{dL}{dt} = \frac{dH}{dt} = \frac{dG}{dt} = \frac{dh}{dt} = \frac{dg}{dt} = 0 \tag{3.20}$$

For the general case, we can denote the Hamiltonian flow of (3.18) as:

$$T_4(t;t_0) : \langle L(t_0), l(t_0), G(t_0), g(t_0), H(t_0), h(t_0)\rangle \rightarrow \langle L(t), l(t), G(t), g(t), H(t), h(t)\rangle$$

For the Kepler problem, this simplifes to the form:

$$T_4^{\text{kepler}}(t;t_0)(L_0, l_0, G_0, g_0, H_0, h_0) = \left\langle L_0, \left(l_0 + \frac{\mu^2}{L_0^3}(t-t_0)\right), G_0, g_0, H_0, h_0 \right\rangle$$

The transformation from the observation space to the time-evolved Delaunay space is then given by:

$$T(t;t_0) = T_4(t;t_0) \circ T(t_0;t_0) \tag{3.21}$$

where $T(t_0;t_0)$ is defined in (3.17). This is well-defined because $T_4(t_0;t_0)$ is the identity transformation.

3.4.2 The Jacobian Matrix

As we shall see, the Jacobian Matrix of the transformation T will be used to compute variations in the Delaunay variables with respect to variations in the $(\rho, \dot{\rho})$ plane. Since the variables $(\alpha, \dot{\alpha}, \delta, \dot{\delta})$ are taken to be known, we only need consider the first two columns of the Jacobian of T. We will denote this 6×2 matrix as Φ. We construct Φ by composition. First we define:

$$\Phi_1 = \frac{\partial(x, y, z, \dot{x}, \dot{y}, \dot{z})}{\partial(\rho, \dot{\rho})} = \begin{bmatrix} \cos\alpha\cos\delta & 0 \\ \sin\alpha\cos\delta & 0 \\ \sin\delta & 0 \\ -\dot{\alpha}\sin\alpha\cos\delta - \dot{\delta}\cos\alpha\sin\delta & \cos\alpha\cos\delta \\ \dot{\alpha}\cos\alpha\cos\delta - \dot{\delta}\sin\alpha\sin\delta & \sin\alpha\cos\delta \\ \dot{\delta}\cos\delta & \sin\delta \end{bmatrix}$$

where the partial derivatives have been computed using the transformation relations (3.14)-(3.15).

Next, the Jacobian of the transformation T_2 is computed:

$$\Phi_2 = \frac{\partial(a, e, i, \omega, \Omega, M)}{\partial(x, y, z, \dot{x}, \dot{y}, \dot{z})}$$

A very efficient and elegant algorithm to compute this matrix is given in Montenbruck & Gill [100], §7.1.2 - §7.1.3. Since this computation is well known, we will not go through the details of it here.

The Jacobian of the transformation matrix T_3 is then computed:

$$\Phi_3 = \frac{\partial(L, l, G, g, H, h)}{\partial(a, e, i, \omega, \Omega, M)} = \begin{bmatrix} L_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ L_a\sqrt{1-e^2} & G_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ L_a\sqrt{1-e^2}\cos i & G_e\cos i & -G\sin i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

where

$$L_a = \frac{1}{2}\sqrt{\frac{\mu}{a}} \qquad \text{and} \qquad G_e = \frac{-Le}{\sqrt{1-e^2}}$$

Finally, the Jacobian corresponding to the time evolution of the system (3.18) can be determined by integrating the system of differential equations (2.7) (with Hamiltonian $H = \mathcal{F}$ given by (3.19)) to obtain:

$$\Phi_4(t; t_0) = \frac{\partial(L(t), l(t), G(t), g(t), H(t), h(t))}{\partial(L(t_0), l(t_0), G(t_0), g(t_0), H(t_0), h(t_0))}$$

This Jacobian Matrix, since it corresponds to a time evolution transformation, is also known as the State Transition Matrix (STM) of the evolution map. For the Kepler

case, this STM takes the simple analytic form:

$$\Phi_4^{\text{kepler}}(t;t_0) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -3\mu^2(t-t_0)/L_0^4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.22}$$

The 6×2 matrix Φ, which is the first two columns of the Jacobian of T, is then given by:

$$\Phi(t;t_0) = \Phi_4(t;t_0) \cdot \Phi_3 \cdot \Phi_2 \cdot \Phi_1$$

and it maps a variation in the admissible region to a variation in the Delaunay variables, i.e.

$$\begin{bmatrix} \delta L \\ \delta l \\ \delta G \\ \delta g \\ \delta H \\ \delta h \end{bmatrix} = \Phi_{6\times 2} \cdot \begin{bmatrix} \delta\rho \\ \delta\dot{\rho} \end{bmatrix}$$

Additionally, the following three 2×2 symplectic submatrices of $\Phi(t;t_0)$ will be considered:

$$\begin{aligned} \mathbb{L}(t;t_0) &= \begin{bmatrix} I_2 & 0_2 & 0_2 \end{bmatrix} \cdot \Phi(t;t_0) \\ \mathbb{G}(t;t_o) &= \begin{bmatrix} 0_2 & I_2 & 0_2 \end{bmatrix} \cdot \Phi(t;t_0) \\ \mathbb{H}(t;t_0) &= \begin{bmatrix} 0_2 & 0_2 & I_2 \end{bmatrix} \cdot \Phi(t;t_0) \end{aligned}$$

where 0_2 and I_2 are the 2×2 matrix of zeros and identity matrix, respectively. These are defined so that:

$$\begin{bmatrix} \delta L \\ \delta l \end{bmatrix} = \mathbb{L} \cdot \begin{bmatrix} \delta\rho \\ \delta\dot{\rho} \end{bmatrix}, \qquad \begin{bmatrix} \delta G \\ \delta g \end{bmatrix} = \mathbb{G} \cdot \begin{bmatrix} \delta\rho \\ \delta\dot{\rho} \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} \delta H \\ \delta h \end{bmatrix} = \mathbb{H} \cdot \begin{bmatrix} \delta\rho \\ \delta\dot{\rho} \end{bmatrix}$$

Notice that, because of the especially simple form $\Phi_4(t; t_0)$ takes in the Kepler case (3.22), the dynamic time evolution of the system *only affects the submatrix* $\mathbb{L}$ of the full STM $\Phi(t; t_0)$. Thus, once the matrices $\mathbb{G}$ and $\mathbb{H}$ are computed for a VD particle, they are constant in time. Note that they still depend on the VD particle, as they depend on $\alpha, \dot{\alpha}, \delta, \dot{\delta}, \rho, \dot{\rho}$, the location of the optical observer on Earth, and the time of the measurement (which gives the observer's position in *inertial* space).

3.4.3 Area Expansion of the Delaunay Map

In §2.3.2 we discussed the mathematical theory of the area expansion and projection factors that arise when mapping differential area elements from one space to another. We show how these factors can be computed in terms of determinants (for the case of the area projection factors) and Gram determinants (for the total area expansion factor) of the Jacobian matrix of the transformation. In this section we will apply these results to the map $F_{\mathfrak{X}} : \mathcal{C} \to \mathcal{D}$ from topocentric spherical coordinates (observation space) to the Delaunay canonical coordinates, cooresponding to an extended admissible vector $\mathfrak{X}$, where $\mathcal{C}$ is the associated admissible region defined in §3.3 and $\mathcal{D} \cong \mathbb{R}^6$ is the Delaunay space. We will consider the admissible vector $A = \langle \alpha, \delta, \dot{\alpha}, \dot{\delta} \rangle = \langle 0, \pi/6, 0.1, 0.03 \rangle$. The spherical coordinates of the observatory at the time of observation are given by $\Phi = 0.1$ and $\Theta = \pi/3 + 0.1$. This is *n*ot a zenith observation. Let Φ be the corresponding 6×2 Jacobian matrix of the map $F_{\mathfrak{X}}$. The admissible region $\mathcal{C}$ is shown in the bottom of Fig. 3.2. $F_{\mathfrak{X}}(\mathcal{C})$ is a two-dimensional manifold embedded in six-dimensional Delaunay space $\mathcal{D}$. The area expansion factor

is given by $\alpha(\rho, \dot{\rho}) = \sqrt{\mathfrak{G}(\Phi)} = \sqrt{\det(\Phi^T \cdot \Phi)}$. This is plotted versus the admissible region in Fig. 3.2. One notices there is a singularity in area expansion at the point

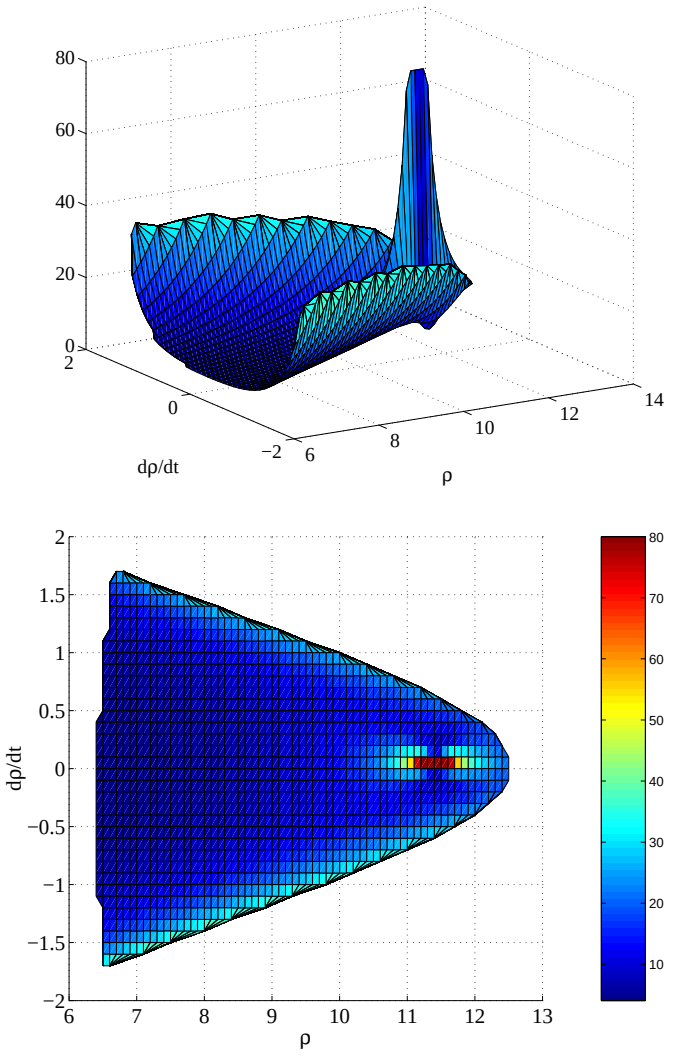

Figure 3.2: Area Expansion Factor vs. Admissible Region

given by a topocentric range of 11.5 Earth radii and a range-rate of 0 Earth radii per hour. (The area expansion factor was artificially cut-off at 80. Several of the grid points near the circular orbit singularity realized an area expansion factor as high as 200). The singularity in the area expansion factor near the circular orbit also exists in the map from the topocentric coordinates to the Poincaré elements as well.

Similarly we can look at the area projection factor, the signed area of the projection of the surface onto each of the Delaunay planes. First let us consider the projection onto the (L, l) plane. If we define Π_κ as in (2.11), this area projection factor is given by $\alpha_L(\rho, \dot{\rho}) = \det(\Pi_1^T \cdot \Phi)$. We plot this projection factor versus the admissible region *and* the projection of the virtual debris field on the (L, l) plane in Fig. 3.3. The

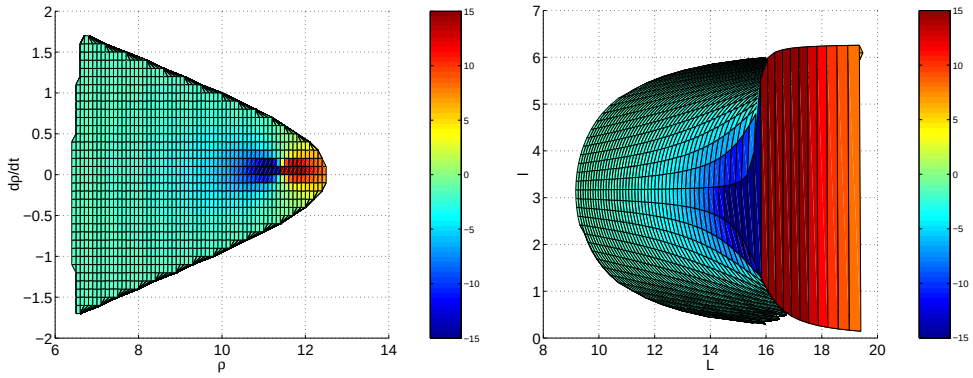

Figure 3.3: (L, l) Area Projection Factor

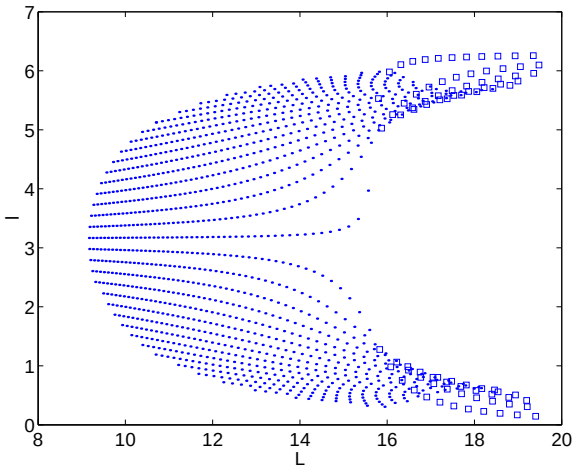

Figure 3.4: Image of VD field on (L, l) Delaunay Plane

actual locations of the individual virtual debris particles are given in Fig. 3.4. VD particles whose L-area projection factors are negative are represented by dots. VD particles with positive area projection factors are squares. The magnitudes of these

factors can be seen best in Fig. 3.3. One sees that the surface actualy "folds over" on the two wings. Notice that in the case of the (L, l) area *projection* factor, the circular orbit is actually a saddle point in $\alpha_L(\rho, \dot{\rho})$, and the singularity itself is bipolar. The circular orbit is the separation point between the positive area projection singularity and a negative area projection singularity.

Similarly one can consider the (G, g) area projection factor $\alpha_G(\rho, \dot{\rho}) = \det(\Pi_2^T \cdot \Phi)$. This is plotted in Fig. 3.5. The individual virtual debris particles are plotted in Fig. 3.6, with squares representing positive area and dots representing negative area.

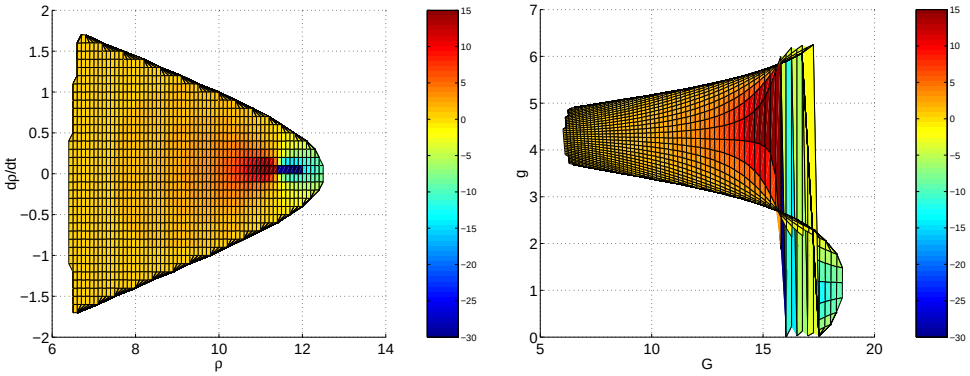

Figure 3.5: (G, g) Area Projection Factor

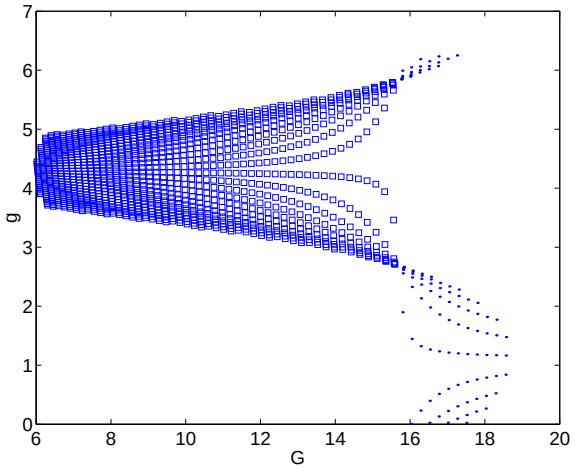

Figure 3.6: Image of VD field on (G, g) Delaunay Plane

The circular orbit singularity is again bipolar, separating a positive and negative singularity in the (G, g) area projection factor. Notice, *as must be the case*, the sign of the two singularities switch when comparing the two different projection factors. If ρ^* is the circular orbit radius, the *positive* singularity in $\alpha_L(\rho, \dot{\rho})$ occurs for $\rho > \rho^*$, whereas the *positive* singularity in $\alpha_G(\rho, \dot{\rho})$ occurs for $\rho < \rho^*$. The reason this "must be the case" is due to the symplecticity condition of the Jacobian matrix:

$$\det(\Pi_1^T \cdot \Phi) + \det(\Pi_2^T \cdot \Phi) + \det(\Pi_3^T \cdot \Phi) = 1$$

Fig. 3.7 shows a plot of the (H, h) area projection factor $\alpha_H(\rho, \dot{\rho}) = \det(\Pi_3^T \cdot \Phi)$. Notice that there is no circular orbit singularity in the area expansion projected onto the (H, h) plane.

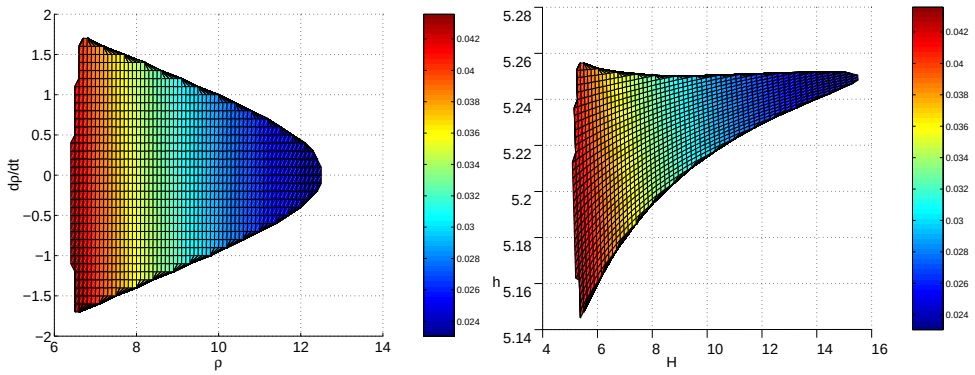

Figure 3.7: (H, h) Area Projection Factor

3.5 Intersection Theory Analysis (ITA)

In this section we describe how the Delaunay space can be used in fitting an orbit determination between two previously uncorrelated tracks (UCT's). In subsequent sections we will illustrate this technique for a sample set of observations, considering separately zenith and non-zenith observations, and additionally the effect of the J_2 term in Earth's potential on the proposed orbit determination process.

3.5.1 Mapping the Admissible Region to Delaunay Space

Let $\mathcal{D} \cong \mathbb{R}^6$ be the six-dimensional Delaunay space and $\mathcal{C}(\mathfrak{X})$ the corresponding admissible region in the topocentric range/range-rate plane for a fixed attributable vector and spatiotemporal observation location, as defined in (3.13). For a fixed epoch time τ, we define the map $F_{\mathfrak{X}}^{\tau} : \mathcal{C} \to \mathcal{D}$ as the restriction of the map $T(\tau; t)$ defined in (3.21) for a fixed $\mathfrak{X}$, where t is the time of the observation (one of the components of $\mathfrak{X}$), so that:

$$F_{\mathfrak{X}}^{\tau} : \langle \rho, \dot{\rho} \rangle \to \langle L, l, G, g, H, h \rangle$$

Thus $F_{\mathfrak{X}}^{\tau}(\mathcal{C}) \subset \mathcal{D}$ is a two-dimensional submanifold of Delaunay space $\mathcal{D} \cong \mathbb{R}^6$. We further define the three *Delaunay projective spaces* $\mathcal{D}_L \cong \mathcal{D}_G \cong \mathcal{D}_H \cong \mathbb{R}^2$, so that the Delaunay space $\mathcal{D}$ has the direct sum decomposition $\mathcal{D} = \mathcal{D}_L \oplus \mathcal{D}_G \oplus \mathcal{D}_H$. These three projective spaces are the projections of Delaunay space onto the three symplectic Delaunay planes. If Π_L, Π_G, and Π_H are the projection operators from the full six-dimensional Delaunay space onto the symplectic Delaunay planes, so that $\Pi_L(L, l, G, g, H, h) = \langle L, l \rangle$, etc., then $\mathcal{D}_L = \Pi_L(\mathcal{D})$, $\mathcal{D}_G = \Pi_G(\mathcal{D})$, and $\mathcal{D}_H = \Pi_H(\mathcal{D})$. The map $F_{\mathfrak{X}}^{\tau}$ can thus be thought of in either of two ways, as a one-to-one mapping from the two-dimensional admissible region $\mathcal{C}$ to the six-dimensional Delaunay space $\mathcal{D}$ or as a one-to-three mapping from the two-dimensional admissible region $\mathcal{C}$ to the three two-dimensional Delaunay projective spaces $\mathcal{D}_L$, $\mathcal{D}_G$, and $\mathcal{D}_H$. While at first glance, such a distinction seems pedantic, it is actually an important one, as intersections are inherently easier to both visualize and carry out in two-dimensional spaces then they are in six-dimensional spaces.

3.5.2 Necessary Conditions for Correlation between Two UCT's

As discussed previously, each uncorrelated track (UCT) provides a set of data $\mathfrak{X}$ that contains the attributable vector and information on when and where the observatory was at the time of the measurements of the track. The Space Situational Awareness (SSA) problem is to determine which UCT's belong to the same physical object, carry out an orbit determination for each orbit, and then to add known orbits to the space debris catalog. In this section we discuss how to compare pairwise UCT's to determine whether they correlate to the same object.

Given two extended attributable vectors $\mathfrak{X}_1$ and $\mathfrak{X}_2$, one determines the corresponding admissible regions $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. These admissible regions cannot be compared directly, as they are subsets of two different sets of topocentric spherical coordinates, affixed to the Earth at different locations and different times. Even if both attributable vectors were recorded by the same observatory, if the time between the two tracks, modulo 24 hours, is not zero, that single observatory would have been at two different positions in inertial space when it measured the two different tracks. We now push the admissible regions forward into Delaunay space, and dynamically evolve or regress both uncertainty distributions in time to a common epoch τ, so that $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)$ are both two-dimensional submanifolds of six-dimensional Delaunay space $\mathcal{D}$, dynamically mapped to a common epoch time.

If $\mathfrak{X}_1$ and $\mathfrak{X}_2$ correspond to the same object, then $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1) \bigcap F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2) \neq \emptyset$.

Since $\mathfrak{X}_1$ and $\mathfrak{X}_2$ each contain four pieces of information (two angles and two angle rates), the system is overdetermined. Unless there is some redundancy in the information, if both tracks correspond to the same physical object, it is likely that the uncertainty manifolds $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)$ will intersect at a single point.

Suppose now that this is the case, i.e. the intersection $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1) \bigcap F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2) = \{\boldsymbol{\Delta}^*\}$

is at the single point $\boldsymbol{\Delta}^* \in \mathcal{D}$. Then the two extended attributable vectors $\mathfrak{X}_1$ and $\mathfrak{X}_2$ corresponding to the two separate UCT's do *not necessarily belong* to the same object. Rather one can say two things. If they belong to the same object, that object's orbit is given by $\boldsymbol{\Delta}^*$. Secondly, they almost certainly do belong to the same object. This is more deeply expounded upon over the next several paragraphs.

They do not necessarily correlate to the same object for the following reason. The orbit corresponding to the first UCT can still lie anywhere on $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and the orbit corresponding to the second UCT can still lie anywhere on $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$. We do not yet know the two UCT's correlate to the same object, so even though both uncertainty manifolds intersect at a single point, this may not be the correct orbit for either object.

They almost certainly do belong to the same object for the following reason. $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ are two separate two-dimensional manifolds embedded into the same six-dimensional Delaunay space $\mathcal{D}$. The probability that they, *by accident*, happen to touch tangentially at a single intersection point $\boldsymbol{\Delta}^*$ is extremely low, unless they are correlated and the orbit for both objects is given by that common intersection point $\boldsymbol{\Delta}^*$. One can therefore, with great confidence, make a preliminary orbit determination from $\boldsymbol{\Delta}^*$ and consider the two tracks as correlated. One then places this preliminary orbit determination into a separate holding catalog and awaits confirmation by a third consistent track of data, at which time the orbit is added to the standard catalog.

The topic of determining this intersection point is taken up in the following subsection.

3.5.3 Intersection Theory Analysis

Supposing two extended attributable vectors $\mathfrak{X}_1$ and $\mathfrak{X}_2$ correspond to the same physical object, a unique intersection point $\mathbf{\Delta}^*$ of the two submanifolds $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ can be found. To do this, we will consider the three-fold projection of these surfaces onto the Delaunay planes. To reduce complexity of notation let us define

$$\begin{aligned} S^i_L &= \Pi_L(F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i)) \subset \mathcal{D}_L \\ S^i_G &= \Pi_G(F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i)) \subset \mathcal{D}_G \\ S^i_H &= \Pi_H(F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i)) \subset \mathcal{D}_H \end{aligned}$$

For each of the symplectic Delaunay surface projections, there is an overlap region

$$S^1_L \cap S^2_L \qquad S^1_G \cap S^2_G \qquad S^1_H \cap S^2_H$$

See, for example, Fig. 3.14 (the overlap regions are not highlighted) in §3.7. We know that if the two tracks were of the same object, the true orbit must be in each of these intersected regions, i.e. $\Pi_L(\mathbf{\Delta}^*) \in S^1_L \cap S^2_L$ and similarly for the G and H projections. However, since each projection is a unique view of the same two two-dimensional surfaces, more information can be extracted. To obtain the unique intersection point, one carries out the following algorithm, which we have named Intersection Theory Analysis (ITA):

1. Select a Delaunay plane P, where $P \in \{\text{“}L\text{”}, \text{“}G\text{”}, \text{“}H\text{”}\} \subset$ the alphabet.

2. The projection of $\mathbf{\Delta}^*$ must lie in the intersection of the two projected uncertainty surfaces, i.e. $\mathbf{\Delta}^* \in S^1_P \cap S^2_P$. Both projections S^1_P and S^2_P are discretized by a population of virtual debris (VD) particles that have been mapped into the Delaunay space. Omit all VD particles that do not lie in the overlap region.

Now define

$$\diamond F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i) \subset F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i) \subset \mathcal{D}$$

by the relation

$$\Pi_P(\diamond F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i)) = S^1_P \cap S^2_P$$

3. Now reproject both surfaces $\diamond F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $\diamond F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)$ onto the Delaunay planes. Define their projections as:

$$\begin{aligned} \diamond S^i_L &= \Pi_L(\diamond F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i)) \subset \mathcal{D}_L \\ \diamond S^i_G &= \Pi_G(\diamond F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i)) \subset \mathcal{D}_G \\ \diamond S^i_H &= \Pi_H(\diamond F^{\tau}_{\mathfrak{X}_i}(\mathcal{C}_i)) \subset \mathcal{D}_H \end{aligned}$$

   See Fig. 3.15 for an example of this. Notice that, in this figure, all nonoverlap points on the Delaunay plane $\mathcal{D}_H$ have been omitted.

4. Repeat Steps 1-3. For step 1 choose a different P. It is alright if you've used that P before, as long as you do not use the same P twice. For steps 2-3, add an extra diamond to each strand of diamonds to indicate that an additional reduction has taken place. Continue until you are left with a single (approximate) intersection point.

The diamond operator is identified with omitting all nonoverlap regions in a particular Delaunay plane. The process of orbit determination is therefore reduced to the development of an efficient computer algorithm that will determine this overlap region for two overlapping discretized laminas on $\mathbb{R}^2$. This procedure is illustrated over the next several sections of the paper, as an initial feasibility study of the ITA algorithm. For the current work, the overlap regions were computed by manual computation and by trial and error. Creating a computer algorithm that determines these overlap regions automatically will be a focus of future research.

3.6 Orbit Determination I: Kepler Orbit with Two Zenith Observations

3.6.1 Concurrent Plot of Two Zenith Observations on the Delaunay Planes

In this section we consider the admissible region corresponding to the zenith observation $A = (0, \pi/6, 0.1, 0.03)$, made at $t = 0$ from a point on the Earth's surface $\Theta = \pi/3$, $\Phi = 0$. The admissible region of the $(\rho, \dot{\rho})$ plane is the inner discretized region plotted in Fig. 3.1. Each district of the discretization is referred to as a VD particle. We will think of the uncertainty region as a two-dimensional surface in a six-dimensional space:

$$\mathcal{A} := \{\langle \rho, \dot{\rho}, \alpha, \delta, \dot{\alpha}, \dot{\delta} \rangle : (\alpha, \delta, \dot{\alpha}, \dot{\delta}) = A \text{ and } (\rho, \dot{\rho}) \in \mathcal{C}\}$$

where $\mathcal{C}$ is the admissible region defined in (3.13). We call $\mathcal{C}$ the admissible region of the $(\rho, \dot{\rho})$ plane and $\mathcal{A}$ the admissible region of the observation space. Define:

$$T_1\{\mathcal{A}\} := \{\langle r, \dot{r} \rangle \in \mathbb{R}^6 : T_1^{-1}(\langle r, \dot{r} \rangle) \in \mathcal{A}\},$$

i.e. $T_1\{\mathcal{A}\}$ is the image of $\mathcal{A}$ under the mapping T_1. Then $T_1\{\mathcal{A}\}$ is a two-dimensional surface in the geocentric cartesian phase space. Similarly, we define:

$$T(t; t_0)\{\mathcal{A}\} := \{x = \langle L, l, G, g, H, h \rangle \in \mathbb{R}^6 : T(t; t_0)^{-1} \cdot x \in \mathcal{A}\}$$

where $T(t; t_0)$ is defined by the relations in §3.4.1. $T(t; t_0)\{\mathcal{A}\}$ is a two-dimensional surface in Delaunay space. Since $T_1\{\mathcal{A}\}$ is allowed to move about relatively freely, fold, wrap around the planet, etc., we will choose to follow the dynamic evolution of the surface $T(t; t_0)\{\mathcal{A}\}$ instead, which has a much more restricted evolution.

The projections of $T(0; 0)\{\mathcal{A}\}$ onto the three symplectic Delaunay planes are shown in Fig. 3.8. Since the phase flow is Hamiltonian and the Delaunay variables are a set of coordinate-momentum symplectic pairs, the sum of the oriented area projections onto the Delaunay planes is conserved, Arnold [6], Marsden & Ratiu

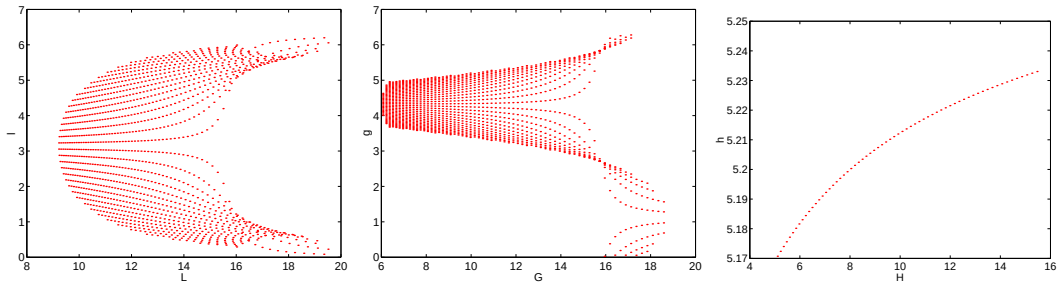

Figure 3.8: $T(0;0)\{\mathcal{A}\}$ projected onto the Delaunay planes

[83]. Since all of the Delaunay variables except for l are constants of motion for the Kepler problem, the total area projection on the $(L-l)$ Delaunay plane will be conserved, unless the surface "folds over," as was discussed in Scheeres et al. [123]. Because the observation was made directly overhead, a certain degeneracy exists that causes the projection of the uncertainty region on the (H,h) Delaunay plane to be a line, as in Fig. 3.8.

The dynamics for the Kepler problem, in terms of Delaunay variables, is governed by the equations of motion (3.20). In particular, $dl/dt \propto L^{-3}$. All of the dots in the (L,l) plane will march up the graph. Since the angle l is given modulo 2π, when a dot reaches $l = 2\pi$, it is reset to $l = 0$. The regions for smaller L will move at a greater constant rate than the regions for larger L. In this way a shearing effect takes place. The dynamics literally shreds the region into thin strips. The longer you wait, the more thin strips the uncertainty region will be cut up into. After 70 hrs, the surface $T(70;0)\{\mathcal{A}\}$ is projected onto the symplectic Delaunay planes, and is shown in Fig. 3.9. Notice the projections onto the (G,g) and (H,h) plane are unchanged.

At time $t = 70$ hrs we will assume that we have another zenith observation of the same particle of space debris. MATLAB randomly selected VD field particle #893 to correspond to the actual physical piece of debris. If VD particle #893 were to be observed again at time $t = 70$ by an Earthbound optical observer with

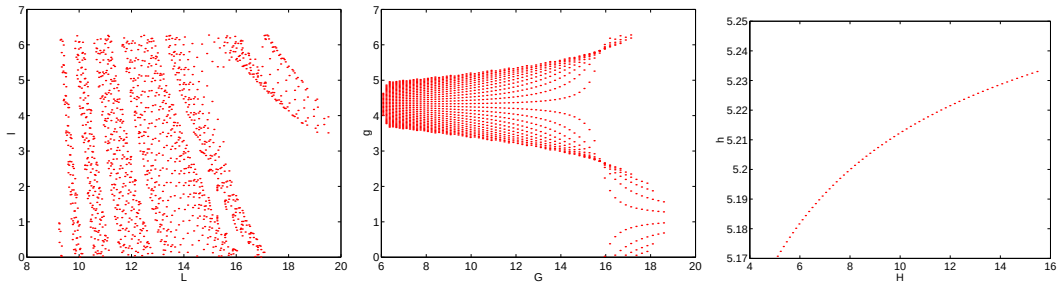

Figure 3.9: $T(70;0)\{\mathcal{A}\}$ projected onto the Delaunay planes

inertial polar angle $\Theta = 1.1650$ (measured from the north pole) and azimuthal angle $\Phi = 5.9214$ (measured from the inertial x-axis), its attributable vector would be $A_{70} = (\alpha, \delta, \dot{\alpha}, \dot{\delta}) = (-0.3618, 0.4058, 0.0315, 0.0209)$. Since $\alpha = \Phi - 2\pi$ and $\delta = \pi/2 - \Theta$, this is again a zenith observation. For this second observation, $t_0 = 70$, so the initial transformation of the admissible region of the observation space $\mathcal{A}_{70}$ to the Delaunay space would be $T(70;70)\{\mathcal{A}_{70}\}$. $\mathcal{A}_{70}$ is the *new* observation recorded at time $t = 70$. $T(70;0)\{\mathcal{A}\}$ and $T(70;70)\{\mathcal{A}_{70}\}$ are plotted concurrently on each of the Delaunay planes in Fig. 3.10.

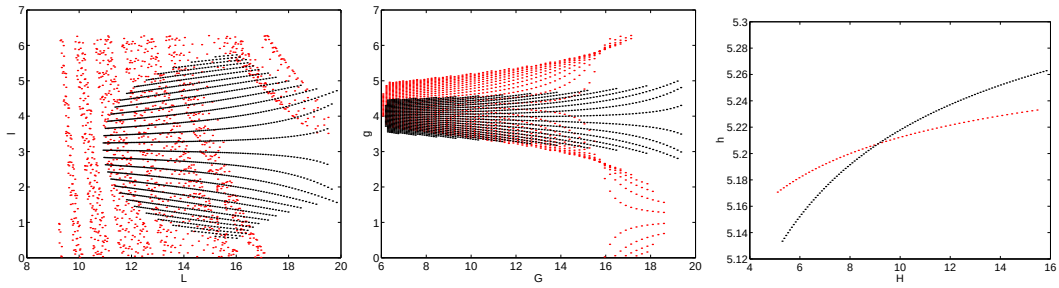

Figure 3.10: $T(70;0)\{\mathcal{A}\}$ (red) and $T(70;70)\{\mathcal{A}_{70}\}$ (black) projected onto the Delaunay planes.

The intersection of the two lines in the $(H-h)$ plane in Fig. 3.10 indicates that the Delaunay variables H and h can be determined exactly. This reduces the uncertainty region to a one-dimensional uncertainty curve, since each point on each line in the (H, h) plane is a curve in the original $(\rho, \dot{\rho})$ admissible region.

3.6.2 Determining the Intersection Point on the $(H - h)$ Plane

As earlier noted, if both observations are zenith observations, the uncertainty region projected onto the $(H - h)$ plane will degenerate to a single line, for both observations. We will therefore begin by determining this intersection point, which will pinpoint the values of H and h that belong to the true debris particle. We will use the Jacobian matrix derived in §3.4.2 as part of a predictor-corrector method in determining this intersection point.

One would have to be fairly lucky to by chance have discretized the initial $(\rho, \dot{\rho})$ plane so that $T(70; 0)\{\mathcal{A}\}$ has a point *exactly* on the intersection in the (H, h) plane. If we were to zoom in on the intersection point in Fig. 3.10, we would be more likely to see something as in Fig. 3.11. Here the red points are the images of VD particles from the initial observation, projected onto the (H, h) plane, i.e. they are points from the set $T(70; 0)\{\mathcal{A}\}$. Similarly, the black points are from the new observation, i.e. they are from the set $T(70; 70)\{\mathcal{A}_{70}\}$. Due to the degeneracy that exists for zenith observations, the preimage of each point on the (H, h) plane is actually a curve in the initial topocentric admissible region in the $(\rho, \dot{\rho})$ space. Our first goal is to determine *a single point* in each admissible region that maps to the intersection point in the Delaunay projective space $\mathcal{D}_H$. We will show how to use this single point to generate the full curve in the admissible region that projects onto this intersection point in the next subsection.

For convenience we will use the coordinates $(\rho, \dot{\rho})$ as coordinates for the initial admissible region $\mathcal{C}$ belonging to the first observation and the coordinates $(\varrho, \dot{\varrho})$ as coordinates for the second admissible region $\mathcal{C}_{70}$ belonging to the second observation.

We begin by choosing an initial guess. We take one of the intersection point's neighboring points $\langle H_0, h_0 \rangle \in T(70; 0)\{\mathcal{A}\}$, which is mapped from $\langle \rho, \dot{\rho} \rangle \in \mathcal{C}$. The

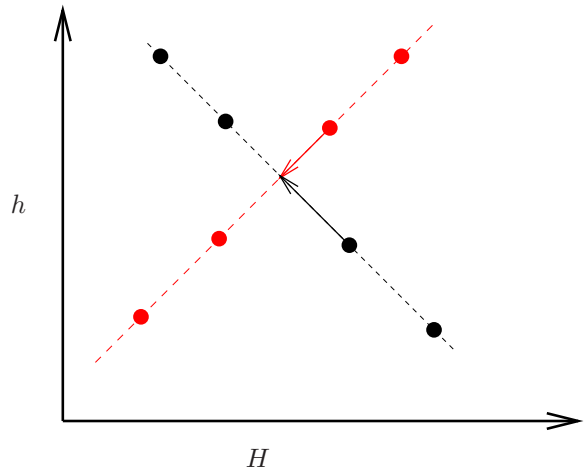

Figure 3.11: Example Schematic for Locating the Intersection Point on the $(H-h)$ Plane

partial Jacobian matrix $\mathbb{H}$ tells us the effect of varying $\langle \rho, \dot{\rho} \rangle$ on the point $\langle H, h \rangle$, i.e. for the black points on Fig. 3.11:

$$\begin{bmatrix} \delta H_0 \\ \delta h_0 \end{bmatrix} = \mathbb{H}_0 \begin{bmatrix} \delta \rho \\ \delta \dot{\rho} \end{bmatrix}$$

Similarly for the second observation, we will consider the nearby point $\langle H_{70}, h_{70} \rangle$ which belongs to the point $\langle \varrho, \dot{\varrho} \rangle$ of the admissible region. Thus, for the red points on Fig. 3.11:

$$\begin{bmatrix} \delta H_{70} \\ \delta h_{70} \end{bmatrix} = \mathbb{H}_{70} \begin{bmatrix} \delta \varrho \\ \delta \dot{\varrho} \end{bmatrix}$$

So the game now is to find a set of points on both admissible regions, $\langle \delta\rho, \delta\dot{\rho} \rangle$ and $\langle \delta\varrho, \delta\dot{\varrho} \rangle$, so that their images under the mapping $T(70;0)$ and $T(70;70)$, respectively, lies on the intersection point. The condition is:

$$\begin{bmatrix} H_{70} \\ h_{70} \end{bmatrix} + \mathbb{H}_{70} \cdot \begin{bmatrix} \delta \varrho \\ \delta \dot{\varrho} \end{bmatrix} = \begin{bmatrix} H_0 \\ h_0 \end{bmatrix} + \mathbb{H}_0 \cdot \begin{bmatrix} \delta \rho \\ \delta \dot{\rho} \end{bmatrix}$$

This can be rearranged as follows:

$$\begin{bmatrix} \Delta H \\ \Delta h \end{bmatrix} := \begin{bmatrix} H_{70} - H_0 \\ h_{70} - h_0 \end{bmatrix} = \mathbb{H}_0 \cdot \begin{bmatrix} \delta \rho \\ \delta \dot{\rho} \end{bmatrix} - \mathbb{H}_{70} \cdot \begin{bmatrix} \delta \varrho \\ \delta \dot{\varrho} \end{bmatrix} \tag{3.23}$$

Since the image of the admissible region $\mathcal{A}$ under the transformation $T(t;t_0)$ is a line when projected onto the (H,h) plane, both state transition matrices $\mathbb{H}_0$ and $\mathbb{H}_{70}$ have a single zero eigenvector. Due to this degeneracy, there is no one unique predictor vector. However, we are only looking for a single point in the admissible region that corresponds to the intersection point on the (H,h) plane. The easiest approach is to do the following. If $\langle 1,0\rangle^T \notin \text{null}(\mathbb{H}_0)$, we take $\delta\dot{\rho}=0$; otherwise we take $\delta\rho=0$. Additionally, if $\langle 1,0\rangle^T \notin \text{null}(\mathbb{H}_{70})$, we take $\delta\dot{\varrho}=0$; otherwise we take $\delta\varrho=0$. For considerations here, we will assume both $\langle 1,0\rangle^T \notin \text{null}(\mathbb{H}_0)$ and $\langle 1,0\rangle^T \notin \text{null}(\mathbb{H}_{70})$, so that we can take $\delta\dot{\rho}=0$ and $\delta\dot{\varrho}=0$. If either of these conditions fails, the procedure presented here can be easily modified accordingly.

Begin by defining $\mathbf{z}=\langle \Delta H,\Delta h\rangle^T$. Let $\mathbf{h}_0$ and $\mathbf{h}_{70}$ be the first columns of the matrices $\mathbb{H}_0$ and $\mathbb{H}_{70}$, respectively. Let $\mathbf{h}_0^\perp$ and $\mathbf{h}_{70}^\perp$ be unit vectors perpendicular to the vectors $\mathbf{h}_0$ and $\mathbf{h}_{70}$, respectively. Then by prescribing the conditions $\delta\dot{\rho}=0$ and $\delta\dot{\varrho}=0$, (3.23) reduces to:

$$\mathbf{z}=(\delta\rho)\mathbf{h}_0-(\delta\varrho)\mathbf{h}_{70}$$

Dotting this equation with $\mathbf{h}_{70}^\perp$ and solving for $\delta\rho$ we obtain:

$$\delta\rho=\frac{\mathbf{h}_{70}^\perp\cdot\mathbf{z}}{\mathbf{h}_{70}^\perp\cdot\mathbf{h}_0}$$

Similarly, by dotting with $\mathbf{h}_0^\perp$, we can obtain the following for $\delta\varrho$:

$$\delta\varrho=-\frac{\mathbf{h}_0^\perp\cdot\mathbf{z}}{\mathbf{h}_0^\perp\cdot\mathbf{h}_{70}}$$

This now gives us a new approximation for the intersection point. We reapply as necessary.

Both of the 2×2 $\mathbb{H}$-matrices will have a degeneracy in the form of a zero eigenvalue. The corresponding eigenvector we call the *zero eigenvector*. The zero eigenvector

itself is *not* the zero vector, rather it is the eigenvector that corresponds to the zero eigenvalue, i.e. the vector whose span is the null space of $\mathbb{H}$. $\mathbb{H}$ clearly has a zero eigenvector, if the initial observation is made at zenith, because the two-dimensional uncertainty region $\mathcal{C}$ on the $(\rho, \dot{\rho})$ plane reduces to a one-dimensional line on the (H, h) plane. Suppose $\xi(\rho, \dot{\rho})$ is the zero eigenvector of $\mathbb{H}$ at $(\rho, \dot{\rho}) \in \mathcal{C}$. The above algorithm provides a single point $(\rho^*, \dot{\rho}^*) \in \mathcal{C}$ that maps to the intersection point on the (H, h) plane. There exists a one-dimensional curve $\gamma(s) : (E \subset \mathbb{R}) \to \mathcal{C}$, such that $\gamma(0) = (\rho^*, \dot{\rho}^*)$, $\gamma(s) = (\rho(s), \dot{\rho}(s))$, and such that the projection onto the (H, h) plane of the image of γ under the mapping $T(70; 0)$ is the single intersection point of the two admissible curves on the (H, h) plane. The curve γ is then generated by the condition that $\gamma'(s) = \xi(\rho(s), \dot{\rho}(s)),$ for all $s \in E$. So to generate a discretized sequence of points along γ, we integrate the zero-eigenvector of $\mathbb{H}$, starting from $(\rho^*, \dot{\rho}^*)$, until the curve exits the admissible region $\mathcal{C}$. In this way, once we determine a single point $(\rho^*, \dot{\rho}^*)$ on the admissible region $\mathcal{C}$ that corresponds to the intersection point in (H, h) space, a reduced admissible region $\mathcal{C}^r \subset \mathcal{C}$ can then be defined.

If the observation is made at zenith, we will have that $\dot{\rho} = \dot{r}$, i.e. the rate of change of the radial coordinate in the frame attached to the observation location will coincide with the rate of change of the radial coordinate in geocentric spherical coordinates. Since the debris particle's angular momentum is independent of $\dot{r}$, we find that the Delaunay variables G, H, and h will all be independent of $\dot{\rho}$. Because of this, zenith observations will have the property that $\mathbb{H}$ will have a constant zero eigenvector of $\langle 0, 1 \rangle$, throughout the admissible region $\mathcal{C}$. As a consequence of the above theorem, the reduced admissible region will be the intersection of the vertical line $\rho = \rho^*$ with the admissible region $\mathcal{C}$. We call the reduced admissible region $\mathcal{C}^r$

and likewise define

$$\mathcal{A}^r := \{\langle \rho, \dot{\rho}, \alpha, \delta, \dot{\alpha}, \dot{\delta} \rangle : (\alpha, \delta, \dot{\alpha}, \dot{\delta}) = A \text{ and } (\rho, \dot{\rho}) \in \mathcal{C}^r\}$$

to be the (one dimensional) admissible region of the observation space. By construction, the projection of $T(t; t_0)\{\mathcal{A}^r\}$ on the (H, h) plane will correspond to a single point: the intersection point as seen in Fig. 3.10 and Fig. 3.11. A similar statement can be made about $\mathcal{C}_{70}$ and $\mathcal{A}_{70}$, which are defined analagously for the second observation.

The reduced admissible regions $\mathcal{C}^r$ and $\mathcal{C}^r$ for the initial and second observation are plotted in Fig. 3.12. Their images under the transformation T, projected onto each of the three Delaunay planes, is shown in Fig. 3.13.

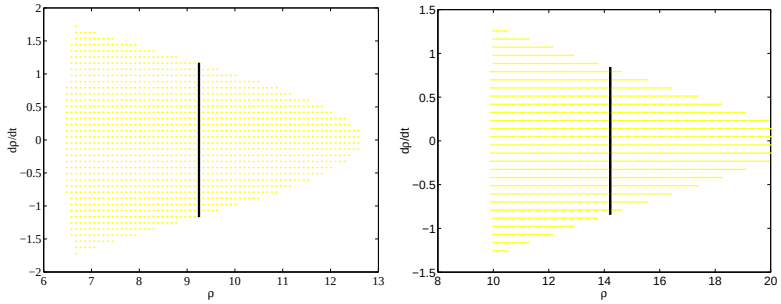

Figure 3.12: $\mathcal{C}$ (yellow) with $\mathcal{C}^r$ (black), left; $\mathcal{C}_{70}$ (yellow) with $\mathcal{C}^r_{70}$ (black), right

3.6.3 Orbit Determination

By reducing the admissible regions $\mathcal{C}$ and $\mathcal{C}_{70}$ to the preimage of the intersection point on the (H, h) Delaunay plane, we found that the dimensionality of the admissible region can be reduced from 2 to 1. One can see that an additional reduction can be made by considering the projection of these curves on the (L, l) Delaunay plane, as seen in Fig. 3.13. These curves have five distinct intersection points on the (L, l) plane, therefore the actual uncertainty distribution has been reduced from a two-dimensional sheet to that of five distinct points in phase space. Additional

information is also available from the overlap region of the two reduced curves as projected onto the (G, g) plane. In the case we are considering, only one of the five intersection points on the (L, l) plane actually lines up as an intersection point on the (G, g) plane. Therefore, the orbit is determined uniquely.

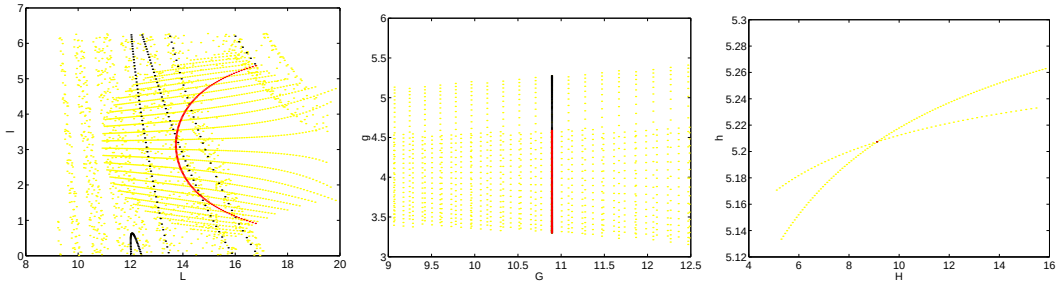

Figure 3.13: Fig. 3.10 plots (yellow) with $T(70;0)\{\mathcal{A}^r\}$ (red) and $T(70;70)\{\mathcal{A}^r_{70}\}$ (black)

3.7 Orbit Determination II: Kepler Orbit with Two Near-Zenith Observations

In this section we will study the case of two near-zenith observations. If the observation is not made when the space debris particle is directly overhead, the projection of the uncertainty region on the (H, h) plane will no longer be one-dimensional. We will consider the same 2 attributable vectors that were observed in §3.6, but nudge the inertial location of the observer so that the observations do not correspond to zenith observations. The first attributable vector is given by $A = (0, \pi/6, 0.1, 0.03)$, made at time $t = 0$ from a point on the Earth's surface $\Theta = \pi/3+0.1$, $\Phi = 0.1$. Assuming particle #1000 is the true space debris particle, a possible second observation (nonzenith) might be given by the attributable vector $A_{70} = (1.1516, 0.4790, 0.2262, -0.0809)$, made at time $t = 70$ from a point on the Earth's surface $\Theta = 1.2516$, $\Phi = 1.1918$. The intersections of the admissible regions, as projected onto the Delaunay planes, is shown in Fig. 3.14. Because the true debris particle did not fly directly over zenith on either of the measurements, the admissible regions now have two-dimensional pro-

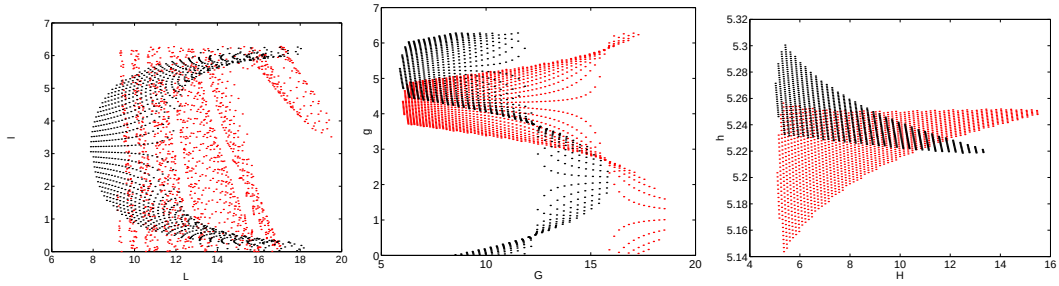

Figure 3.14: $T(70;0)\{\mathcal{A}\}$ (red) and $T(70;70)\{\mathcal{A}_{70}\}$ (black) projected onto the Delaunay planes, nonzenith observations

jections on the (H,h) plane. Since we are considering the Kepler case, the original uncertainty distribution's projections on the (G,g) and (H,h) plane are static. Our goal now is to systematically reduce the uncertainty region, by considering each Delaunay plane in sequence, as much as possible until it is reduced to either a single point (complete orbit determination) or a one-dimensional line.

By examination of the concurrent Delaunay plots of the uncertainty region projections (Fig. 3.14), we choose to begin the orbit determination process by cutting off the non-overlap sections of the surface in the (H,h) plane. The Delaunay projections of the remaining piece of surface is shown in Fig. 3.15. We see in Fig. 3.15

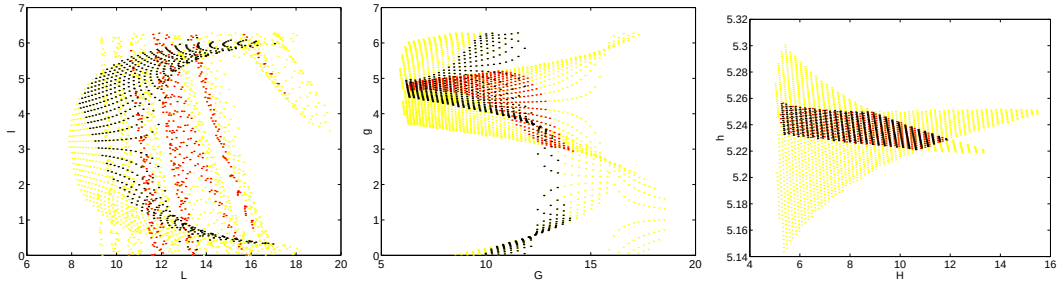

Figure 3.15: $T(70;0)\{\mathcal{A}^r\}$ (red) and $T(70;70)\{\mathcal{A}^r_{70}\}$ (black) projected onto the Delaunay planes, nonzenith observations

that there is again an overlap and non-overlap region in the (G,g) plane. Removing the non-overlap region further reduces the admissible region, as shown in Fig. 3.16. Interestingly, the (H,h) projection can be again used to cut out more of the uncer-

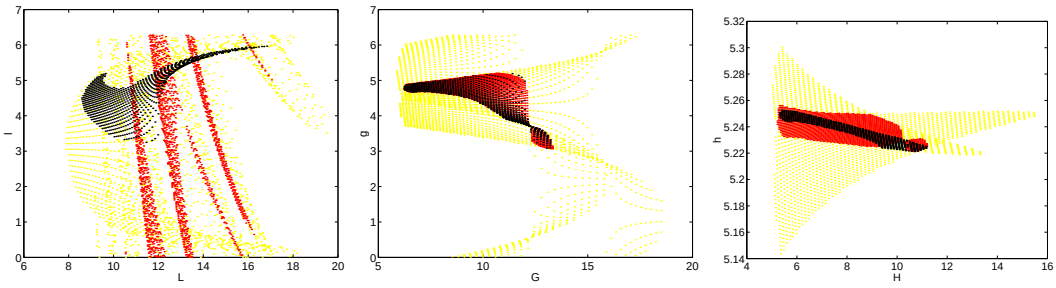

Figure 3.16: $T(70;0)\{\mathcal{A}^{rr}\}$ (red) and $T(70;70)\{\mathcal{A}^{rr}_{70}\}$ (black) projected onto the Delaunay planes, nonzenith observations

tainty surface, resulting in a third reduction, as shown in Fig. 3.17. We now turn to

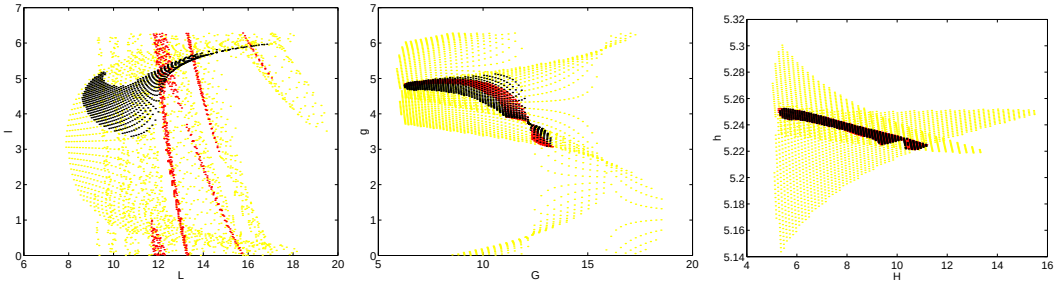

Figure 3.17: $T(70;0)\{\mathcal{A}^{rrr}\}$ (red) and $T(70;70)\{\mathcal{A}^{rrr}_{70}\}$ (black) projected onto the Delaunay planes, nonzenith observations

the projection of the thrice reduced uncertainty region on the (L, l) plane. The first three reductions have eliminated all but three overlap regions on the (L, l) plane. We consider each in term. The systematic projection of each overlap region onto each of the three Delaunay planes is shown in Fig. 3.18. We see that the far right overlap region on the (L, l) plane (the overlap that is almost confined to a single point) does not overlap on the (G, g) plane. This overlap region thus cannot correspond to the actual debris particle and is now ruled out. The middle overlap region on the (L, l) plane does not overlap on the (G, g) or (H, h) plane, so it is ruled out. Finally, the leftmost overlap region has a small intersection on both the (G, g) and (H, h) plane. The orbit is thus determined to within a small uncertainty about a single point in Delaunay space. It is possible that further reductions can be made by continuing this

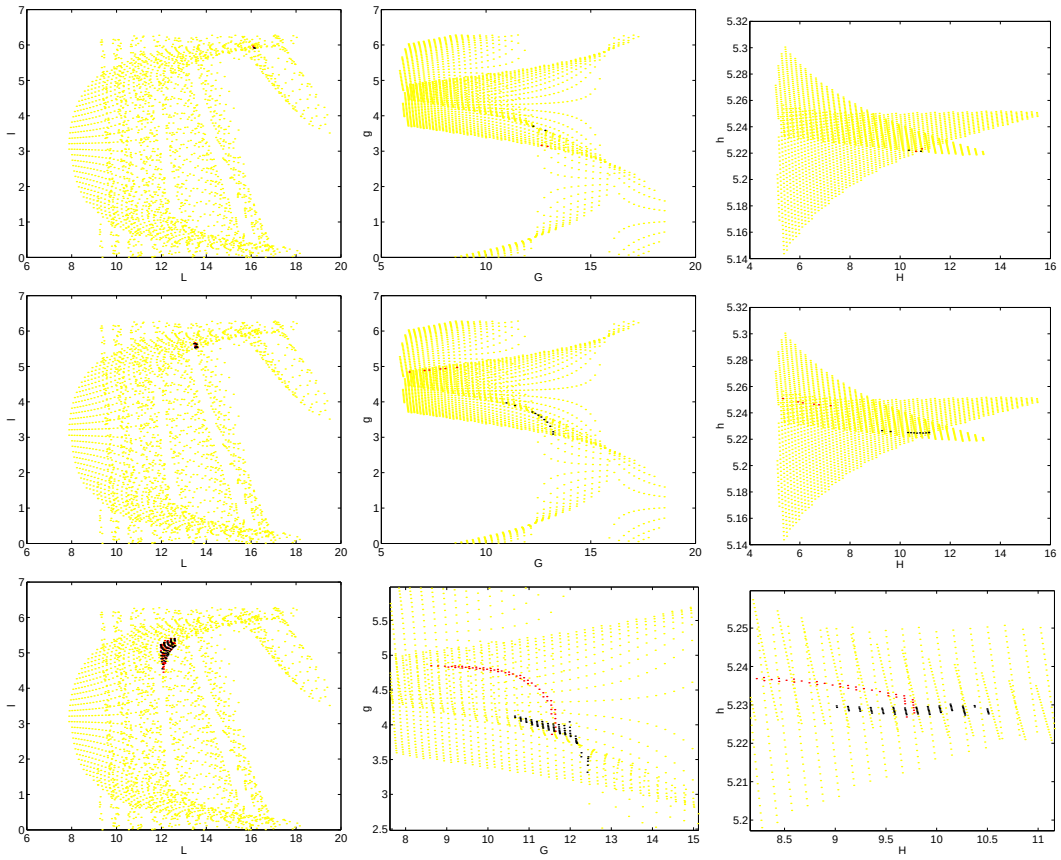

Figure 3.18: $T(70;0)\{\mathcal{A}^{rrrr}\}$ (red) and $T(70;70)\{\mathcal{A}^{rrrr}_{70}\}$ (black) projected onto the Delaunay planes, nonzenith observations

process: cut away the nonoverlap region of the (G, g) plane, then do the same for the new nonoverlap region of the (H, h) plane, and continue to ping-pong back and forth until the intersection is known to within the desired uncertainty. Alternatively, since the actual orbit is now known to within a small neighborhood of a single point, a least squares solution can be carried out.

3.8 Orbit Determination III: J_2 Orbit

The intersection of two admissible regions (eg. Fig. 3.10) can take on a variety of different appearances. The purpose of this section is twofold. The main purpose is to present the reader with a menagerie of qualitatively different examples to give

the reader a broader feel for how these overlap regions can appear. We will do this in the context of the J_2 problem, so that we can also show how the case where one treats perturbations differs from the associated Kepler problem.

3.8.1 Dynamics of the J_2 Orbit

It is known that the gravitational potential of an axisymmetric body can be expanded in a series of the form:

$$V = -\frac{GM}{r}\left(1 - \sum_{n=2}^{\infty} \frac{J_n P_n(\cos(\theta))}{r^n}\right)$$

where θ is measured from the axis of symmetry, and $P_n(x)$ is the n-th Legendre polynomial. Taking into account the first order correction of the gravitational field of the Earth, due to its oblateness, the potential can be approximated by the potential:

$$V \approx -\frac{\mu}{r} + \frac{\mu J_2(3\cos^2\theta - 1)}{r^2}$$

where μ is the gravitational parameter of the Earth and $J_2 \approx 1.08\times10^{-3}$ is the Earth's J_2 term. The disturbing function for the averaged potential for the J_2 problem can be written as (Danby [40]):

$$R = \frac{\mu J_2}{2a^2(1-e^2)^{3/2}}\left(\frac{3}{2}\sin^2 i - 1\right)$$

The full Hamiltonian for the averaged J_2 problem can be written as:

$$\mathcal{H} = -\frac{\mu}{2a} + \frac{\mu J_2}{2a^2(1-e^2)^{3/2}}\left(\frac{3}{2}\sin^2 i - 1\right)$$

Substituting the orbit elements with Delaunay variables (3.16), we can rewrite the potential as follows:

$$\mathcal{F} = -\frac{\mu^2}{2L^2} + \frac{\mu^4 J_2}{2L^3G^3}\left(\frac{1}{2} - \frac{3H^2}{2G^2}\right) \tag{3.24}$$

Applying Hamilton's equations (3.18) to the J_2 Hamiltonian (3.24), we obtain the following dynamic equations of motion for a particle in the Earth's J_2 field:

$$\begin{aligned} \frac{dl}{dt} &= \frac{\mu^2}{L^3} - \frac{3\mu^4 J_2}{2L^4 G^3}\left(\frac{1}{2} - \frac{3H^2}{2G^2}\right) \\ \frac{dg}{dt} &= \frac{15\mu^4 J_2 H^2}{4L^3 G^6} - \frac{3\mu^4 J_2}{4L^3 G^4} \\ \frac{dh}{dt} &= -\frac{3\mu^4 J_2 H}{2L^3 G^5} \end{aligned} \tag{3.25}$$

The conjugate momenta are conserved:

$$\frac{dL}{dt} = 0 \qquad \frac{dG}{dt} = 0 \qquad \frac{dH}{dt} = 0$$

3.8.2 STM of the J_2 Dynamics

The STM corresponding to the transformation $T_4(t; t_0)$ will no longer be given by (3.22). The solution to the dynamic equations of motion (3.25) for the J_2 problem are simply:

$$L(t) = L_0, \qquad G(t) = G_0, \qquad H(t) = H_0$$

$$l(t) = l_0 + \Lambda \cdot (t - t_0), \qquad g(t) = g_0 + \Gamma \cdot (t - t_0), \qquad h(t) = h_0 + \Xi \cdot (t - t_0)$$

where we define

$$\begin{aligned} \Lambda &= \frac{\mu^2}{L^3} - \frac{3\mu^4 J_2}{2L^4 G^3}\left(\frac{1}{2} - \frac{3H^2}{2G^2}\right) \\ \Gamma &= \frac{15\mu^4 J_2 H^2}{4L^3 G^6} - \frac{3\mu^4 J_2}{4L^3 G^4} \\ \Xi &= -\frac{3\mu^4 J_2 H}{2L^3 G^5} \end{aligned}$$

so that the STM is given by:

$$\Phi_4^{J_2}(t;t_0) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \Lambda_L(t-t_0) & 1 & \Lambda_G(t-t_0) & 0 & \Lambda_H(t-t_0) & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \Gamma_L(t-t_0) & 0 & \Gamma_G(t-t_0) & 1 & \Gamma_H(t-t_0) & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \Xi_L(t-t_0) & 0 & \Xi_G(t-t_0) & 0 & \Xi_H(t-t_0) & 1 \end{bmatrix}$$

where Λ_L, Λ_G, and Λ_H are the partial derivatives of $\Lambda(L,G,H)$ with respect to L, G, and H, respectively; and similarly for Γ and Ξ.

3.8.3 Concurrent Plot of Two Zenith Observations on the Delaunay Planes

In this section we will assume the same initial observation of the debris particle as considered in §3.6, i.e. the attributable vector $A = (0, \pi/6, 0.1, 0.03)$ is recorded at $t = 0$ from the point $\Theta = \pi/3$, $\Phi = 0$, on the Earth's surface. The projections of the corresponding admissible region on the Delaunay planes are shown in Fig. 3.8. Taking into account the J_2 perturbation due to the Earth's oblateness on the debris particle's orbit, the time-evolved Delaunay projections are shown in Fig. 3.19. (The time-evolved Delaunay projections of the same admissible region in the Kepler case were shown in Fig. 3.9). One sees that at these time scales, the J_2 effect on the (L,l) and (G,g) planes is fairly insignificant. On the (H,h) plane, the J_2 perturbation causes the projection of the uncertainty surface to widen from a line to a narrow two-dimensional region, thus regenerating the degenerate surface projection.

We consider now the three cases that the actual debris particle is virtual debris particle #400, #600, and #1000. Assuming a second zenith observation of the debris particle is made after 70 hours, the two concurrent admissible region projections will appear as in Fig. 3.20. On the other hand, if the second observation is instead made

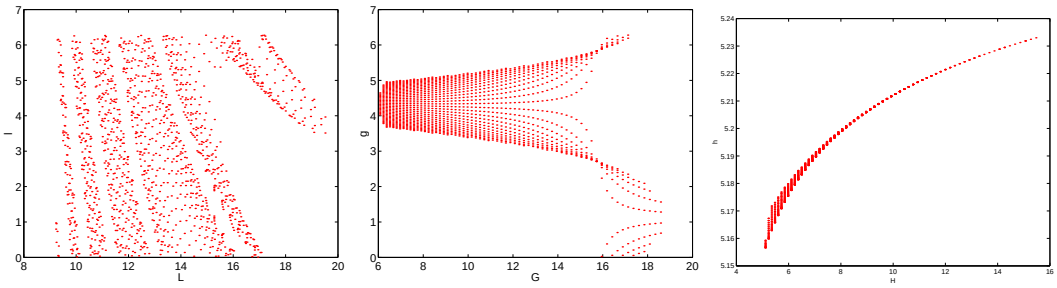

Figure 3.19: $T(70;0)\{\mathcal{A}\}$ projected onto the Delaunay planes, J_2 problem

after 140 hours, the concurrent projections of the two admissible regions will appear as in Fig. 3.21. Each of these 6 cases assume the same initial attributable vector with admissible region Fig. 3.15. Notice that the projection of $T(t;0)\{A\}$ on the (L,l) plane becomes more shredded as it dynamically evolves. In fact, modulo the perturbations, most of the dynamical evolution of the original uncertainty surface is contained within this shredding. We show these cases to give the reader a broader feel of the variety in which these uncertainty intersections can appear. Since we treat the intersection procedure in §3.6 and §3.7, we will not discuss it again here.

3.9 A Conceptual Algorithm

The purpose of this paper has been to introduce and illustrate the viability of this orbit determination technique. As such, all surface intersection reductions were carried out by hand. Future research must be done on the development of technology that efficiently automates this process. As intersections of two-dimensional surfaces must be performed, and not higher-dimensional surfaces, it is feasible to develop computationally efficient approaches for this. In this section we discuss an algorithm and indicate how one might use this technology in the orbit determination process and the subsequent inclusion of these new orbits in the space debris catalog when faced with a large number of observations per night. As was mentioned in §3.3.1,

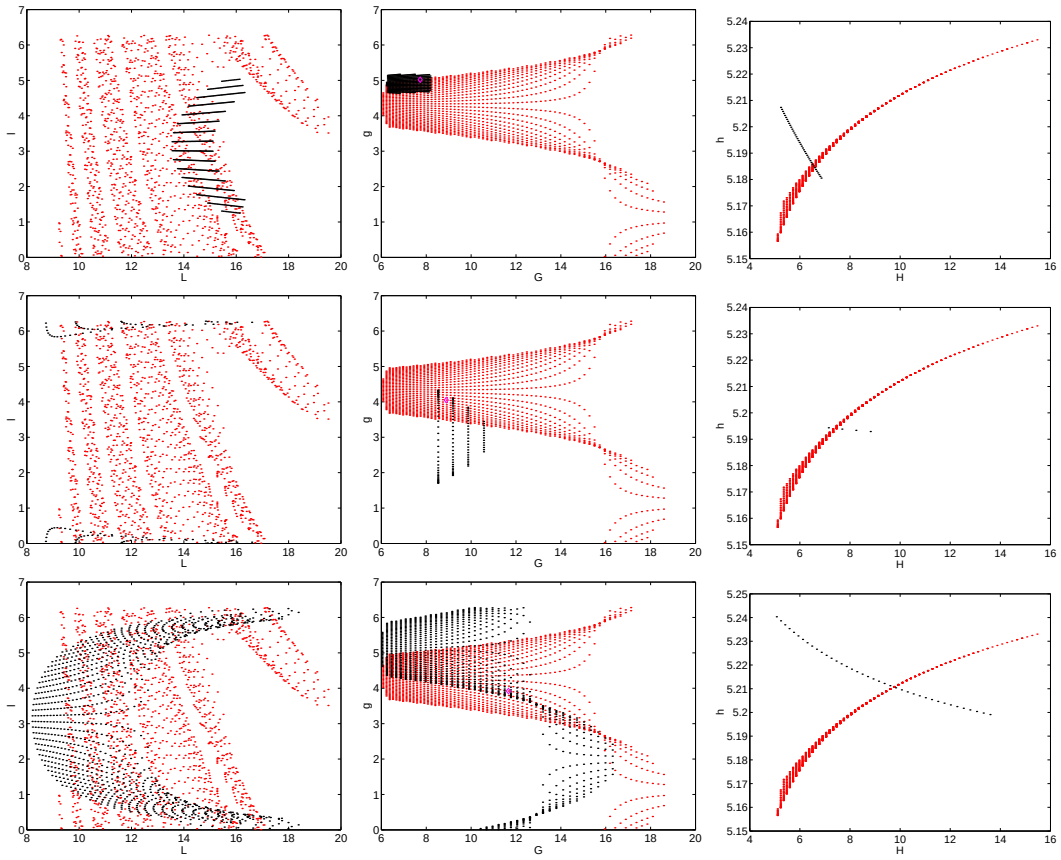

Figure 3.20: $T(70;0)\{\mathcal{A}\}$ (red) and $T(70;70)\{\mathcal{A}_{70}\}$ (black) projected onto the Delaunay planes, J_2 problem. Assumes the true particle is number #400, #600, and #1000, from top down.

each observation should be recorded as an observation vector:

$$x = (A, t, L) \in \mathbb{R}^5 \times \mathbb{N}$$

containing an attributable vector, the observation time, and the observatory's location. Each new observation should be checked against the catalog of known objects. If the observed attributable vector does not match any of the orbital particles in the catalog, it will be saved as an uncorrelated observation. A rolling observation window can be defined (for instance, one week) within which it is compared to all other uncorrelated observations. For these comparisons, a standard epoch time can be defined and all uncorrelated observations made within the observation window can then be mapped into the Delaunay planes and then dynamically evolved or regressed to the

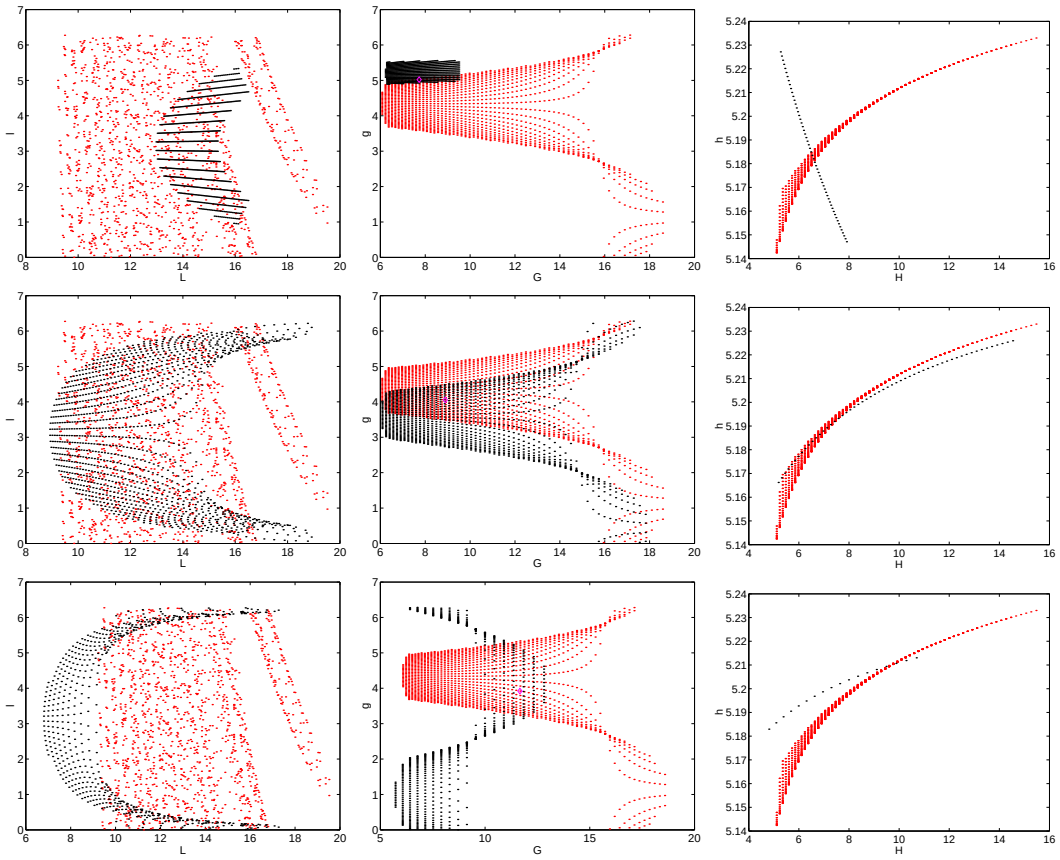

Figure 3.21: $T(140;0)\{\mathcal{A}\}$ (red) and $T(140;140)\{\mathcal{A}_{70}\}$ (black) projected onto the Delaunay planes, J_2 problem. Assumes the true particle is number #400, #600, and #1000, from top down.

epoch time. These uncertainty projections can then be stored and intersected with all other such observations to discover which observations are correlated. For each orbit correlation that is found, the corresponding observations can then be saved in a secondary catalog, which is a temporary holding catalog, until the orbit is confirmed, at which time the data can be promoted to the primary catalog of correlated data.

3.10 Alternative Approaches

As we discussed in §3.5.2, the correlation and orbit determination between two data tracks is tantamount to finding the unique intersection point $\mathbf{\Delta}^*$ of two two-dimensional submanifolds, $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)$, of six-dimensional Delaunay space

$\mathcal{D} \cong \mathbb{R}^6$. It is the scope of this paper to discuss the feasibility of an approach, presented in this current work, known as Intersection Theory Analysis (ITA). The ITA algorithm was introduced in §3.5.3 and further illustrated by means of example in §3.6, 3.7, and 3.8, for the cases of zenith observations, non-zenith observations, and J_2 orbital perturbations, respectively. In this current section we will discuss some viable alternative approaches for determining the intersection point $\mathbf{\Delta}^*$, thus obtaining an orbit determination. A thorough analysis of these alternative approaches will be the study of future research; it is our goal here only to demonstrate the voracity of a subvolume intersection approach to orbit determination problems, in the sense that there are competing methodologies within this approach, each with its own merits, that could lead to fruitful results.

3.10.1 TITA

In this paper, we choose the Delaunay space to carry out the ITA process for several reasons. As opposed to orbital element space, the Delaunay space is a symplectic space. This characteristic is beneficial for two reasons. The symplecticity of this space brings with it added structure and geometric invariants that may prove useful in understanding the resulting dynamics. But moreover, on a more mundane but simultaneously more practical level, there is a natural pairing of coordinates, yielding a natural fracture of the space into three separate two-dimensional projective spaces, in which intersections are more effictively carried out. Since the true space is a splicing of these individual projective spaces, the intersection process can be repeated until a single intersection point emerges as victor. The other benefit of using Delaunay coordinates is that it provides a standard space for comparison of competing uncertainty manifolds. For each observation window (e.g. perhaps a rolling two week time span), a common epoch time is defined and all uncertainty

manifolds are mapped to the Delaunay space and then dynamically evolved or regressed to that epoch time. These uncertainty manifolds can therefore be pairwise compared in this common space.

As an alternative, one can perform TITA - topocentric Intersection Theory Analysis. As the name suggests, TITA involves performing the same Intersection Theory Analysis routine in the topocentric observation geometry as opposed to the Delaunay space. As the topocentric viewing geometry is continuously transforming, there is no common space in which to compare all uncertainty distributions. Instead, for each pair of tracks, one does the following. First compute the admissible region corresponding to each track of data. Pull the second admissible regions back into geocentric cartesian coordinates, dynamically map the resulting distribution forward or backward in time to the time of the first track, then push the resulting manifold forward to the topocentric coordinates in which the first track was viewed. This approach has the possible disadvantage that *for each pair* of tracks, an uncertainty surface must be mapped into a new space, entailing a total of $O(N^2)$ surface mappings for a set of N uncorrelated tracks. Recall that for regular ITA, each uncertainty surface is *mapped once* into a common space for comparison, yielding a total of $O(N)$ such mappings. In the TITA approach, one then carries out the ITA algorithm in the topocentric observation projective spaces $\mathcal{O}_\rho = (\rho, \dot{\rho})$, $\mathcal{O}_\alpha = (\alpha, \dot{\alpha})$, and $\mathcal{O}_\delta = (\delta, \dot{\delta})$. TITA has the advantage that, when performing ITA in the observation projective spaces, that *one* of the uncertainty surfaces degenerates to *a single point* in the projective spaces $\mathcal{O}_\alpha$ and $\mathcal{O}_\delta$. Thus, ITA degenerates to checking to see if the second uncertainty manifold contains in it a four-vector $\vec{\theta} = \langle \alpha, \dot{\alpha}, \delta, \dot{\delta} \rangle$ in an epsilon-neighborhood of the original attributable vector A of the first observation. If so, one then checks if the corresponding $\langle \rho, \dot{\rho} \rangle$ of the second uncertainty surface lie on the admissible region of the

first. Thus the ease with which one performs ITA in the topocentric projective space and added computation associated with the additional surface mappings must be compared with the relative complexity in performing intersections in the Delaunay projective spaces and the associated computational advantage of performing fewer surface mappings.

3.10.2 Metric Approaches

As the ultimate goal is to find the intersection point shared by two two-dimensional surfaces $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)$, each represented by a discretized VD field mapped to Delaunay space, a natural approach might be to find the pair of points $\mathbf{\Delta}^*_1 \in F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $\mathbf{\Delta}^*_2 \in F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)$ which are "closest together" in some sense. The accomplishment of this objective depends on the suitable choice of a metric for the Delaunay space $\mathcal{D}$. Taking the example of two zenith-observations considered in §3.6, there are literally hundreds of point pairs, $\{\mathbf{\Delta}_{1i} \in F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1); \mathbf{\Delta}_{2i} \in F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)\}_{i=1}^{100's}$, whose Euclidean separation distance $||\mathbf{\Delta}_{1i} - \mathbf{\Delta}_{2i}||_{\text{Euclid}}$ is less than the distance between the actual approximate intersection points $\mathbf{\Delta}^*_1$ and $\mathbf{\Delta}^*_2$ yielded by ITA and known to be correct by construction. Many of these false positives don't even hit the obvious intersection point on the (H, h) plane (see Fig. 3.10). Clearly a much more judicious choice of metric is needed. One could imagine that with the correct metric, the pair of points that lies closest together will always be the intersection point. Such a discovery would greatly reduce the computational complexity of repeatedly performing successive intersections between overlapping planar laminas.

Typically when one speaks of a metric one thinks of it as being defined globally, i.e. we would have:

$$\mathbf{g} : T\mathcal{D} \times T\mathcal{D} \to \mathbb{R}$$

We will, however, require a metric, $\mathbf{g}_1$, that is defined only on the restriction of $\mathcal{D}$ to the submanifold $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$, so that:

$$\mathbf{g}_1 : T\mathcal{D}|_{F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)} \times T\mathcal{D}|_{F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)} \to \mathbb{R}$$

Note $\mathbf{g}_1$ takes as inputs full six-dimensional vectors from $T\mathcal{D}$, only it is defined solely at points located on the submanifold $F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$; it is *not* the tangent bundle to this submanifold. For any point $\mathbf{\Delta}_1 \in F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$, the metric $\mathbf{g}_1(\mathbf{\Delta}_1)$ is a good approximation for the metric in a local neighborhood $\mathcal{U} \subset \mathcal{D}$ of $\mathbf{\Delta}_1$. Thus for any $\mathbf{\Delta} \in \mathcal{U}$, the distance between $\mathbf{\Delta}$ and $\mathbf{\Delta}_1$ is approximated by

$$d(\mathbf{\Delta}, \mathbf{\Delta}_1) \approx \sqrt{(\mathbf{\Delta} - \mathbf{\Delta}_1)^T \cdot \mathbf{g}_1(\mathbf{\Delta}_1) \cdot (\mathbf{\Delta} - \mathbf{\Delta}_1)}$$

However if the point $\mathbf{\Delta}$ is far away, we care less about the actual measure of this distance than about the fact that it is far away. Hence one can simply seek the points $\mathbf{\Delta}_1^*$ and $\mathbf{\Delta}_2^*$ that minimize

$$D_{\min} = \inf_{\mathbf{\Delta}_1 \in F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)} \left[\inf_{\mathbf{\Delta}_2 \in F^{\tau}_{\mathfrak{X}_2}(\mathcal{C}_2)} \left(\sqrt{(\mathbf{\Delta}_2 - \mathbf{\Delta}_1)^T \cdot \mathbf{g}_1(\mathbf{\Delta}_1) \cdot (\mathbf{\Delta}_2 - \mathbf{\Delta}_1)} \right) \right]$$

Since both submanifolds are discretized by VD fields, one simply compares this quantity for pairwise sets of points.

As mentioned above, the successful actualization of this method depends on the correct choice of metric $\mathbf{g}_1$. Such a metric might be constructed as to preserve the integrity and structure of the observation geometry in which the attributable vector was first recorded. The Euclidean metric induces the following metric on the spherical

coordinates $\langle \rho, \dot{\rho}, \alpha, \delta, \dot{\alpha}, \dot{\delta} \rangle$ used in the topocentric (TC) frame:

$$\mathbf{g}_{\mathrm{TC}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho\cos\delta & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho\cos\delta & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho \end{pmatrix}$$

One can simply push this metric forward to Delaunay space in the following sense. For any $\mathbf{\Delta} \in F^{\tau}_{\mathfrak{X}_1}(\mathcal{C}_1)$ and $\mathbf{v}_1,\ \mathbf{v}_2 \in T_{\mathbf{\Delta}}\mathcal{D}$, define

$$\mathbf{g}_1(\mathbf{\Delta})(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{g}_{\mathrm{TC}}({F^{\tau}_{\mathfrak{X}_1}}^{-1}(\mathbf{\Delta}))\left((dF^{\tau}_{\mathfrak{X}_1})^{-1} \cdot \mathbf{v}_1, (dF^{\tau}_{\mathfrak{X}_1})^{-1} \cdot \mathbf{v}_2\right)$$

Since the matrix of the linear transformation $dF^{\tau}_{\mathfrak{X}_1}$ is simply the Jacobian matrix Φ of the transformation, computation of which was discussed in §3.4.2, which is easily computed, we have:

$$\mathbf{g}_1(\mathbf{\Delta})(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{g}_{\mathrm{TC}}({F^{\tau}_{\mathfrak{X}_1}}^{-1}(\mathbf{\Delta}))\left(\Phi^{-1} \cdot \mathbf{v}_1, \Phi^{-1} \cdot \mathbf{v}_2\right) = \mathbf{v}_1^T \cdot (\Phi^{-1})^T \cdot \mathbf{g}_{\mathrm{TC}} \cdot \Phi^{-1} \cdot \mathbf{v}_2$$

we have the matrix of the metric $\mathbf{g}_1$ is given by:

$$\mathbf{g}_1 = (\Phi^{-1})^T \cdot \mathbf{g}_{\mathrm{TC}} \cdot \Phi^{-1}$$

This approach is not as of yet entirely robust, as it still returns false positives for the intersection point. However, it seems to preserve the visual sense of closeness one has from visual examination of the Delaunay planes, i.e. points close together actually look close together. This has not always been the case with other metrics we experimented with. Further investigation and development of this metric will be a topic of future research.

3.11 Conclusion

In this paper, we presented a method for the orbit determination of two previously uncorelated observations of space debris particles. We considered the case of a Kepler orbit with two zenith measurements, the case of a Kepler orbit with two non-zenith measurements, and also the case where the J_2 perturbation was included. Each observation is to be treated as a two-dimensional uncertainty surface, and can be mapped into the symplectic Delaunay space. For the unperturbed problem the projections of this surface onto the (G, g) and (H, h) Delaunay planes are static, and the projection on the (L, l) plane shifts at a rate dependent upon only (L). The Delaunay space is actually the action-angle space, and is isomorphic to $T^3 \times \mathbb{R}^3 = S^1 \times S^1 \times S^1 \times \mathbb{R}^3$. Since the angles $l,\ g,\ h$ are modulo 2π, and each $L = const.$ strip of the (L, l) projection of the uncertainty surface progresses along at a constant rate in time, the surface projection on the (L, l) plane becomes more and more "shredded" as time evolves. Since the other two symplectic projections of the surface are static for the Kepler problem, this shredding occurs *without* an increase or decrease to the total projected area on the (L, l) plane. In order to correlate two separate observations, their respective uncertainty surfaces are to be mapped to the same epoch time and then projected concurrently onto the three Delaunay planes. If both observations are zenith observations, the projections of these surfaces onto the (H, h) plane degenerate to a single line and a unique orbit determination can be made as outlined in 3.6. This orbit determination process is robust as a similar process can be used to determine the orbit in the case of two nonzenith observations, as we showed in §3.7. Our goal in the present work has been to present a qualitative overview of this process and indicate the existence and robustness of this process. Future work is needed in the

development of an efficient algorithm that will automate the uncertainty reduction process we illustrated here. Furthermore, uncertainty in the attributable vector must also be taken into account. This too will be the topic of future work.

CHAPTER IV

The Eccentric Frame Decomposition for Central Force Fields

4.1 Introduction

The rosette-shaped motion of a particle in a central force field is known to be classically solvable by quadratures. In this chapter we will present a new approach for describing and characterizing such motion based on the eccentricity vector of the two body problem. In general, this vector is not an integral of motion. However, the orbital motion, when viewed from the nonuniformly rotating frame defined by the orientation of the eccentricity vector, can be solved analytically and will either be a closed periodic circulation or libration. The motion with respect to inertial space is then given by integrating the argument of periapsis with respect to time. Finally we will apply the decomposition to a modern central potential, the spherical Hernquist-Newton potential, which models dark matter halos of galaxies with central black holes.

4.1.1 Central Force Fields

The motion of a particle in a central force field is known to be classically solvable by quadratures. Due to the spherical symmetry of the force field, an angular momentum integral exists and the ensuing motion is confined to a single orbital plane so that, without loss of generality, we can assume the system to have 2 degrees of freedom.

In polar coordinates, the Hamiltonian may be expressed as:

$$H = \frac{1}{2}\left(\dot{r}^2 + \frac{h^2}{r^2}\right) - U(r),$$

where $h = r^2\dot{\theta}$ is the angular momentum and $U(r)$ is the potential energy function. We will use the standard convention that dots refer to time derivatives, whereas primes refer to spatial derivatives. The corresponding Hamiltonian system,

$$\dot{r} = v_r \qquad \dot{v}_r = \frac{h^2}{r^3} + U'(r) \qquad \dot{\theta} = \frac{h}{r^2} \qquad \dot{h} = 0,$$

is integrable by quadratures:

$$\int_{r(0)}^{r(t)} \frac{\pm dr}{\sqrt{2H + 2U(r) - h^2/r^2}} = t - t_0 \tag{4.1}$$

$$\theta(t) = \theta_0 + \int_{t_0}^{t} \frac{h\ dt}{r(t)^2}. \tag{4.2}$$

The ensuing motion follows rosette-shaped paths (Arnold [6, 4], Whittaker [133], etc.). For some current research related to central force fields, also see Brun & Pacheco [26], Celletti & Chierchia [31], and Lei & Santorprete [76].

4.1.2 Osculating Orbital Elements

One could, alternatively, proceed using Variation of Parameters and Lagrange's Planetary Equations (Brouwer & Clemence [25], Roy [118]). In this case, one can write down differential equations of motion for the six osculating classical orbital elements and then solve them by quadrature for all time. These equations are nonlinear and furthermore depend upon a choice of the "planetary" gravitational parameter μ. One assumes the potential is a perturbation of a Newtonian potential:

$$U(r) = \frac{\mu}{r} + R(r),$$

where $R(r)$ is known as the *disturbing function*. For a general central force field where there is no nominal attracting body, such a choice is somewhat arbitrary.

For instance, there is no primary "planet" when considering motion in a galactic halo. A gravitational parameter for the unperturbed motion can nonetheless be artificially contrived, perhaps based on the total halo mass (if the motion evolves in the outskirts of the galaxy) or based on the mass of a central galactic bulge or black hole (if the motion evolves near the galactic core). Whatever the choice of gravitational parameter, a complete set of osculating orbital elements arises and the ensuing motion can be determined.

4.1.3 The Eccentric Frame

We will define a gravitational parameter based on the central force field's potential, with no reference to a main attracting body and perturbation theory. Following the classical analogy, we define $\mu(r)$ such that:

$$U(r) = \frac{\mu(r)}{r}.$$

We will show that this gives rise to a nonstatic eccentricity vector that rotates at a nonuniform rate. The eccentricity vector (Runge-Lenz vector) associated with this spatially variable gravitational parameter function defines a preferred coordinate system which we call the eccentric frame. With respect to this frame, we will show that the motion follows a closed orbit. Depending on the value of energy, the particle will make closed circulations or librations in the eccentric frame. The eccentric frame decomposition gives rise to a set of orbital elements. We will discuss their physical implications and the key features of how they arise. In particular, one can have circular orbits in inertial space with nonzero osculating eccentricity. This feature is not unique to our method, it can arise from any choice of osculating orbital elements. The eccentric frame decomposition, however, illuminates the behavior and gives rise to a new standard description that better fits orbits of central force field potentials.

4.2 The Eccentric Frame Decomposition

We first define the eccentric frame by means of specifying the nonstatic eccentricity vector associated with the gravitational parameter function $\mu(r) = rU(r)$. We then show that the particle traces a closed orbit as viewed from this noninertial frame. Finally we compute the set of osculating orbital elements that belong to this system.

4.2.1 Motion with respect to the Eccentric Frame

Given a spherically symmetric potential energy field, we can recast the Hamiltonian into the following form, reminiscent of its classical analogy:

$$E = \frac{1}{2}\left(v^2 + \frac{h^2}{r^2}\right) - \frac{\mu(r)}{r}, \tag{4.3}$$

where $h = r^2\dot{\theta}$ is the magnitude of the angular momentum vector,

$$\mathbf{H} = \mathbf{r} \times \dot{\mathbf{r}}, \tag{4.4}$$

and (r, v, θ, h) are the symplectic coordinates, with $v = \dot{r}$. This gives rise to the following Hamiltonian equations of motion:

$$\begin{aligned} \dot{r} &= v & \dot{v} &= \frac{h^2}{r^3} + \frac{\mu'(r)}{r} - \frac{\mu(r)}{r^2} \\ \dot{\theta} &= \frac{h}{r^2} & \dot{h} &= 0 \end{aligned}$$

which can be recast in the following form

$$\ddot{\mathbf{r}} = \left(\ddot{r} - \frac{h^2}{r^3}\right)\mathbf{e}_r = \left(\frac{\mu'(r)}{r} - \frac{\mu(r)}{r^2}\right)\mathbf{e}_r, \tag{4.5}$$

where $\mathbf{r} = r\mathbf{e}_r$. Consider now the eccentricity vector

$$\mathbf{B} = \dot{\mathbf{r}} \times \mathbf{H} - \mu(r)\mathbf{e}_r. \tag{4.6}$$

Its evolution is governed by the following equations of motion:

$$\begin{aligned}
\dot{\mathbf{B}} &= \ddot{\mathbf{r}} \times \mathbf{H} - \mu'(r)\dot{r}\mathbf{e}_r - \mu(r)\dot{\theta}\mathbf{e}_\theta \\
&= \left(\frac{\mu'(r)}{r} - \frac{\mu(r)}{r^2}\right) r^2\dot{\theta}\mathbf{e}_r \times \hat{\mathbf{H}} - \mu'(r)\dot{r}\mathbf{e}_r - \mu(r)\dot{\theta}\mathbf{e}_\theta \\
&= -\left(\frac{\mu'(r)}{r} - \frac{\mu(r)}{r^2}\right) r^2\dot{\theta}\mathbf{e}_\theta - \mu'(r)\dot{r}\mathbf{e}_r - \mu(r)\dot{\theta}\mathbf{e}_\theta \\
&= -\mu'(r)\dot{\mathbf{r}}.
\end{aligned}$$

The vector $\mathbf{B}$ itself works out to be

$$\begin{aligned}
\mathbf{B} &= r^2\dot{\theta}(\dot{r}\mathbf{e}_r + r\dot{\theta}\mathbf{e}_\theta) \times \hat{\mathbf{H}} - \mu(r)\mathbf{e}_r \\
&= -r^2\dot{r}\dot{\theta}\mathbf{e}_\theta + r^3\dot{\theta}^2\mathbf{e}_r - \mu(r)\mathbf{e}_r \\
&= -h\dot{r}\mathbf{e}_\theta + \left(\frac{h^2}{r} - \mu(r)\right)\mathbf{e}_r.
\end{aligned}$$

We thus find the magnitude of $\mathbf{B}$ is:

$$B = \sqrt{2h^2E + \mu(r)^2}. \tag{4.7}$$

We define the argument of periapsis, ω, to be the angle made between the inertial x-axis and the $\mathbf{B}$-vector. The $\mathbf{B}$-vector defines a rotating reference frame, which we call the eccentric frame. We define $\hat{\mathbf{B}}$ to be a unit vector in the $\mathbf{B}$ direction. Hats will denote unit vectors. Let X and Y be the cartesian coordinates of the particle with respect to the eccentric frame and let x and y be the cartesian coordinates of the particle with respect to the inertial frame. The axes of the inertial frame are determined by the stationary unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. The polar angle of the particle measured with respect to the $\hat{\mathbf{B}}$ direction is known as the true anomaly f. This notation is also used in Roy [118]. The polar angle of the particle in the inertial frame is related to the true anomaly by the following relation:

$$\theta = f + \omega. \tag{4.8}$$

In some of the literature, the true anomaly f is denoted by ν; and the inertial polar angle (argument of latitude) θ is denoted by u. Decomposing the eccentricity vector $\mathbf{B}$ in the inertial frame, we see that

$$\begin{aligned}\mathbf{B} &= \left(\dot{r}h\sin\theta + \left(\frac{h^2}{r} - \mu(r)\right)\cos\theta\right)\hat{\mathbf{i}} \\ &\quad + \left(-\dot{r}h\cos\theta + \left(\frac{h^2}{r} - \mu(r)\right)\sin\theta\right)\hat{\mathbf{j}}.\end{aligned}$$

However, by definition, $\mathbf{B} = B(r)(\cos\omega\hat{\mathbf{i}} + \sin\omega\hat{\mathbf{j}})$. Hence:

$$\begin{aligned}\begin{bmatrix}\cos\omega \\ \sin\omega\end{bmatrix} &= \frac{1}{B(r)}\mathbf{A}\cdot\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} \\ &= \frac{1}{B(r)}\mathbf{A}\cdot\mathbf{F}\cdot\begin{bmatrix}\cos\omega \\ \sin\omega\end{bmatrix} \qquad (4.9)\end{aligned}$$

where we have defined

$$\mathbf{A} = \begin{bmatrix}(h^2/r - \mu(r)) & \dot{r}h \\ -\dot{r}h & (h^2/r - \mu(r))\end{bmatrix}$$

and

$$\mathbf{F} = \begin{bmatrix}\cos f & -\sin f \\ \sin f & \cos f\end{bmatrix}$$

and have further made use of the trigonometric identities

$$\begin{aligned}\cos\theta &= \cos(f+\omega) = \cos f\cos\omega - \sin f\sin\omega \\ \sin\theta &= \sin(f+\omega) = \sin f\cos\omega + \cos f\sin\omega.\end{aligned}$$

We recognize that the matrix premultiplying the vector $\langle\cos\omega, \sin\omega\rangle$ on the right hand side of (4.9) must be the identity matrix. Hence we have found an explicit expression relating the true anomaly and the radius:

$$\cos f = \frac{1}{B(r)}\left(\frac{h^2}{r} - \mu(r)\right) \qquad (4.10)$$

$$\sin f = \frac{1}{B(r)}\dot{r}h. \qquad (4.11)$$

Thus we see that the particle traces out a closed path in the eccentric frame. By carefully considering (4.11), we see that periapsis is always achieved at $f = 0$, i.e. when $\mathbf{r}$ and $\mathbf{B}$ are parallel; and that apoapsis is achieved at $f = \pi$, i.e. when $\mathbf{r}$ and $\mathbf{B}$ are anti-parallel.

If the angular momentum is positive, (4.11) tells us that r is increasing when the particle is in the upper half plane and is decreasing when the particle is in the lower half plane. The opposite is true for the case of a negative angular momentum.

4.2.2 The Osculating Eccentricity and Semi-Major Axis

We would also like to point out that one can rearrange (4.10) into the following form:

$$r = \frac{h^2/\mu(r)}{1 + B(r)\cos f/\mu(r)} = \frac{p(r)}{1 + e(r)\cos f}, \tag{4.12}$$

completely analogous to its classical ($\mu(r) = \text{const.}$) form.

Utilizing the relation $p = a(1 - e^2)$, we can define the osculating eccentricity and semi-major axis of the system in closed form as follows:

$$e(r) = \frac{B(r)}{\mu(r)} \tag{4.13}$$

$$a(r) = \frac{h^2\mu(r)}{\mu(r)^2 - B(r)^2}. \tag{4.14}$$

These are given unambiguously as a function of r, without integrating. They represent a standard decomposition of the motion. Using a standard choice of osculating orbital elements, one would first define a semi-arbitrary choice for a fixed μ. Thus, there is no unique standard set of osculating orbital elements for a general system, rather a one parameter family of orbital elements that describe the motion. By using the radially varying $\mu(r)$, we seek to better normalize the description of motion in such systems.

Since the true anomaly is given by the relations (4.10) and (4.11), one now only need solve for the osculating argument of periapsis to obtain the complete motion as a function of time.

4.2.3 The Osculating Argument of Periapsis

Solving for the osculating argument of periapsis can be done in one of two ways. First, one may integrate (4.1)-(4.2) by quadratures. Once r and θ are known, f, a, and e can be extracted by the above relations (4.10), (4.11), (4.13), (4.14); then the osculating argument of periapsis can be solved by means of the relations $\theta = f + \omega$. On the other hand one can solve the quadrature we derive below.

To determine the rotation of the eccentric frame, consider the angular momentum integral:

$$h = r^2\dot{\theta} = r^2(\dot{f} + \dot{\omega}) = r^2 f'(r)\dot{r}\left(1 + \frac{d\omega}{df}\right). \tag{4.15}$$

Differentiating (4.10) and utilizing (4.11), we have that

$$\begin{aligned} -\sin(f(r))f'(r) &= \frac{-\dot{r}f'(r)h}{B(r)} \\ &= \frac{1}{B(r)}\left(\frac{-h^2}{r^2} - \mu'(r)\right) \\ &\quad -\frac{B'(r)}{B(r)^2}\left(\frac{h^2}{r} - \mu(r)\right). \end{aligned}$$

We can now solve (4.15) for $\omega'(f)$:

$$\frac{d\omega}{df} = \frac{h}{f'(r)\dot{r}r^2} - 1 = \frac{\Phi(r)}{B(r)h^2 - \Phi(r)} \tag{4.16}$$

where we define $\Phi(r)$ as:

$$\Phi(r) = \mu(r)r^2B'(r) - B'(r)rh^2 - \mu'(r)r^2B(r). \tag{4.17}$$

We thus have

$$\omega(f) = \omega(0) + \int_0^f \frac{\Phi(r)}{B(r)h^2 - \Phi(r)}\, d\tilde{f}. \tag{4.18}$$

where we recognize $r = r(\tilde{f})$ in the integrand, by the relations (4.10) and (4.11). Together with (4.13) and (4.14), this constitutes a full set of osculating orbital elements that are well-defined for the orbit for all time.

The full motion is then completely specified in terms of the parameter f by the relation:

$$\theta(f) = f + \omega(f).$$

4.3 The Zero Velocity Curve

The central force problem is a 2 degree of freedom problem with 2 integrals of motion, E and h. It is therefore integrable and, in fact, reduces to motion on a Liouville torus. The symplectic coordinates of the system are (r, v, θ, h). The coordinate h is conserved, and the motion therefore takes place on the $h =$ const. hyper-plane. Motion in the (r, v) plane is constrained to the curve $\Gamma_{h,E}$ defined by (4.3), with fixed E and h. Meanwhile, θ cycles along S^1 according to $h = r^2\dot{\theta}$. Motion in the reduced (r, v, θ) space can therefore be visualized as follows: it is constrained to the surface obtained by revolving the curve $\Gamma_{h,E}$ around the v axis. This resulting surface is (obviously) topologically equivalent to the Liouville torus, but obtained directly without the arduous task of computing action-angle variables.

For a fixed h, as one varies the energy, one encounters various bifurcation points where the system undergoes changes.

4.3.1 Periapsis and Apoapsis

Computation of the periapsis and apoapsis radii is accomplished by the standard technique of plotting the zero-velocity curve on the $E - r$ plane. The plot is obtained by setting $v = \dot{r} = 0$ in (4.3). The resulting equation is:

$$E_{\text{zv}}(r) = \frac{1}{2}\frac{h^2}{r^2} - \frac{\mu(r)}{r}. \tag{4.19}$$

For a fixed energy E, the solutions to this equation represent the periapsis r_p and apoapsis r_a radii. Maximum and minimum values of of $E_{\text{zv}}(r)$ correspond to unstable and stable circular orbits, respectively. If there are multiple "wells," the corresponding roots of this equation alternate $r_{p1}, r_{a1}, r_{p2}, r_{a2}, \ldots$, and the forbidden regions of the inertial x-y plane are concentric, circular annuli.

4.3.2 Circular Orbits

As one decreases the energy for a fixed angular momentum, the curves $\Gamma_{h,E}$ on the (r, v)-plane shrink until they degenerate to a single point on the r-axis which corresponds to a circular orbit in the (r, θ) polar plane. This occurs at the local minima of E on the (r, v) plane, and hence is given by $\nabla E = 0$, where E is given by (4.3) and h is held fixed. This condition amounts to

$$v = 0 \tag{4.20}$$

$$\frac{h^2}{r^3} + \frac{\mu'(r)}{r} - \frac{\mu(r)}{r^2} = 0 = \ddot{r}. \tag{4.21}$$

The root of (4.21), r_{circ}, corresponds to the radius of the circular orbit which occurs at the minimum energy $E_{\text{circ}} := E_{\text{zv}}(r_{\text{circ}})$.

4.3.3 Escape Orbits

If $U(r) \to$ const. as $r \to \infty$, a series of unbounded orbits are present in the solution space. Such orbits are classified as escape orbits. Typically one takes the potential at infinity to be zero, so that $U(r) \to 0$ as $r \to \infty$, so that orbits with negative energies are gravitationally bounded to the center of the potential, whereas orbits with positive energies have enough energy to escape to infinity.

4.4 Circulations vs. Librations in the Eccentric Frame

As one decreases the energy from E_{esc} to the minimum energy E_{circ}, one encounters a bifurcation in the eccentric frame at E_{crit}, as the orbits (as seen from the eccentric frame) change from circulations to librations. This is a necessary transition that must occur, as one lowers the energy, before one can reach a circular orbit. It will be our goal in this section to understand the how this bifurcation comes about and to give a qualitative description of motion in the eccentric frame for fixed h as one varies E.

4.4.1 The Critical Energy

We now define a critical radius and critical energy. The critical radius is defined as the root to the right hand side of (4.10), which occurs when:

$$h^2 = r\mu(r). \tag{4.22}$$

For a fixed h, let the solution to (4.22) be r_{crit}. Further, let us define the critical energy as follows:

$$E_{\text{crit}} = E_{\text{zv}}(r_{\text{crit}}) = -\frac{1}{2}\frac{\mu(r_{\text{crit}})}{r_{\text{crit}}}, \tag{4.23}$$

where $E_{\text{zv}}(r)$ is given by (4.19). In the following subsections, we will see how passing through this value of energy brings about a bifurcation in our system.

The critical radius r_{crit} has an important physical signifigance in terms of the eccentric frame. From (4.10), we see that $\cos f = 0$, i.e. the particle is crossing the Y-axis in the eccentric frame, exactly when $r = r_{\text{crit}}$. It is interesting to note that r_{crit} is *independent* of the energy of the system. Thus, as one changes the energy, the particle passes between the left and right hand planes through the same two portals ($Y = \pm r_{\text{crit}}$).

As one decreases the energy, the zero velocity curves $r = r_p$ and $r = r_a$ come closer together. Eventually, one will coincide with r_{crit}. This occurs at the critical

energy E_{crit} and brings about the bifurcation in the system. For $E < E_{\text{crit}}$, the points $Y = \pm r_{\text{crit}}$ both lie in the forbidden region, thus a transition from the left half to right half plane is no longer possible. If the apoapsis zero-velocity curve $r = r_a$ reaches r_{crit} before the periapsis zero-velocity curve $r = r_p$ does, the particle follows periapsis librations (i.e. librations around periapsis on the right half plane) in the eccentric frame. Alternatively, if the periapsis zero-velocity curve r_p reaches r_{crit} first, the particle follows apoapsis librations.

4.4.2 The Route to Periapsis Librations

We will first consider the case where it is the apoapsis radius that coincides with r_{crit} at the bifurcation energy E_{crit}. This event brings about periapsis librations for all energies $E < E_{\text{crit}}$. In Fig. 4.1 and Fig. 4.2, the path of the particle, for various values of energy, is plotted *with respect to the eccentric frame*, i.e. the X-axis is coincident with the eccentricity vector $\hat{\mathbf{B}}$. As viewed from this nonuniformly rotating frame, the trajectory of the particle makes closed orbits.

If $E >> E_{\text{crit}}$, (4.10) and (4.11) produce a well-defined closed orbit in the eccentric frame, as seen in Figure 4.1a.

As E approaches E_{crit} from above, the apoapsis radius slowly approaches the critical radius, and an orbit such as the one seen in Figure 4.1b is present. Notice the left half of this orbit is nearly circular. This presents some numerical difficulty if one discretizes the radius r and not the true anomaly f. However, this difficulty can be overcome by analytically approximating the left half of the orbit with an ellipse $r_{\text{approx}}(f),\ f \in [-\pi/2, -\pi] \cup [\pi/2, \pi]$ fitted to the data points $r_{\text{approx}}(\pm\pi/2) = r_{\text{crit}}$ and $r_{\text{approx}}(\pi) = r_a$.

At $E = E_{\text{crit}}$, the apoapsis radius and the critical radius coincide, as shown in Figure 4.1c. The particle thus reaches the Y-axis of the eccentric frame at the pre-

cise moment it reaches the zero velocity curve. Recall that (4.11) implies that r is increasing in the upper half plane and decreasing in the lower half plane for the case $h > 0$. This bifurcation point is rather interesting, as one only has a half orbit in the eccentric frame. The motion begins at periapsis, but when it reaches the Y-axis, i.e. apoapsis, it "hops" π-radians to the corresponding point on the lower half plane and then returns to periapsis. To compensate there is a corresponding π-radian hop in the argument of periapsis, so that the true polar angle θ is a continuous function of time. This is allowed as the $\frac{d\omega}{df}$ equation, (4.16), is actually undefined for $\dot{r} = 0$. This is permissible because $B(r_{\text{crit}}) = 0$ exactly if $E = E_{\text{crit}}$, i.e. the eccentricity vector actually vanishes at these endpoints, and then reappears pointing in the opposite direction.

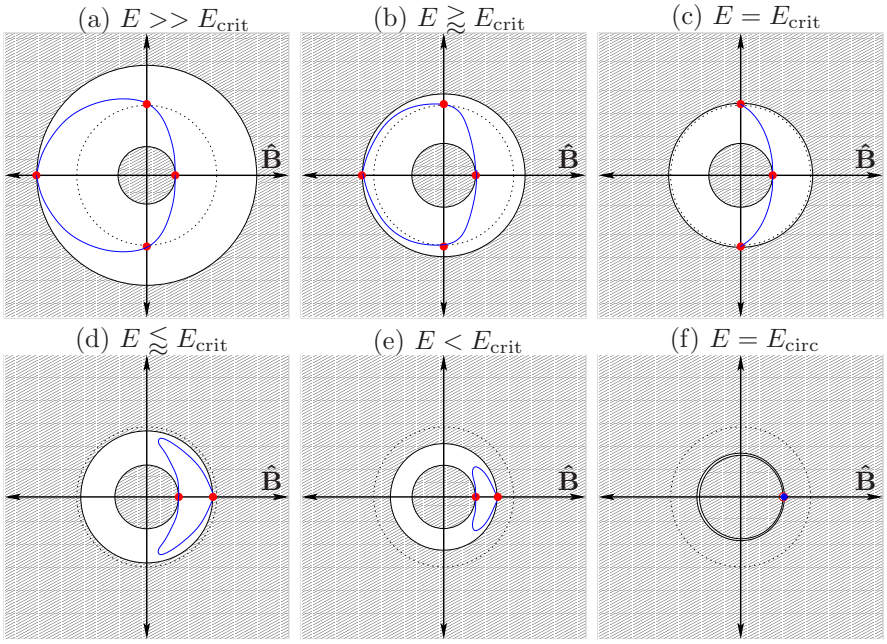

Figure 4.1: The Route to Periapsis Librations

For $E < E_{\text{crit}}$, one sees that $r_a < r_{\text{crit}}$. Thus the portal $X = 0,\ Y = \pm r_{\text{crit}}$ from the right- to the left- half plane lies in the forbidden region. Motion is therefore constrained to the right-half, where periapsis librations arise in the eccentric frame,

Figure 4.1d-4.1e.

Finally, at $E = E_{\text{circ}}$, the periapsis and apoapsis radius coincide and the trajectory in the eccentric frame degenerates to a single point $X = r_{\text{crit}},\ Y = 0$, as seen in Figure 4.1f. The eccentric frame now rotates at a uniform rate and a circular orbit is present in the actual inertial space.

4.4.3 The Route to Apoapsis Librations

As one sees from Figure 4.2, the case where the periapsis radius and the critical radius coincide at the bifurcation energy $E = E_{\text{crit}}$ leads to apoapsis librations in the left half plane.

An easy test to determine whether the librations will be periapsis or apoapsis librations is as follows:

$$r_{\text{circ}} < r_{\text{crit}} \implies \text{periapsis librations}$$

$$r_{\text{circ}} > r_{\text{crit}} \implies \text{apoapsis librations.}$$

Again one sees that there is a π-radian hop in both true anomaly f and argument of periapsis ω at the bifurcation energy $E = E_{\text{crit}}$. Just before the bifurcation, the right half of the orbit (the half closest to periapsis) is nearly circular.

4.5 Symmetry of the Rotation

One can exploit the form of the dynamical equation for ω (4.16) to reduce the numerical integration to one over only one half of an orbit. By examining the differential equation (4.16), one sees that $\omega'(f)$ depends only upon the radial coordinate r. Due to the periodicity of the orbit, we have that $\omega'(f)$ is 2π periodic. Moreover, for $f \in [\pi, 2\pi]$, we have that $\omega'(f) = \omega(2\pi - f)$, since the orbits in the eccentric frame are symmetric with respect to the x-axis.

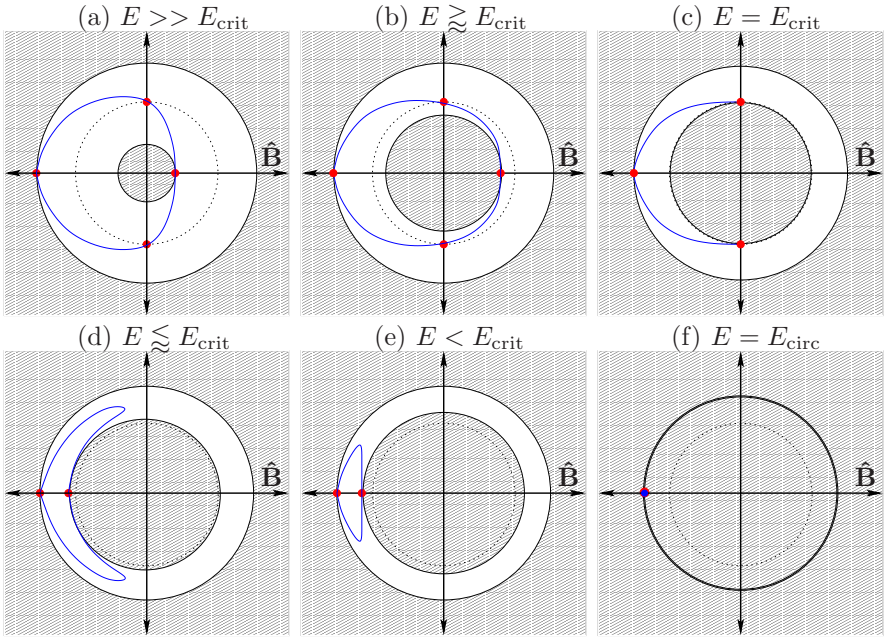

Figure 4.2: The Route to Apoapsis Librations

4.5.1 Circulations

Prior to the bifurcation ($E > E_{\text{crit}}$) the trajectory makes closed circulations in the eccentric frame. During the circulations, there is a secular growth in the argument of periapsis ω. Define the following:

$$\tilde{\Omega} = \int_0^f \frac{d\omega}{df}(r(\tilde{f}))\, d\tilde{f} \qquad f \in [0, \pi],$$

such that $\tilde{\Omega}$ is the argument of periapsis restricted to the domain $f \in [0, \pi]$. We will show that once one has $\tilde{\Omega}$, one can systematically find $\omega(f)$ for all future f, without integration.

The condition that $\omega'(f) = \omega'(2\pi - f)$ for $f \in [\pi, 2\pi]$ suggests that the function $\omega(f)$ is odd with respect to the axes $f = \pi$ and $\omega = \omega(\pi)$ on the interval $[0, 2\pi]$. Thus, given $\tilde{\Omega}$ (which we presume has been found by a numerical algorithm), one

defines

$$\Omega(f) = \begin{cases} \tilde{\Omega}(f) & f \in [0, \pi] \\ 2\tilde{\Omega}(\pi) - \tilde{\Omega}(2\pi - f) & f \in (\pi, 2\pi] \end{cases}.$$

The net secular growth in $\omega(f)$ over one nominal orbit $0 \leq f \leq 2\pi$ is given by

$$\Delta\Omega = \Omega(2\pi).$$

$\omega(f)$ can subsequently be found by applying the following:

$$\omega(f) = n\Delta\Omega + \Omega(f \bmod 2\pi),$$

where n is the orbit number, i.e. $n = 0$ if $f \in [0, 2\pi]$, $n = 1$ if $f \in [2\pi, 4\pi]$, etc.

4.6 The Hernquist-Newton Potential

To illustrate the theory in the context of a modern problem, we will consider motion of a particle (star) in a spherical galaxy, modelled with the Hernquist potential, with a central black hole. These results could be similarly applied to a black hole at the center of a globular cluster, or various other astrophysical configurations that yield spherical or azimuthal symmetry. In this context, the central black hole provides a classical point potential, but no general relativistic effects are included.

4.6.1 Galactic Halos with Central Black Holes

The Hernquist potential has achieved some acclaim in recent years for its ability to analytically model galactic dark matter halos, see Hernquist [64]. A spirographic approximation was used to describe particle motion in the Hernquist potential in Adams and Bloch [2]. For more background on modeling galaxy matter distributions see Binney and Tremaine [15]. We will consider here a coupling between the spherical Hernquist profile and a Newtonian point mass, assumed to model a black hole at the

center of the galaxy. Some numerical modelling of triaxial galaxies with central black holes has already been carried out, as in Poon & Merrit [115].

Let $\mu_{\rm BH}$ and $\mu_{\rm halo}$ be the gravitational parameters of the central black hole and the galactic dark matter halo, respectively; and let b be a length scale of the galaxy (so that $M(b) = M_{\rm tot}/4$, see Hernquist [64]). Then the Hernquist-Newton potential can be written:

$$U(r) = \frac{\mu_{\rm halo}}{R+b} + \frac{\mu_{\rm BH}}{R}.$$

By defining:

$$\mu_0 = \mu_{\rm halo} + \mu_{\rm BH} \qquad \tilde{\mu} = \frac{\mu_{\rm halo}}{\mu_{\rm halo} + \mu_{\rm BH}},$$

the Hernquist-Newton potential can be recast into the following equivalent form:

$$U(r) = \frac{\mu_0}{R}\left(1 - \frac{\tilde{\mu}}{1+R/b}\right)$$

with associated Hamiltonian:

$$\mathscr{E} = \frac{1}{2}\left[\left(\frac{dR}{dT}\right)^2 + \frac{H^2}{R^2}\right] - \frac{\mu_0}{R}\left(1 - \frac{\tilde{\mu}}{1+R/b}\right). \tag{4.24}$$

where $H = R^2\frac{d\theta}{dT}$ is the angular momentum. As this is a central force field, the angular momentum and energy will be conserved quantities. Observe that when $\tilde{\mu} = 0$, the potential energy reduces to that of a Newtonian point mass. When $\tilde{\mu} = 1$, the potential energy is equivalent to the Hernquist potential. For $0 < \tilde{\mu} << 1$, the model represents a Newtonian point mass with a surrounding "Hernquist cloud" and for $0 << \tilde{\mu} < 1$, we have the Hernquist potential with a relatively weak point mass at the origin, which could be used to model a spherical Hernquist galaxy with a central black hole.

4.6.2 Nondimensionalization

Carrying out the following change of variables:

$$R = rb \qquad T = \sqrt{\frac{b^3}{\mu_0}}t,$$

and thus, consequently,

$$H = \sqrt{b\mu_0}h \qquad \mathscr{E} = \frac{\mu_0}{b}E,$$

we can recast the Hamiltonian (4.24) into the following form:

$$E = \frac{1}{2}\left(\dot{r}^2 + \frac{h^2}{r^2}\right) - \frac{\mu(r)}{r}, \tag{4.25}$$

where we define

$$\mu(r) = \left(1 - \frac{\tilde{\mu}}{1+r}\right). \tag{4.26}$$

We have thus recast the Hernquist-Newton potential to a one-parameter family of potentials, with $\tilde{\mu} = 1$ corresponding to a the Hernquist potential and $\tilde{\mu} = 0$ corresponding to a pure Newtonian point mass.

We note that the Hernquist-Newton potential is similar to analogous work on the Manev problem, which considers a potential of the form $U(r) = A/r + B/r^2$. In fact, work has been carried out for the anisotropic Manev problem, which replaces the radial coordinate r with an "elliptic radius" $m = \sqrt{\mu x^2 + y^2}$ (e.g., Craig et al. [36], Diacu & Santoprete [44]). In this type of potential, one obtains a large class of chaotic orbits as well as nonchaotic orbits.

4.6.3 Zero Velocity Curves

Using the relationship for circular orbits (4.21) and the critical energy condition (4.22), sample values of circular radius and energy, critical energy, and the critical periapsis and apoapsis are shown in Table 4.1, where we have taken $h = 0.1$. The

critical radius r_{crit} coincides with the critical energy periapsis radius $r_{\text{crit}}^{\text{peri}}$ in each case, so that the bifurcation always leads to apoapsis librations.

$\tilde{\mu}$	r_{circ}	E_{circ}	E_{crit}	$r_{\text{crit}}^{\text{peri}}$	$r_{\text{crit}}^{\text{apo}}$
1	0.2500	-0.7200	-0.4524	0.1051	1.1932
0.99	0.2274	-0.7539	-0.5000	0.1000	1.000
0.95	0.1508	-0.9372	-0.7440	0.0820	0.4460
0.90	0.0938	-1.3206	-1.1960	0.0647	0.1645

Table 4.1: Various physical quantities for $h = 1$

The zero-velocity curves are plotted in Fig. 4.3 below for $h = 0.1$ and for the same values of $\tilde{\mu}$ as in Table 4.1.

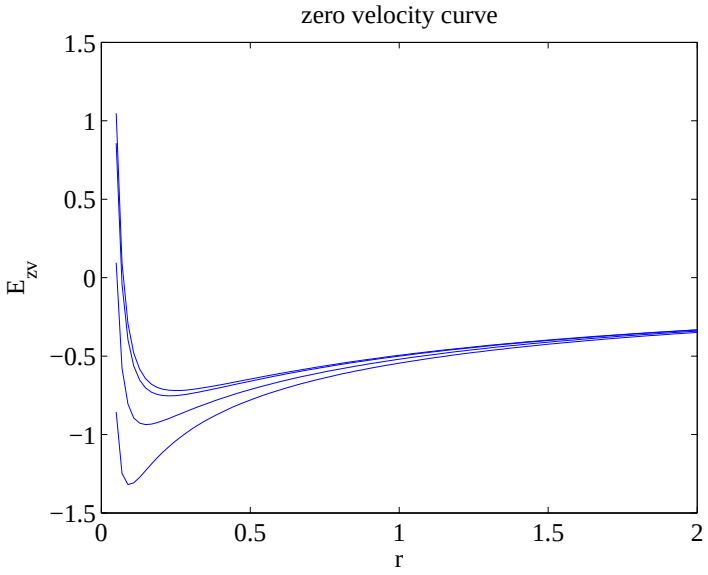

Figure 4.3: Zero Velocity Curves with $h = 1$ and, from top down, $\tilde{\mu} = 1, 0.99, 0.95, 0.9$

4.6.4 Orbits for $h = 0.1$

We examine a sample of orbits in the eccentric and inertial frames for various energies at $h = 0.1$. The zero velocity curve for this angular momentum is plotted in Fig. 4.4 below.

For a sample orbit with $E > E_{\text{crit}}$, we take $E = -0.6$. The orbit as seen from the eccentric and inertial frames is shown in Fig. 4.5.

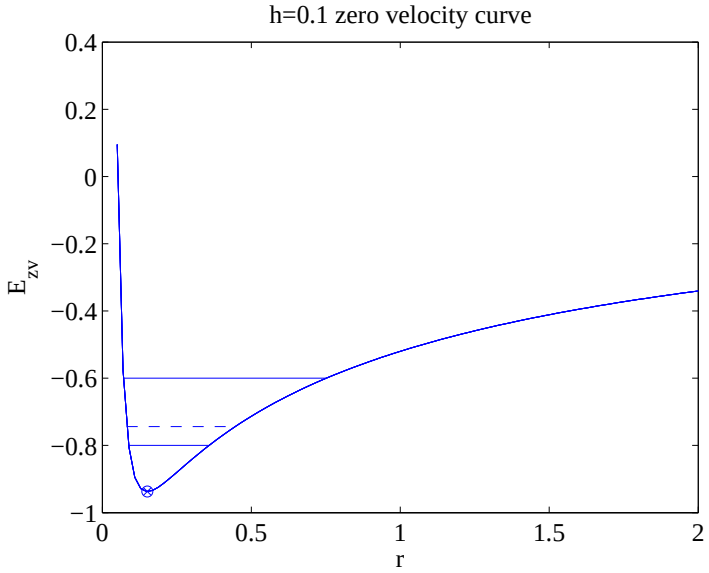

Figure 4.4: Zero Velocity Curves and sample orbits at $h = 0.1$; $E = -0.6, E_{\text{crit}}, -0.8, E_{\text{circ}}$

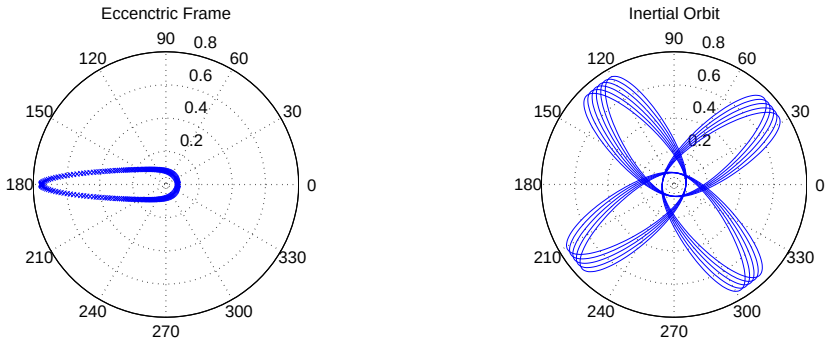

Figure 4.5: $E = -0.6$ orbit in eccentric (left) and inertial (right) frames

Upon integrating (4.18), one obtains $\omega(f)$, which can be seen for this orbit plotted in Fig. 4.6. One sees $\Delta\omega = \omega(2\pi) - \omega(0)$ is the *turning angle* of the rosette. For energies prior to (above) the critical bifurcation energy, we find a secular retrograde rotation of the eccentric frame. $\theta(t)$, $\omega(t)$, and $f(t)$ are plotted against time over three standard orbits on the right. One sees a secular prograde growth in the inertial polar angle θ, with turning angle $\Delta\theta = 2\pi + \Delta\omega$. (Recall that $\Delta\omega < 0$).

Finally, we compute the osculating semi-major axis and eccentricity vs. time

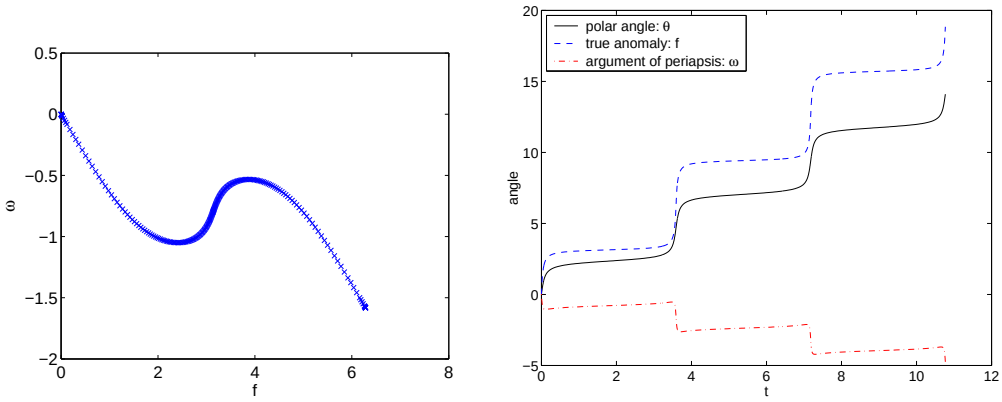

Figure 4.6: $\omega(f)$ (left) and $\theta(t),\ \omega(t),\ f(t)$ (right) for $E = -0.6$

(Fig. 4.7) over one nominal orbit (as seen from the eccentric frame). The solid lines represent the osculating elements as provided by the eccentric frame method, see (4.13)-(4.14). The dashed curves are a standard set using the osculating orbital element transformation as defined by classical perturbation theory, using $\mu_0 = 1$ for the "planet" mass, i.e. μ_0 is the gravitational parameter of the *total* halo mass *plus* the mass of the central black hole.

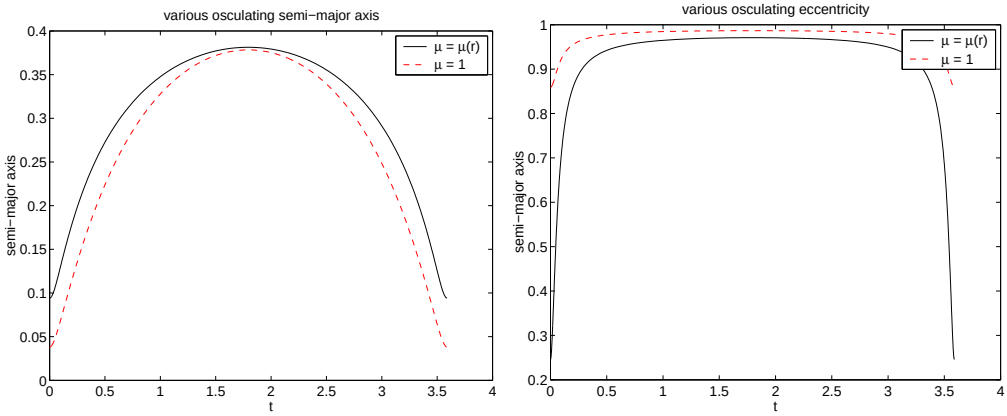

Figure 4.7: Osculating orbital elements for $E = -0.6$

For a sample orbit with $E < E_{\text{crit}}$, we take $E = -0.8$. The orbit as seen from the eccentric and inertial frames is shown in Fig. 4.8. Notice that the particle now makes librations in the eccentric frame. The libration paths become smaller and

smaller until they degenerate to a single point at $E_{\rm circ}$, to be considered next.

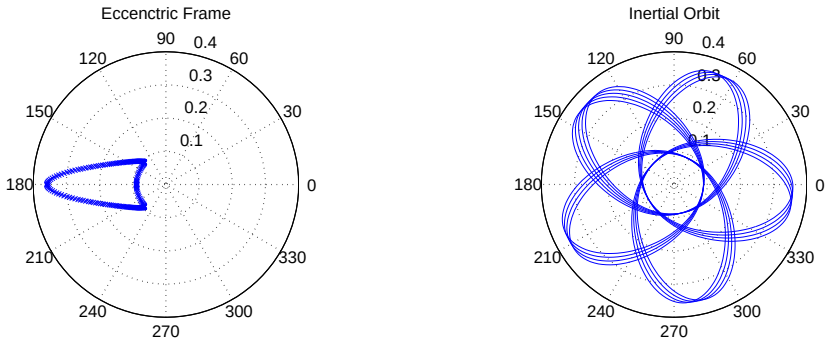

Figure 4.8: $E = -0.8$ orbit in eccentric (left) and inertial (right) frames

$\omega(f)$ for this orbit is plotted in Fig. 4.9. Notice that f librates around $f = \pi$ and there is secular prograde growth in ω. The turning angle is still given by $\Delta\omega$. For energies after (below) the critical bifurcation energy, we find a secular prograde rotation of the eccentric frame. $\theta(t)$, $\omega(t)$, and $f(t)$ are plotted against time over three standard orbits on the right. One now sees a secular posigrade growth in the argument of periapsis, coupled with librations in true anomaly f.

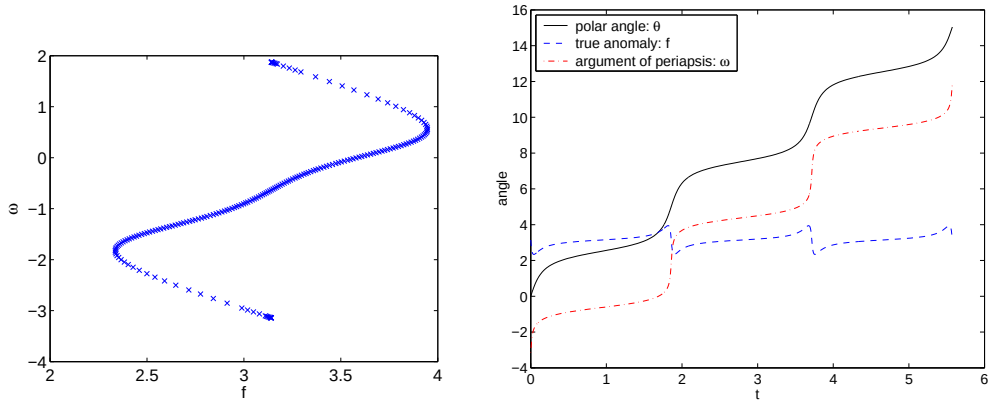

Figure 4.9: $\omega(f)$ (left) and $\theta(t)$, $\omega(t)$, $f(t)$ (right) for $E = -0.8$

Finally, we compute the osculating semi-major axis and eccentricity vs. time (Fig. 4.10) over one nominal orbit (as seen from the eccentric frame). The solid lines represent the osculating elements as provided by the eccentric frame method, see

(4.13)-(4.14). As before, the dashed curves are a standard set using the osculating orbital element transformation as defined by classical perturbation theory, using $\mu_0 = 1$ for the "planet" mass, i.e. μ_0 is the gravitational parameter of the *total* halo mass *plus* the mass of the central black hole.

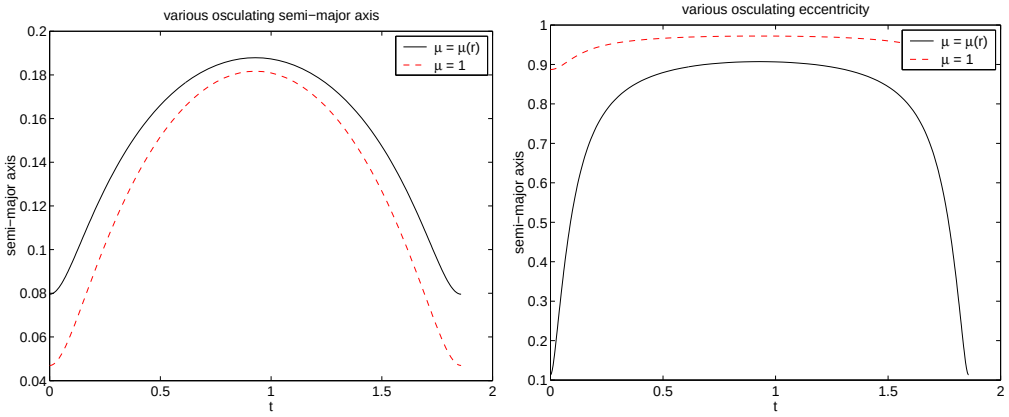

Figure 4.10: Osculating orbital elements for $E = -0.8$

At the circular energy $E = E_{\text{circ}}$, the orbit in the eccentric frame degenerates to a single point. The orbit is circular in the inertial plane, Fig. 4.11. The eccentric frame now precesses at a uniform rate.

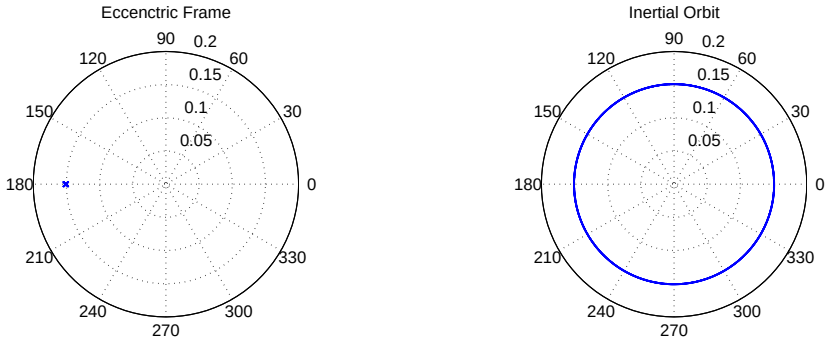

Figure 4.11: $E = -0.9372$ orbit in eccentric (left) and inertial (right) frames

$\omega(f)$ (left) and $\theta(t), \omega(t), f(t)$ are plotted below in Fig. 4.12. Now f is virtually constant and there is uniform growth in ω.

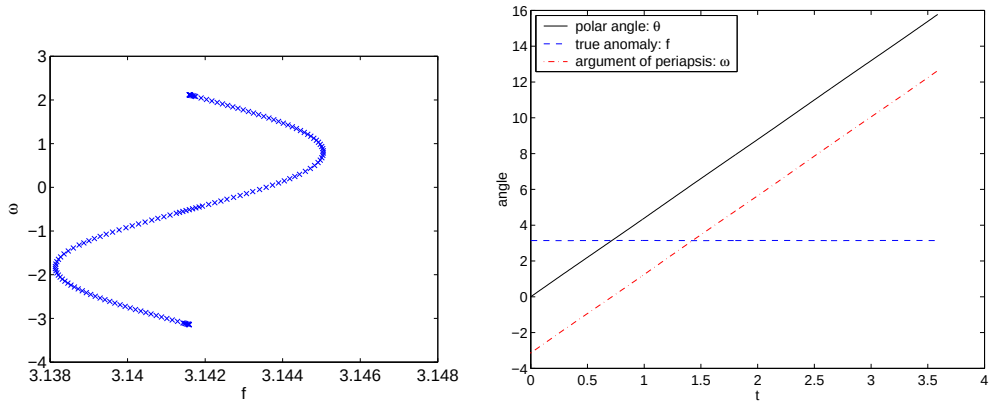

Figure 4.12: $\omega(f)$ (left) and $\theta(t),\ \omega(t),\ f(t)$ (right) for $E = -0.9372$

Finally, when observing the semi-major axis and eccentricity of the orbit (Fig. 4.13), we see something counterintuitive. The osculating eccentricity is close to 0.62. If one, on the other hand, used a classical definition of osculating orbital elements, as previously discussed, the osculating eccentricity would be close to 0.94. We thus see a circular orbit (in inertial space) with high osculating eccentricity. The osculating ellipse is a highly eccentric one that always touches the true path at apoapsis. In this way, the osculating ellipse rotates synchronously with the particle so that the particle is always at apoapsis and the true motion is a circular path.

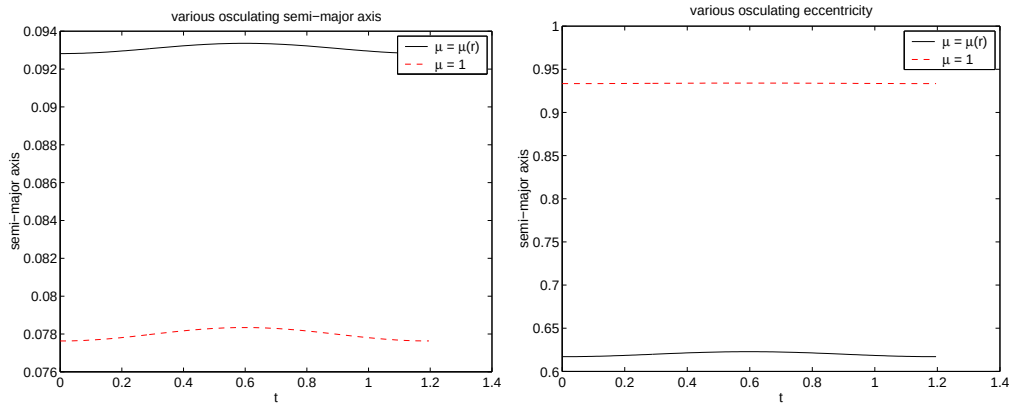

Figure 4.13: Osculating orbital elements for $E = -0.9372$

4.7 Stability Analysis of Equatorial Rosettes in Axi-symmetric Potentials

In this section we will consider the stability of the planar equatorial motion of a particle in an axi-symmetric potential. On the plane of symmetry, the potential reduces to a two-dimensional central force field. Since the motion is periodic with respect to the eccentric frame, the technique justifies application of Floquet theory for the stability analysis of the in-plane motion, where the period used is the period of the orbit in the eccentric frame.

4.7.1 Floquet Theory

Suppose now we are considering a conservative system with axi-symmetric equipotential contours, so that, in cylindrical coordinates, the potential energy is given by $U(r, z)$. We can define a variable gravitational parameter by $\mu(r, z) = rU(r, z)$. The Hamiltonian of the system is thus given by:

$$H = \frac{1}{2}\left(v_r^2 + \frac{h_z^2}{r^2} + v_z^2\right) - U(r, z)$$

where the symplectic coordinates are $\langle r, v_r, \theta, h_z, z, v_z\rangle$. The equations of motion are:

$$\begin{aligned} \dot{r} &= v_r & \dot{v}_r &= \frac{h_z^2}{r^3} + \frac{\partial U(r,z)}{\partial r} \\ \dot{\theta} &= \frac{h_z}{r^2} & \dot{h}_z &= 0 \\ \dot{z} &= v_z & \dot{v}_z &= \frac{\partial U(r,z)}{\partial z}. \end{aligned}$$

In particular, equatorial motion with $z \equiv 0$ is well defined, and this reduces to the problem of motion in a central force field. The resulting motion follows a rosette-shaped path in the equatorial plane. We now ask whether this motion is stable under a small out-of-plane perturbation δz. To do this, we consider the 2×2 out-of-plane

State Transition Matrix (STM) Φ, which, by definition, gives:

$$\begin{bmatrix} \delta z(t) \\ \delta \dot{z}(t) \end{bmatrix} = \Phi \cdot \begin{bmatrix} \delta z(0) \\ \delta \dot{z}(0) \end{bmatrix}.$$

The STM is determined by integrating the following differential equation:

$$\dot{\Phi}(t) = \begin{bmatrix} 0 & 1 \\ U_{zz}(r,0) & 0 \end{bmatrix} \cdot \Phi(t), \qquad \Phi(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where the coefficient matrix is evaluated along the nominal orbit ($z = 0$) in the equatorial plane. Since the equatorial motion reduces to a central force field problem, there exists an eccentric frame decomposition, in which the motion is periodic. Let T be the period of a single orbit in the eccentric frame. The coefficient matrix, above, only depends on $r(t)$, and thus it is T-periodic. We are therefore justified in using Floquet theory in the stability analysis. See Cesari [32] for an introduction on this. Let λ_1, λ_2 be the eigenvalues of $\Phi(T)$. Since $\lambda_1\lambda_2 = 1$, either the eigenvalues are complex conjugates on the unit circle in the complex plane, or real-valued with $\lambda_2 = \lambda_1^{-1}$. The rosettal motion on the equatorial plane is therefore stable if and only if both eigenvalues are on the unit circle. A bifurcation from stable to unstable must occur when $\lambda_1 = \lambda_2 = 1$.

4.7.2 Application to a toy axi-symmetric potential

To show how this theory might be applied, we will consider the following toy potential:

$$U(R) = \frac{1}{R}\left(1 - \frac{1}{1+R}\right) = \frac{\mu(R)}{R}, \quad \text{where} \quad R = \sqrt{r^2 + \frac{z^2}{a^2}}.$$

Here, $r = \sqrt{x^2 + y^2}$ and $a > 0$ is a parameter. When $a < 1$, the potential is oblate spheroidal, and when $a > 1$, it is prolate. This potential is motivated by replacing

r with R in (4.26), but does not have direct physical significance. Its utility to us is only to illustrate the theory. The potential reduces to the Hernquist potential when restricted to the equatorial plane. The question now arises, for various values of the parameter a, when is the equatorial motion due to out of plane perturbations stable? Clearly, for $a = 1$, the motion is stable due to the angular momentum integral. To proceed, we compute the out of plane dynamics:

$$\ddot{z} = \frac{\partial U}{\partial z} = \frac{-\mu(R)}{R^3}\frac{z}{a^3} + \frac{\mu'(R)}{R}\frac{z}{Ra^2}.$$

For a small perturbation δz, we obtain:

$$\delta\ddot{z} = \omega(t)\delta z = \left(\frac{-\mu(r)}{a^2r^3} + \frac{\mu'(r)}{a^2r^2}\right)\delta z.$$

The coefficient $\omega(t)$ is a function of time because we have an explicit solution for $r(t)$ for the nominal motion along the equatorial plane. The out of plane State Transition Matrix (STM) for $\langle \delta z, \delta\dot{z}\rangle^T$ can then be written as:

$$\dot{\Phi} = \begin{bmatrix} 0 & 1 \\ \omega(t) & 0 \end{bmatrix} \cdot \Phi, \qquad \Phi(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The STM Φ can now be integrated along with the nominal planar solution, as its dynamic coefficient matrix depends only on $r(t)$. We can now systematically integrate $\dot{\Phi}$ between $t \in [0, T]$, where T is the eccentric frame orbital period. Computing the eigenvalues of $\Phi(T)$ reveals the stability of the planar equatorial orbit.

We followed this procedure for a sampling of different energy levels and axis ratios $1 : 1 : a$. The result for oblate potentials is shown in Fig. 4.14. The grid points with dots correspond to parameter values of a and E for which the equatorial motion is unstable. For prolate potentials, the result is similarly depicted in Fig. 4.15. Notice in both plots, for $a = 1$, that the planar motion is stable for all energies.

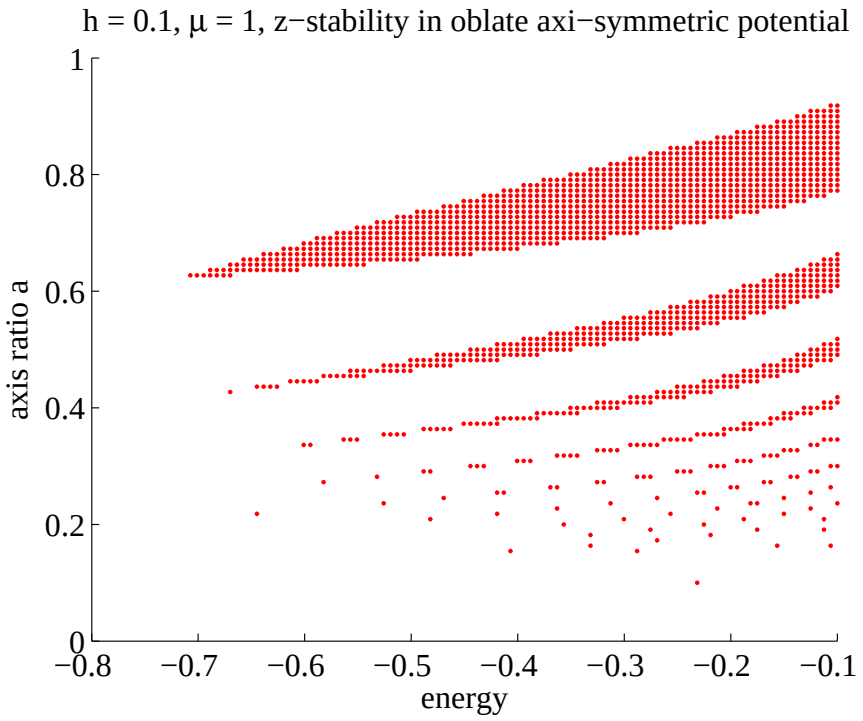

Figure 4.14: Equatorial stability plot for oblate potential

4.8 Conclusion

In this paper we presented a preferred, nonuniformly rotating frame that exists for motion in any central force field, with respect to which the orbital motion is periodic. We showed that for high values of energy, the particle trajectories in the eccentric frame make circulations. However, when the energy drops beneath a certain critical level, the trajectories follow librations in the eccentric frame. This is not a true bifurcation of the system, as there is no distinguishable physical change when the orbits are viewed with respect to inertial space, yet it is a necessary transition that must occur as one nears the minimum circular orbit energy. For circular orbits (in inertial space), motion in the eccentric frame degenerates from librations to a single fixed point. For this case, the eccentric frame rotates at a constant rate, and a circular orbit in inertial space is observed. Further we showed that, even in the

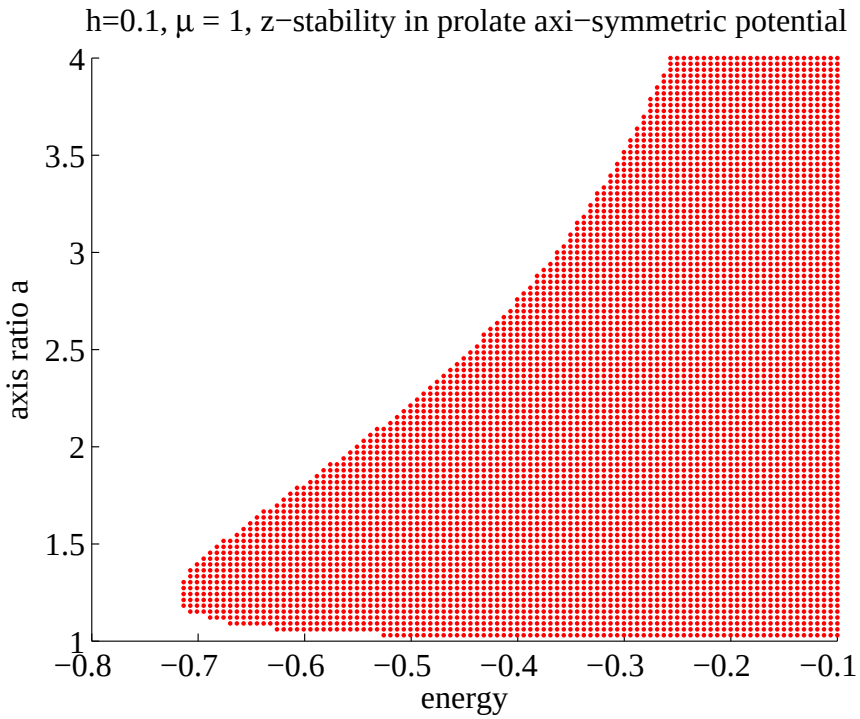

Figure 4.15: Equatorial stability plot for prolate potential

case of a circular orbit, the osculating eccentricity can be very high. This occurs because the particle is "stuck" to the periapsis or apoapsis of the osculating ellipse. The osculating ellipse has static (high) eccentricity and rotates at a uniform rate. We also presented a model for the potential energy of a Hernquist galaxy with a central black hole, analyzed the rosette-shaped orbits, and then compared them to the orbit as seen from the eccentric frame for various parameter values. Finally we indicated how one might use the eccentric frame to determine stability of planar orbits in axi-symmetric potentials.

CHAPTER V

The Variational Principles of Nonholonomic Mechanics

In this chapter we will begin our discussion on the second main theme of this thesis-nonholonomic systems. In particular, we will strive to formulate a precise geometric understanding of the role quasi-velocities have on nonholonomic mechanical and control systems on manifolds. We will discuss the nature of these constraints at a differential level; in particular, the transitivity choice one must make in defining variations of curves in nonholonomic systems; and relate this to the proper application of variational principals in obtaining equations of motion for the system. Finally we will show how one can apply quasi-velocity techniques to unconstrained systems whose underlying configuration manifold is a Lie group. Then in Chapter VI we will extend the techniques discussed in this present chapter to a higher order form that is suitable for optimal control problems.

5.1 Introduction

5.1.1 Nonholonomic Mechanics

A *nonholonomic* system is one in which there are a set of nonintegrable constraints on the velocities. In this way, every point in your phase space or on your configuration manifold can be dynamically obtained, in principal, by appropriately "steering" your system from one configuration to another. So it is not the set of points that is

restricted, rather the types of motion that can ensue. The canonical example of a nonholonomic system is that of an ice skate. We will take as generalized coordinates the set (x, y, θ), which represent the two-dimensional position of the ice skate, and the angle the blade makes with a standard axis. By experience you know that you could, starting at an arbitrary position and angle, reach any other position and angle, assuming that you can ice skate. In this way, an n degree of freedom system with m nonholonomic constraints still requires n generalized coordinates.

This is to be contrasted with a system whose constraints are integrable. Take for example the spherical pendulum. One could take (x, y, z), the position of the bob, as generalized coordinates. However, the constraint $x^2 + y^2 + z^2 = 1$ limits the set of points the system can access. In this way, one could reduce the dimensionality of the system by the number of holonomic constraints. One could instead use the polar and azimuthal angles (θ, ϕ) as generalized coordinates for the system and be done with it. An n degree of freedom system with m *holonomic* constraints requires only $n - m$ generalized coordinates. Because of this, without loss of generality, we will only consider systems with nonholonomic constraints. The difference between nonholonomic and holonomic systems is discussed in most books on mechanics, in particular see Bloch [16], Goldstein [54], Greenwood [56, 57, 58], and Papastavridis [113].

We will also discuss the variational principles of nonholonomic mechanics. Variational principles play a key role in theoretical physics, as all of the laws of mechanics are deriveable from them (Lanczos [72]) as well as laws of electromagnetism and gravitation (Landau and Lifshitz [73], Pauli [114]). To obtain the correct dynamical equations of motion, one must enforce the variational principle known as the *Principle of Virtual Work*, which states that virtual displacements to the actual curve must

satisfy the nonholonomic constraints. We will discuss the meaning of this statement and how one enforces it throughout the chapter. The Principle of Virtual Work leads to the fundamental nonholonomic form of Lagrange's equations, given in Theorem 32.

The birth of nonholonomic mechanics was surrounded with an air of confusion, as the correct dynamical equations of motion *do not satisfy Hamilton's Principle*. Hamilton's Principle states that the correct dynamical path between two points in the configuration manifold is the one that minimizes the action

$$I = \int L(q(t), \dot{q}(t))\ dt$$

The equations of motion that a curve must satisfy in order to be an extremal of the above action are given in Theorem 33, and are referred to as the *vakonomic equations of motion*. The term vakonomic was originally introduced by V.I. Arnold [4]. The term itself was based off the acronym VAK, which stood for motion of the *variational axiomatic kind*. In general the vakonomic equations of motion produce the *incorrect* dynamical equations of motion; however, in some special cases the two sets of equations produce the same result, see Favretti [48] or Cortes et al. [34] for more on this. One must be weary, as the vakonomic equations of motion erroneously appear in place of the fundamental nonholonomic form of Lagrange's equations even in the classical text Goldstein [54], for example.

Even though the motion of mechanical systems with nonholonomic constraints does not satisfy Hamilton's principle in general, vakonomic dynamics has found applications in control theory (Bloch and Crouch [17]), economics (Samuelson [121] and Sato and Ramachandran [122]), and the motion of microorganisms (Shapere and Wilczek [125]). Some recent papers that discuss nonholonomic dynamical and vakonomic systems from a geometric viewpoint are Cardin and Favretti [29], de Leon et

al. [41], Ehlers et al. [47], Gracia et al. [55], Lewis and Murray [79], and the reference contained therein. Several important texts on the subject are Bloch [16], Greenwood [58], Neimark and Fufaev [106], and Papastavridis [113]. In particular, the texts by Greenwood and Neimark and Fufaev talk extensively on a series of techniques based on quasi-velocities. Using a quasi-velocity approach affords one an arsenal of different methodologies with which one can write down the dynamical equations of motion. Each of these approaches has its own advantages and disadvantages, but in each you can write down the equations of motion using fewer differential equations than would be obtained using the fundamental nonholonomic form of Lagrange's equations. It is not our goal here to discuss and compare these different approaches, as this has been done. Rather, our goal is to understand quasi-velocities and the variational principles of nonholonomic and vakonomic mechanics from a more geometric basis. We will also extend the quasi-velocity techniques written for nonholonomic dynamics to the case of vakonomic motion. This seems at first a step in vain, as the vakonomic motion is the incorrect dynamical motion; however it has been shown (Bloch and Crouch [17]) that the vakonomic motion is equivalent to an optimal control problem under certain circumstances. The vakonomic equations of motion are also related to mathematical problems in the calculus of variations, see Kirk [70], Sagan [120], Smith [127]. Moreover, by extending these techniques to their corresponding vakonomic form, we gained insight in the geometry of the variational principles themselves and were able to further generalize these techniques, using this insight, to a higher order version suitable to dynamical optimal control problems. We will discuss this generalization in Chapter VI.

5.1.2 Chapter Overview

§5.2 and 5.3 are preliminary sections in which we provide the basic definitions for quasi-velocities, nonholonomic constraints, and variations. In §5.4 we introduce a new connection with respect to which the basis $\mathscr{E}$ is covariantly constant. We discuss its basic properties and relate it to Ehresmann connections, and then use it to derive the transpositional relations, a set of equations key to understanding the variational principles of nonholonomic mechanics. In §5.3.4 we discuss the geometry of constrained variations in more depth, and the corresponding transitivity choice one must make in defining variations in nonholonomic systems. The transitivity choice allows for two main possible definitions of variations, though one may define variations of mixed type. One must be completely aware of which transitivity choice one is using in order to correctly execute any variational principle in obtaining the equations of motion. There is then still freedom left in how one carries out the variational principles, a topic discussed in §5.6. In this section we will strive to elucidate the difference between the mechanical and vakonomic variational principles, and derive the corresponding equations of motion for each case. In §5.8 we discuss Maggi's equations, and generalize them to the vakonomic case. We do the same with the Boltzmann-Hamel equations in §5.9.

5.1.3 Summation Convention

To aid in notation, we will invoke the summation convention of Table 5.1. This convention will be used throughout the current chapter and the next.

Table 5.1: Summation Convention

letter type	summation over
$\alpha, \beta, \gamma, \ldots$	$1, \ldots, m$
$A, B, C, \ldots$	$m+1, \ldots, n$
$a, b, c, \ldots$	$1, \ldots, n$

5.2 Nonholonomic Constraints and Quasi-Velocities

Let Q be the configuration manifold of our system, with $\dim Q = n$, and TQ its corresponding phase space or tangent bundle. Let $L : TQ \to \mathbb{R}$ be a Lagrangian function defined on the tangent bundle. In many physical systems, the Lagrangian takes the form

$$L(q, \dot{q}) = \frac{1}{2} g_{ij} \dot{q}^i \dot{q}^j - V(q)$$

where g_{ij} is the kinetic energy metric and $V(q)$ is the potential. In the discussion that follows, however, we will not require the Lagrangian to take this form.

We further suppose our system is subject to m linear nonholonomic constraints. We will further take the constraints to be scleronomic, i.e., time does not appear explicitly in the constraint equations. The constraints then take the following form:

$$a_i^\sigma(q)\dot{q}^i = 0 \tag{5.1}$$

where rank $a = m$. In the classical study of nonholonomic mechanics, a number of techniques have been developed based on entities called *quasi-velocities*. For a review of these techniques see both Neimark and Fufaev [106] and Greenwood [58]. A more modern definition of quasi-velocities can be found in Bullo and Lewis [27], though there they are referred to as pseudo-velocities. These functions are defined as follows:

Definition 9. *Let $c : [a, b] \to Q$ be a smooth curve on the differentiable manifold Q and let $\mathscr{X} = \{X_1, \ldots, X_n\}$ be a set of vector fields in TQ that form a basis for the tangent space T_qQ, for every $q \in Q$. Define the n functions $u^k : I \to \mathbb{R}$ by the relation $c'(t) = u^k(t)X_k(\gamma(t))$. The n functions $u^k(t)$ are the **quasi-velocities** of the curve c with respect to the bases given by the vector fields $\mathscr{X}$.*

Define now a mapping $\mathbf{\Psi} : TQ \to TQ$ on the tangent bundle, such that for each point $q \in Q$, the transformation $\mathbf{\Psi}(q) : T_qQ \to T_qQ$ is a vector space isomorphism,

with the corresponding matrix of transformation Ψ. $\mathbf{\Psi}(q)$ is defined so that $\Psi_i^\sigma = a_i^\sigma$ and so that $\mathbf{\Psi}(q)$ is invertible, with $\mathbf{\Phi}(q) = \mathbf{\Psi}^{-1}(q)$. In this way, the leading m rows of the matrix Ψ coincide with the constraint coefficient matrix a_i^σ. The remaining rows can be chosen freely, so long as Ψ is invertible.

A useful identity that we will make use of manifold is derived from the relation $\Psi_s^i \Phi_j^s = \delta_j^i$. Differentiating with respect to q^k we obtain:

$$\Psi_s^i \frac{\partial \Phi_j^s}{\partial q^k} = -\Phi_j^s \frac{\partial \Psi_s^i}{\partial q^k} \tag{5.2}$$

Now define a set of n differential forms

$$\omega^j = \Psi_i^j dq^i \tag{5.3}$$

Notice that if we write the velocity of the system in terms of a coordinate basis as:

$$\mathbf{v} = \dot{q}^k \frac{\partial}{\partial q^k}$$

then the nonholonomic constraints (5.1) can be rewritten as:

$$\omega^\sigma(\mathbf{v}) \equiv 0$$

It is for this reason that the m differential forms ω^σ are often referred to as the *constraint one-forms*.

The linear mapping $\mathbf{\Psi} : TQ \to TQ$ defines a set of n vector fields $\mathscr{E} = \{E_1, \ldots, E_n\}$ on TQ by the relation:

$$E_i(q) = \Phi_i^j(q) \frac{\partial}{\partial q^j} \in T_q Q \tag{5.4}$$

For each $q \in Q$ these vectors form a basis of the tangent space $T_q Q$. Notice that at least m of the the dual basis one-forms $E^i(q) \in T^*Q$ are not exact. In fact the first m one-forms of the dual basis coincide with the constraint one-forms. The n components u^i of the velocity vector $\mathbf{v}$ with respect to this basis constitute a set

of quasi-velocities for the system. Notice that, by construction, the first m quasi-velocities $(u^1, \ldots, u^m)$ must vanish identically along any trajectory consistent with the nonholonomic constraints. Since the dual basis vectors are not exact, these functions are not the time derivative of any function. They are, however, related to the time derivatives of the coordinates $q^i(t)$ of the solution trajectory by the kinematic relations (5.4). The velocity vector can be described equivalently in any of the following forms:

$$\mathbf{v} = \dot{q}^i \frac{\partial}{\partial q^i} = u^j E_j = \left(\Psi_i^j \dot{q}^i\right) E_j = \left(\Phi_j^i u^j\right) \frac{\partial}{\partial q^i} \tag{5.5}$$

from which we see

$$u^j = \Psi_i^j(q)\dot{q}^i \qquad \text{and} \qquad \dot{q}^i = \Phi_j^i(q)u^j \tag{5.6}$$

The relations (5.6) are kinematic relations that allow one to transform from the true velocities to the quasi-velocities and back again. The two bases transform in the obvious way:

$$E_j = \Phi_j^i \frac{\partial}{\partial q^i} \qquad \text{and} \qquad \frac{\partial}{\partial q^i} = \Psi_i^j E_j \tag{5.7}$$

Quasi-velocities arise in a very natural context in the study of rigid body mechanics. One can take as general coordinates for the configuration manifold $SO(3)$ a set of Euler angles. The Euler angle rates are true velocities. However, it is often notationally advantageous to write down the dynamical equations of motion in terms of the components of the body fixed angular velocity vector, $\boldsymbol{\omega} = \langle \omega_x, \omega_y, \omega_z \rangle$. The variables ω_x, ω_y, and ω_z, however, are *not* time derivatives of functions of the generalized coordinates. They therefore constitute a set of quasi-velocities for the system. In solving for the attitude of the body as a function of time, one then needs to integrate a set of kinematic relations that relate the Euler angle rates to these quasi-velocities. We'll come back to this example several times throughout the remainder of this dis-

sertation.

5.3 Variations

In this section we will discuss some of the geometric theory of variations. The mathematical theory of variations is laid out in Do Carmo [30], Dubrovin et al. [46], and Gelfand and Fomin [53]. We will follow the more modern geometric formalism found in Bullo and Lewis [27] and Marsden and Ratiu [83]. We begin with a discussion of variations on curves, and then discuss how one uses this to take variations of functions and functionals.

5.3.1 Variations of Curves

We begin by making the following definition:

Definition 10. *Let Q be a manifold and consider two fixed points $q_1, q_2 \in Q$ and an interval $[a, b] \subset \mathbb{R}$. The **path space** from q_1 to q_2 is the set*

$$\Omega(q_1, q_2, [a, b]) = \{c : [a, b] \to Q |\ c \text{ is a } C^2 \text{ curve}, c(a) = q_1, c(b) = q_2\} \tag{5.8}$$

The set of curves $\Omega(q_1, q_2, [a, b])$ is actually an infinite-dimensional differentiable manifold, see Marsden and Ratiu [83]. One can also easily show that, for a given $c \in \Omega(q_1, q_2, [a, b])$, that the tangent space at c is given by:

$$T_c\Omega(q_1, q_2, [a, b]) = \left\{ v : [a, b] \to TQ \;\middle|\; \begin{array}{c} v \text{ is a } C^2 \text{ map}, \;\; \pi_Q \circ v = c, \\ \text{and } v(a) = v(b) = 0 \end{array} \right\} \tag{5.9}$$

Here, $\pi_Q : TQ \to Q$ is the projection operator. To show this, let us first define what we mean by a *variation of c*.

Definition 11. *Consider a smooth curve $c : [a, b] \to Q$. A **variation** of c is a differentiable function $\vartheta : [-\varepsilon, \varepsilon] \times [a, b] \to Q$ that satisfies the following conditions:*

(i) $\vartheta(0,\tau) = c(\tau),\ \forall\ \tau \in [a,b]$

(ii) $\vartheta(s,a) = c(a)$ *and* $\vartheta(s,b) = c(b),\ \forall\ s \in [-\varepsilon,\varepsilon]$.

The parameterized curve $\vartheta_s(\tau) := \vartheta(s,\tau)$*, where* s *is held fixed, is called a* ***curve in the variation****. The parameterized curve* $\vartheta_\tau(s) := \vartheta(s,\tau)$*, where* τ *is held fixed, is called a* ***transversal curve of the variation****.*

The image of the variation itself is a two-dimensional submanifold of Q which contains the curve c and is parameterized by the variables s and τ. On the curve c, we have $t = \tau$, though this does not necesarily hold elsewhere in the variation. The reason for this distinction is that in mechanics time has a physical meaning. The variation itself, however, is a mathematical construct. The question, 'at what time was the particle at such and such point in the variation,' is ill-posed, unless either the point in question lies on the actual trajectory or we define what we mean by time throughout the entire variation. As a result, the velocity $c'(t)$ is defined *along the curve* c, but not elsewhere in the variation. For a fixed τ, one cannot differentiate the velocity with respect to s without further definition, since the velocity is a priori only defined at $s = 0$. This is not an issue for the mathematical theory of the calculus of variations, holonomic mechanics, and the Hamel formulation of nonholonomic mechanics, so it is not surprising that it has caused so much confusion. Its imortance only arises in the study of nonholonomic mechanics from the Suslov approach or from a fiber bundle approach. The idea that one must define the velocity on the variation, and that there are alternative ways of doing so, was first recognized by Neimark and Fufaev [105, 106].

Notice that if we view the curve c as a point in the manifold $\Omega(q_1, q_2, [a,b])$, that a variation of c is a smooth curve c_s through $\Omega(q_1, q_2, [a,b])$ with $c_0 = c$. This curve

therefore defines a vector in the tangent space, i.e. $dc_s/ds \in T_c\Omega(q_1, q_2, [a, b])$. Since, for each $s \in [-\varepsilon, \varepsilon]$, the curve $c_s(\tau) \in Q$ has fixed endpoints, we have $dc_s/ds(a) = dc_s/ds(b) = 0$. This shows that the tangent space of $\Omega(q_1, q_2, [a, b])$ at c is given by the set (5.9). Elements of the tangent space $v \in T_c\Omega(q_1, q_2, [a, b])$ are usually denoted

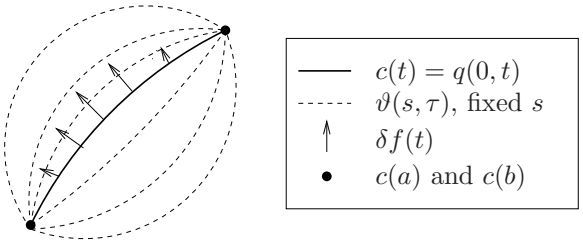

Figure 5.1: A Variation of $c \in \Omega(q_1, q_2, [a, b])$

δc and, in terms of the manifold Q, are referred to as infinitesimal variations of the curve c. Formally we write:

Definition 12. *The **infinitesimal variation** δc corresponding to the variation $\vartheta : [-\varepsilon, \varepsilon] \times [a, b] \to Q$ is a vector field defined along the curve c, i.e. $\delta c : [a, b] \to T_cQ$, that is defined by*

$$\delta c(t) = \left.\frac{\partial \vartheta(s, \tau)}{\partial s}\right|_{s=0,(t=\tau)}$$

If a set of coordinates $\{q^i\}$ are given, we refer to the components of δc with respect to the coordinate basis as:

$$\delta c = \delta q^i \frac{\partial}{\partial q^i}$$

The infinitesimal variation δc is a vector field defined only along the curve c, upon which the variable τ coincides with time. In terms of notation, δ is an operator that represents a derivative, evaluated at the nominal curve c, in the direction of the variation s. We will also use the notation

$$\partial_s = \frac{\partial}{\partial s} \qquad \text{and} \qquad \partial_\tau = \frac{\partial}{\partial \tau}$$

to represent partial derivatives with respect to s and τ, respectively, for any point in the variation. Finally the operator d_t will sometimes be used in place of d/dt to represent the time derivatives.

We can also represent the quasi-infinitesimal variation in terms of the basis $\mathscr{E}$ as follows.

Definition 13. *Let $c : [a,b] \to Q$ be a curve, $\vartheta : [-\varepsilon,\varepsilon] \times [a,b] \to Q$ a variation of c, and $\delta c : [a,b] \to TQ$ its corresponding infinitesimal variation. Then the* ***quasi-infinitesimal variations*** *are the n scalar functions $\zeta^i : [a,b] \to \mathbb{R}$ which are the components of δc with respect to the basis $\mathscr{E}$:*

$$\delta c(t) = \zeta^i(t) E_i(c(t)) \tag{5.10}$$

5.3.2 Variations of Velocity

Consider the curve $c : [a,b] \to Q$ and its variation $\vartheta : [-\varepsilon,\varepsilon] \times [a,b] \to Q$, as defined previously. Both the infinitesimal variation δc and the velocity $d_t q$ are vector fields which are only defined along the curve c itself, and not elsewhere in the variation. In this section we will discuss the meaning of the operators $d_t\delta$ and δd_t. Since δc is defined along c, the derivation $d_t \delta c$ is also defined along c. However, along a given transversal curve of the variation $\vartheta_\tau(s)$, the velocity $d_t q$ is only defined for $s = 0$. Without defining the meaning of $d_t q$ throughout the variation, the quantity $\delta d_t q$ is undefined.

Definition 14. *Given a curve $c : [a,b] \to Q$ on the manifold Q and a variation $\vartheta : [-\varepsilon,\varepsilon] \times [a,b] \to Q$, a vector field $V : [-\varepsilon,\varepsilon] \times [a,b] \to TQ$ is an* ***extended velocity field on q*** *if each of the following conditions hold:*

(i) $V(0,\tau) = c'(\tau)$,

(ii) $V(s,a) = V(s,b) = 0$ *for all* $s \in [-\varepsilon,\varepsilon]$, *and*

(ii) $\pi_Q(\text{image}(V)) = \text{image}(\vartheta) \subset Q$, *where* $\pi_Q : TQ \to Q$ *is the projection operator.*

An infinitesimal variation of the velocity $\delta\dot{c}$ is not defined without defining an extension V of the velocity $\dot{c}$ to the variation q. Notice that this definition does not require that the extended velocity vectors to be tangent to the variation, except of course along the curve c. One possible way of extending the velocity to the variation $q(\varepsilon, \tau)$ is by considering the variation to be *contemporaneous*. A contemporaneous variation is one in which the paramter τ is considered to coincide with time throughout the variation. Contemporaneous variations, typically denoted $\vartheta(\varepsilon, t)$. To define a variation q to be contemporaneous is to define the extended velocity field ubiquitously, namely as:

$$V(s,t) = \frac{\partial \vartheta(s,t)}{\partial t}$$

Due to the continuity of the variation ϑ it follows that the operators d_t and δ must commute, i.e. $\delta V^i = d_t \delta q^i$. See Fig. 5.2. Variations are not always taken to be contemporaneous. Rather this is a definition for the extension of both time and the velocity from the curve c to the entire variation ϑ. In this definition, the extended velocities are the tangent vectors of the curves in the variations. This is, however, not how the velocity field is extended in the Suslov formulation of nonholonomic mechanics. We'll come back to this in §5.3.4.

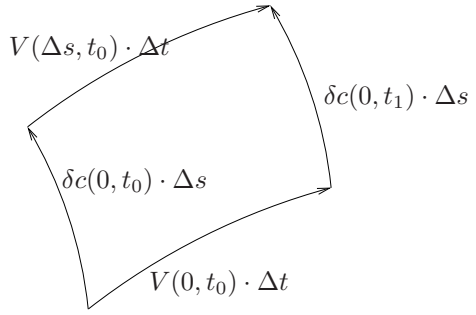

Figure 5.2: Contemporaneous Variations (here $t_1 = t_0 + \Delta t$)

5.3.3 Variations of Functions

In the preceding section we saw how an infinitesimal variation of the velocity is undefined until one defines an extended velocity field throughout the variation. When considering a function $f : TQ \to \mathbb{R}$, it is desirable to define its variation $\delta f : TQ \to TQ$ in a way that is independent of the choice of an extended velocity field, that agrees with the classical definition, and that respects the Lagrange-D'Alembert principle. We will first consider the classical case of functions on $\mathbb{R}^n$ and then extend the definition to arbitrary manifolds.

Variations of Functions on $\mathbb{R}^n$

In the classical theory of Calculus of Variations, variations of functions are defined in the following way. See for instance Gelfand and Fomin [53]. Given a curve $c : [a, b] \to \mathbb{R}^n$ and a function $f : T\mathbb{R}^n \to \mathbb{R}$, one defines a vector field $\eta : [a, b] \to T\mathbb{R}^n$ along the curve c such that $\eta(a) = \eta(b) = 0$. Given a set of coordinates $\{x^i\}$ for $\mathbb{R}^n$, the curve c may be expressed in coordinates as $c = x^i(t)$. One then defines a one parameter family of varied curves by $c_s(t) = x^i(t) + s\eta(t)$, where s is a fixed constant. The infinitesimal variation of the function f is then defined by:

$$\delta f(t) = \frac{\partial}{\partial s} \left(f(x^i(t) + s\eta(t), \dot{x}^i(t) + s\dot{\eta}(t)) \right) \Big|_{s=0}$$

where the dot represents a derivative with respect to t. A brief calculation then yields the following formula for the infinitesimal variation of f:

$$\delta f(t) = \frac{\partial f(x^i(t), \dot{x}^i(t))}{\partial x^i} \eta^i(t) + \frac{\partial f(x^i(t), \dot{x}^i(t))}{\partial \dot{x}^i} \frac{d\eta^i(t)}{dt}$$

Notice that the vector field η *is* the infinitesimal variation δc. Furthermore, when taking variations of a function of both coordinates and velocities, we have

$$\delta f = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial \dot{x}}\frac{d}{dt}\left(\delta x\right), \tag{5.11}$$

$$\underline{\text{and not}} \qquad \delta f = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial \dot{x}}\delta \dot{x} \tag{5.12}$$

As we shall see, the appropriate way to extend this definition to manifolds will be to take infinitesimal variations of functions to be of the form (5.11) and *not* (5.12). This distinction is immaterial when the extended velocity field is choosen so that the operators d_t and δ commute; however it is crucial when these operators do not commute, which we will discuss in the context of Fiber Bundles and Suslov's Principle later on in this manuscript.

Variations of Functions on Manifolds

By direct extension of (5.11) we define the infinitesimal variation of a function on a manifold as follows:

Definition 15. *Let $c : [a, b] \to Q$ be a curve through the configuration manifold Q let $\vartheta : [-\varepsilon, \varepsilon] \times [a, b] \to Q$ be a variation of the curve c. Let δq^i be the components of the infinitesimal variation with respect to the coordinate basis induced by the coordinates $\{q^i\}$. We define the infinitesimal variation of the function $f : TQ \to \mathbb{R}$ as*

$$\delta f = \frac{\partial f}{\partial q^i}\delta q^i + \frac{\partial f}{\partial \dot{q}^i}\frac{d}{dt}\left(\delta q^i\right) \tag{5.13}$$

In addition to being a direct generalization of the classical definition (5.11), this definition is also *independent* of the extended velocity field $V(s, \tau)$. The infinitesimal variation δc is defined along c, and therefore $d_t \delta c$ is defined *independently* of the choice of an extended velocity field. To the contrary, had one defined an infinitesimal

variation of a function to be of the form (5.12), δV would appear in the variation, which does depend on the choice of extended velocity field V.

As mentioned previously, for the case of contemporaneous variations $\vartheta(s,t)$, the extended velocity field is given by:

$$V(s,t) = \frac{\partial \vartheta(s,t)}{\partial t}$$

Due to the smoothness of the variation, its infinitesimal variation is:

$$\delta V(t) = \left.\frac{\partial^2 \vartheta(s,t)}{\partial s \partial t}\right|_{s=0} = \left.\frac{\partial^2 \vartheta(s,t)}{\partial t \partial s}\right|_{s=0} = \frac{d}{dt}\left(\delta c(t)\right)$$

For contemporaneous variations, the above definition is equivalent to the following:

$$\delta f = \frac{\partial f}{\partial q^i}\delta q^i + \frac{\partial f}{\partial \dot{q}^i}\delta \dot{q}^i \qquad \text{for contemporaneous variations} \tag{5.14}$$

Also notice that by defining variations of functions as in Def. 15, integration by parts yields

$$\int_a^b \delta f(t)\ dt = \int_a^b \left[\frac{\partial f}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{q}^i}\right)\right] \delta q^i(t)\ dt,$$

independent of the commutativity of the operators d_t and δ, i.e. independent of the choice of extended velocity field. In this way, the Lagrange-D'Alembert equation always follows from a variational principle, namely:

$$\int \delta L\ dt = 0 \qquad \Rightarrow \qquad \left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)\right] \delta q^i = 0 \tag{5.15}$$

This holds for both holonomic and nonholonomic mechanics. However, in nonholonomic mechanics, the variations δq^i are *not* independent. Rather they must be chosen to satisfy the constraints. Neimark and Fufaev [106] used functional variations of the form (5.14) as their definition. They resolved the issue by saying the left side of (5.15) presupposes commutivity of the operators d_t and δ, and that the *fundamental*

variational principle should be taken as:

$$\int \left[\delta L + \frac{\partial L}{\partial \dot{q}^i} \left(\frac{d}{dt}(\delta q^i) - \delta \dot{q}^i \right) \right] \, dt = 0 \qquad \text{(Neimark and Fufaev)}$$

However, this is unnecessary. One must simply take (5.13) as the definition of δf over (5.14).

5.3.4 The Transitivity Choice

In this section, we will discuss the two primary ways of extending the velocity to the variation. Both ways lead to a different but equivalent set of equations of motion. If one has a curve $c = q^i(t)$, its variation ϑ, and an extended velocity field V, one cannot choose both

$$\frac{\partial \delta q^i}{\partial t} - \frac{\partial V^i}{\partial s} = 0 \qquad \text{and} \qquad \frac{\partial \zeta^j}{\partial t} - \frac{\partial u^j}{\partial s} = 0$$

One must choose either one or the other. The fact that one cannot choose both is a direct result of the transpositional relations (5.43), which we will discuss in §5.4.4. This is known as the transitivity choice; this terminology was introduced by Papastavridis [113]. However it was Neimark and Fufaev [106] who first hit upon the idea that the velocity is only defined along the curve of motion, and that the transpositional relations themselves are meaningless without defining the extended velocity field. The transitivity choice itself is actually just the definition of the extended velocity field.

The two transitivity choices we will be discussing in this dissertation are the following.

(T1) The *first transitivity choice* has been the choice of Hamel [60, 61], Hölder [65], and Voronets [131], and is the choice we will assume for the remainder of this paper, except for a brief aside later on about Suslov's Principle. The first transitivity choice is to use contemporaneous variations, which were discussed in

§5.3.2. In this case, the variable τ is identified with time, not only on the curve of motion, but throughout the variation. The extended velocity field consists of the tangent vectors of the varied curves, i.e. we have:

$$V(s,t) = \frac{\partial \vartheta(s,t)}{\partial t} \tag{5.16}$$

Due to the continuity of the variations, the operators ∂_s and ∂_t commute when acting on the variation, i.e. we have:

$$\frac{\partial \delta q^i}{\partial t} - \frac{\partial V^i}{\partial s} \equiv 0 \tag{5.17}$$

However, due to the transpositional relations, not all legs of this quadrilateral can be kinematically admissible. This bifurcates the theory into mechanics and vakonomics, and we will discuss extensively how this distinction emerges from this choice in §5.6.

(T2) The *second transitivity choice* is the choice of Suslov [128, 129], Levi-Civita and Amaldi [78], Rumiantsev [119], and Bloch [16]. This is the choice one invokes when employing Suslov's Principle or Ehresmann Connections. The second Transitivity Choice involves separating the velocity components into a set of independent $\dot{q}^I$ and dependent $\dot{q}^\sigma$ variables. In the language of fiber bundles, the independent variables are the variables of the base, whereas the dependent variables are the fiber variables. This is a natural division if the $m \times n$ constraint matrix a_i^σ is of the form $[a_i^\sigma] = [\mathbb{I}_{m\times m} | A_I^\sigma(q)]$. If the constraint matrix a_i^σ is not already in this form, one can replace the constraints by an equivalent set of constraints by replacing a_i^σ with its row-reduced echelon form. Either way, the nonholonomic constraints $a_i^\sigma \dot{q}^i = 0$ can be rewritten as:

$$a_i^\sigma \dot{q}^i = \dot{q}^\sigma + A_I^\sigma \dot{q}^I = 0$$

Then the first m velocity components $\dot{q}^\sigma = -A_I^\sigma \dot{q}^I$ are the dependent, fiber variables; and the remaining $(n-m)$ velocity components are the independent base variables.

The (T2) choice of an extended velocity field is one that is kinematically admissible *throughout* the variation. One sets the independent components of the variation to be contemporaneous, so that the picture Fig. 5.2 is still valid *when restricted to the independent directions*. The dependent components of the velocity field are then choosen as to satisfy the constraints. With these choices, the extended velocity field is given by:

$$V^I(s,\tau) = \frac{\partial \vartheta^I(s,\tau)}{\partial \tau} \tag{5.18}$$

$$V^\sigma(s,\tau) = -A_I^\sigma(q(s,\tau))V^I(s,\tau) \tag{5.19}$$

As a result, we have:

$$\frac{\partial \delta q^I}{\partial t} = \frac{\partial V^I}{\partial s}, \qquad \text{but} \qquad \frac{\partial \delta q^\sigma}{\partial t} \neq \frac{\partial V^\sigma}{\partial s} \tag{5.20}$$

Additionally, one is then free to take the following:

$$\frac{\partial \zeta^i}{\partial t} - \frac{\partial u^i}{\partial s} \equiv 0$$

We will use the second transitivity choice in our discussion of Suslov's Principle in §5.5; otherwise we will be focusing primarily on the contemporaneous variations of Hamel.

Notice that the transitivity choice defines what is meant by the term δV, which can formally be thought of as $\delta\dot{q}$ or $\delta d_t q$. In Hamel's formulation of nonholonomic mechanics, δV is understood to be the directional derivative of *the tangent vectors of the varied curves* in the direction of the infinitesimal variation $\delta q(t)$. In this formalism,

the velocity vectors $V(s,\tau)$ are only kinematically admissible along the actual curve at $s = 0$ (unless one is studying the vakonomic equations). In contrast, Suslov's understanding of the term δV is that it is a directional derivative of *kinematically admissible velocity vectors, whose base components are tangent to the varied curves* in the direction of the infinitesimal variation $\delta q(t)$.

5.4 The Transpositional Relations

The transpositional relations are an important feature in correctly understanding the dynamics of nonholonomic and vakonomic mechanics. The transpositional relations are a formula for the commutator $[d_t, \delta]q^i$. They were first used in the context of nonholonomic mechanics by Hamel [60, 61, 62] and Suslov [128]. The texts by Greenwood [58], Neimark and Fufaev [106], and Papastavridis [113] discuss the transpositional relations in depth for nonholonomic systems. We will arrive at a geometric form of these by first consider a connection that will later prove to have useful properties in our analysis.

5.4.1 A Certain Connection

In this section we will introduce a connection that is nonsymmetric (nonvanishing torsion) with zero curvature that will play an important role in understanding the transpositional relations and the geometry of the underlying variational principles of nonholonomic mechanics and vakonomic motion. We will refer to this connection and its associated Christoffel symbols with tildes, to accent that this is *not* the connection induced from the kinetic energy metric or the constraint distribution.

Definition 16. *Given the basis $\mathscr{E} = \{E_i\}_{i=1}^n$ defined on TQ, define a connection $\tilde{\nabla} : TQ \times TQ \to TQ$ by the property*

$$\tilde{\nabla}_{E_i} E_j = 0 \tag{5.21}$$

$\tilde{\nabla}$ *is the unique connection with respect to which the basis* $\mathscr{E}$ *is covariantly constant.*

Theorem 17. *With respect to the regular basis* $\left\{\frac{\partial}{\partial q^i}\right\}_{i=1}^{n}$, *the Christoffel Symbols* $\tilde{\Gamma}_{ij}^{k}$ *associated with the connection* $\tilde{\nabla}$ *are given by:*

$$\tilde{\Gamma}_{ij}^{k} = \Phi_{s}^{k}\frac{\partial \Psi_{j}^{s}}{\partial q^{i}} \tag{5.22}$$

Proof. The Christoffel symbols are defined by the relations

$$\tilde{\nabla}_{\frac{\partial}{\partial q^i}}\left(\frac{\partial}{\partial q^j}\right) = \tilde{\Gamma}_{ij}^{k}\frac{\partial}{\partial q^k}$$

By the relations (5.7) we have:

$$\begin{aligned}\tilde{\nabla}_{\frac{\partial}{\partial q^i}}\left(\frac{\partial}{\partial q^j}\right) &= \tilde{\nabla}_{\Psi_i^p E_p}\left(\Psi_j^q E_q\right)\\ &= \Psi_i^p\Psi_j^q\tilde{\nabla}_{E_p}\left(E_q\right) + \Psi_i^p E_p(\Psi_j^q)E_q\\ &= \Psi_i^p\Phi_p^s\frac{\partial\Psi_j^q}{\partial q^s}\Phi_q^k\frac{\partial}{\partial q^k} = \Phi_q^k\frac{\partial\Psi_j^q}{\partial q^i}\frac{\partial}{\partial q^k}\end{aligned}$$

□

In terms of regular coordinates, this connection defines the following covariant derivative of vector fields $X = X^i\frac{\partial}{\partial q^i}$ and $Y = Y^i\frac{\partial}{\partial q^i}$:

$$\tilde{\nabla}_X Y = \left\{X^i\frac{\partial Y^j}{\partial q^i} + \tilde{\Gamma}_{ik}^{j}X^iY^k\right\}\frac{\partial}{\partial q^j} \tag{5.23}$$

Theorem 18. *The torsion of this connection is given by:*

$$\tilde{T} = -\gamma_{pq}^{s}\ E_s\otimes\omega^p\otimes\omega^q \tag{5.24}$$

where the coefficients γ_{pq}^{s} *are the Hamel coefficients (see Greenwood [58])*

$$\gamma_{pq}^{s} = -\left\{\frac{\partial\Psi_j^s}{\partial q^i} - \frac{\partial\Psi_i^s}{\partial q^j}\right\}\Phi_p^i\Phi_q^j \tag{5.25}$$

The torsion of this connection may also be represented as:

$$\tilde{T} = d\omega^s E_s \tag{5.26}$$

where ω^s *are the one forms defined in* (5.3).

Proof. The torsion components with respect to the basis $\mathscr{E}$ are defined by:

$$T\left(E_p, E_q\right) = T^s_{pq} E_s$$

We have:

$$\begin{aligned} T(E_p, E_q) &= \tilde{\nabla}_{E_p}\left(E_q\right) - \tilde{\nabla}_{E_q}\left(E_p\right) - [E_p, E_q] \\ &= E_p E_q - E_q E_p \\ &= \Phi^j_q \frac{\partial}{\partial q^j}\left(\Phi^i_p \frac{\partial}{\partial q^i}\right) - \Phi^i_p \frac{\partial}{\partial q^i}\left(\Phi^j_q \frac{\partial}{\partial q^j}\right) \\ &= \left\{\Phi^j_q \frac{\partial \Phi^i_p}{\partial q^j}\Psi^s_i - \Phi^i_p \frac{\partial \Phi^j_q}{\partial q^i}\Psi^s_j\right\} E_s = -\gamma^s_{ij} E_s \end{aligned} \tag{5.27}$$

To show that $\tilde{T} = d\omega^s E_s$, consider the following straightforward calculation:

$$\begin{aligned} \tilde{T} &= \Phi^k_s \left\{\frac{\partial \Psi^s_j}{\partial q^i} - \frac{\partial \Psi^s_i}{\partial q^j}\right\} \frac{\partial}{\partial q^k} \otimes dq^i \otimes dq^j \\ &= \frac{\partial \Psi^s_j}{\partial q^i}\, E_s \otimes dq^i \wedge dq^j \\ &= d\omega^s E_s \end{aligned} \tag{5.28}$$

□

The vanishing of the torsion $\tilde{T}$ is equivalent to the n differential forms ω^i being exact. Since the basis $\mathscr{E}$ was constructed from a nonintegrable constraint distribution, the associated torsion tensor will be nonzero. It should be noted that the nonvanishing of the torsion tensor has to do with the nonintegrability of the basis $\mathscr{E}$. Even if there are no constraints, one can still define a nonintegrable set of quasi-velocities. This is commonly done in (unconstrained) rigid body mechanics. The rotational equations of motion (Euler's equations) are expressed in terms of a body fixed basis. The components of the angular velocity vector, when written with respect to this basis, are nonintegrable, and thus constitute a set of quasi-velocities.

In addition, the curvature of $\tilde{\nabla}$ is identically zero. It is interesting to note that the Ehresmann connection defined by the constraint distribution has the opposite properties [16, 29]. It is the vanishing of the curvature associated with the Ehresmann connection that is equivalent to the integrability of the constraint distribution, and the torsion is identicaly zero.

Theorem 19. *The curvature tensor defined by our connection,*

$$\tilde{R}(X,Y,Z) = \tilde{\nabla}_X \tilde{\nabla}_Y Z - \tilde{\nabla}_Y \tilde{\nabla}_X Z - \tilde{\nabla}_{[X,Y]} Z \tag{5.29}$$

is zero.

Proof. We will show the curvature components vanish in the basis $\mathscr{E}$. The curvature components are defined in the usual way:

$$\tilde{R}\left(E_a, E_b, E_c\right) = \tilde{R}^k_{abc} E_k$$

Let us use the notation

$$[E_i, E_j] = h^s_{ij} E_s$$

Due to the connection definition (5.21), the first two terms of (5.29) vanish, and we are left with

$$\tilde{R}\left(E_a, E_b, E_c\right) = h^s_{ab} \tilde{\nabla}_{E_s}\left(E_c\right) = 0$$

Thus with respect to the basis $\mathscr{E}$, the curvature components vanish. Since curvature is a tensor, they vanish with respect to any basis. □

5.4.2 The Rigid Body

As an example, we now compute the Christoffel symbols of this connection for the rigid body. We use the body axis components of the angular velocity as quasi-velocities.

Example: The Rigid Body Equations

Consider the free rigid body, with configuration manifold $SO(3)$. For generalized coordinates we choose the Type-I Euler angles (ψ, θ, ϕ). The body axis components of the angular velocity constitute a set of quasi-velocities of the system:

$$u_1 = \omega_x = -\dot{\psi}\sin\theta + \dot{\phi} \tag{5.30}$$

$$u_2 = \omega_y = \dot{\psi}\cos\theta\sin\phi + \dot{\theta}\cos\phi \tag{5.31}$$

$$u_3 = \omega_z = \dot{\psi}\cos\theta\cos\phi - \dot{\theta}\sin\phi \tag{5.32}$$

The transformation matrices are given by

$$\Psi = \begin{bmatrix} -\sin\theta & 0 & 1 \\ \cos\theta\sin\phi & \cos\phi & 0 \\ \cos\theta\cos\phi & -\sin\phi & 0 \end{bmatrix} \quad \text{and} \quad \Phi = \begin{bmatrix} 0 & \sec\theta\sin\phi & \sec\theta\cos\phi \\ 0 & \cos\phi & -\sin\phi \\ 1 & \tan\theta\sin\phi & \tan\theta\cos\phi \end{bmatrix}$$

The nonzero Christoffel symbols (5.22) are found to be:

$$\Gamma^1_{21} = -\tan\theta \qquad \Gamma^3_{21} = -\sec\theta \qquad \Gamma^2_{31} = \cos\theta \qquad \Gamma^1_{32} = -\sec\theta \qquad \Gamma^3_{32} = -\tan\theta$$

The geodesics corresponding to this connection are given by the geodesic equation $\ddot{q}^i + \Gamma^i_{jk}\dot{q}^j\dot{q}^k = 0$, and work out to be:

$$\ddot{\psi} = \tan\theta\dot{\psi}\dot{\theta} + \sec\theta\dot{\theta}\dot{\phi} \tag{5.33}$$

$$\ddot{\theta} = -\cos\theta\dot{\phi}\dot{\psi} \tag{5.34}$$

$$\ddot{\phi} = \sec\theta\dot{\psi}\dot{\theta} + \tan\theta\dot{\theta}\dot{\phi} \tag{5.35}$$

As can be readily shown u_1, u_2, and u_3, given by (5.30)-(5.32), are integrals of motion of the system (5.33)-(5.35). The geodesic equation therefore produces motion with constant body-axis components of angular velocity. This makes sense since the vectors E_i of the basis $\mathscr{E}$ are covariantly constant along the geodesics (5.21).

□

The above example motivates the following theorem.

Theorem 20. *The geodesics given by the connection $\tilde{\nabla}$ are equivalent to the resulting motion obtained by holding all of the quasi-velocities fixed.*

Proof. The geodesic equation can be written succinctly as

$$\tilde{\nabla}_{\dot{q}}\dot{q} = 0$$

Since $\dot{q} = u^i E_i$ it follows:

$$\tilde{\nabla}_{\dot{q}}\dot{q} = \tilde{\nabla}_{u^i E_i}(u^j E_j) = u^i u^j \tilde{\nabla}_{E^i}(E_j) + u^i E_i(u^j) E_j$$

The first term vanishes due to (5.21). The n geodesic equations are therefore simply

$$u^i E_i(u^j) = 0$$

However the left hand side is just $\dot{u}^j$. Therefore u^j will be constant under the resulting motion. □

5.4.3 Fiber Bundles

A common approach to analyzing nonholonomic constraint distributions is to use Fiber Bundles. See Bloch [16], Bloch, Krishnaprasad, Marsden, & Murray [20], or Favretti [48] for more details. For an n-dimensional configuration manifold Q for a system with m nonholonomic constraints, a Fiber Bundle on Q can be thought of as an assignment of $n - m$ independent coordinates called the base variables r^I and m dependent coordinates called the fiber variables s^α. Formally B is an $(n - m)$-dimensional submanifold of Q called the base space, $\pi : Q \to B$ is the projection operator, sets of the form $\pi^{-1}(b), b \in B$ are fibers, and $V_q = d_q\pi$ is the vertical space at q. Local coordinates can be taken as $q = (r^I, s^\alpha)$.

Definition 21. *An **Ehresmann connection** A is a vector-valued one-form on Q that satisfies:*

*(i) A is **vertical valued**: $A_q : T_qQ \to V_q$ is a linear map for each point $q \in Q$.*

*(ii) A is a **projection**: $A(v_q) = v_q$ for all $v_q \in V_q$.*

A can be represented locally as follows:

$$A = \omega^\alpha \frac{\partial}{\partial s^\alpha} \qquad \text{where} \qquad \omega^\alpha(q) = ds^\alpha + A_I^\alpha(r,s)dr^I$$

Let $v \in T_qQ$ be given in bundle coordinates as:

$$v = \dot{r}^I \frac{\partial}{\partial r^I} + \dot{s}^\alpha \frac{\partial}{\partial s^\alpha}$$

Then

$$A(v) = (\dot{s}^\alpha + A_I^\alpha \dot{r}^I) \frac{\partial}{\partial s^\alpha}$$

We further define the horizontal space by $H_q = \ker A$, so that $T_qQ = H_q \oplus V_q$. The Ehresmann connection is choosen so that H_q coincides with the constraint distribution. The one-forms ω^α can then taken to be an equivalent re-expression of the constraint one-forms.

For a given system with m linear nonholonomic constraints (5.1), an Ehresmann connection can be defined by taken q^α as fiber variables, q^I as base variables, and then by replacing the constraint matrix $a_i^\alpha(q)$ with its reduced row echelon form:

$$\text{rref}[a_i^\alpha] = \left[\begin{array}{c|c} \mathbb{I}_{m\times m} & [A_I^\alpha]_{m\times(n-m)} \end{array}\right]$$

This provides a natural partition of the velocity space where the $(n-m)$ velocity components $\dot{q}^I$ are taken to be independent and the m components $\dot{q}^\delta$ are taken to

be dependent via the relation:

$$\dot{q}^{\delta} = -A_I^{\delta}(q)\dot{q}^I \tag{5.36}$$

The transformation matrix Ψ, which defines the quasi-velocities, and its inverse Φ are then given by:

$$\Psi = \left[\begin{array}{c|c} \mathbb{I}_{m\times m} & [A_I^{\alpha}]_{m\times(n-m)} \\ \hline \mathbb{O}_{(n-m)\times m} & \mathbb{I}_{(n-m)\times(n-m)} \end{array}\right]$$

$$\Phi = \left[\begin{array}{c|c} \mathbb{I}_{m\times m} & [-A_I^{\alpha}]_{m\times(n-m)} \\ \hline \mathbb{O}_{(n-m)\times m} & \mathbb{I}_{(n-m)\times(n-m)} \end{array}\right]$$

The Christoffel symbols (5.22) associated with the connection $\tilde{\nabla}$ defined in Definition 16 are therefore given by:

$$\tilde{\Gamma}_{iJ}^{\alpha} = \frac{\partial A_J^{\alpha}}{\partial q^i}, \qquad \tilde{\Gamma}_{ij}^{A} = 0, \qquad \tilde{\Gamma}_{j\beta}^{i} = 0$$

$\tilde{\Gamma}$ has at most $nm(n-m)$ nonzero components out of its total n^3 components. From (5.28), the components of the torsion tensor with respect to the ordinary basis are:

$$\tilde{T} = \left(\tilde{\Gamma}^i_{jk} - \tilde{\Gamma}^i_{kj}\right) \frac{\partial}{\partial q^i} \otimes dq^k \otimes dq^j$$

$\tilde{T}$ therefore has the following nonzero components:

$$\tilde{T}^\alpha_{IJ} = \frac{\partial A^\alpha_J}{\partial q^I} - \frac{\partial A^\alpha_I}{\partial q^J} \qquad \tilde{T}^\alpha_{\sigma J} = -\tilde{T}^\alpha_{J\sigma} = \frac{\partial A^\alpha_J}{\partial q^\sigma}$$

Let $X, Y \in H_q$, so that X and Y satisfy the constraints. Then

$$\begin{aligned}
\tilde{T}(X,Y) &= \left[\left(\frac{\partial A^\alpha_J}{\partial q^I} - \frac{\partial A^\alpha_I}{\partial q^J}\right) X^I Y^J + \frac{\partial A^\alpha_J}{\partial q^\sigma} X^\sigma Y^J - \frac{\partial A^\alpha_J}{\partial q^\sigma} X^J Y^\sigma\right] \frac{\partial}{\partial q^\alpha} \\
&= \left[\left(\frac{\partial A^\alpha_J}{\partial q^I} - \frac{\partial A^\alpha_I}{\partial q^J}\right) X^I Y^J - A^\sigma_I \frac{\partial A^\alpha_J}{\partial q^\sigma} X^I Y^J + A^\sigma_J \frac{\partial A^\alpha_I}{\partial q^\sigma} X^I Y^J\right] \frac{\partial}{\partial q^\alpha} \\
&= -B^\alpha_{IJ} X^I Y^J \frac{\partial}{\partial q^\alpha}
\end{aligned}$$

where

$$B^\alpha_{IJ} = \frac{\partial A^\alpha_I}{\partial q^J} - \frac{\partial A^\alpha_J}{\partial q^I} - A^\sigma_J \frac{\partial A^\alpha_I}{\partial q^\sigma} + A^\sigma_I \frac{\partial A^\alpha_J}{\partial q^\sigma} \tag{5.37}$$

The coefficients B^α_{IJ} are the coefficients of the curvature of the Ehresmann connection, defined in Bloch [16]. The curvature of the Ehresmann connection is therefore related to how the torsion of the connection $\tilde{\nabla}$ acts on horizontal (kinematically admissible) vectors. It is sometimes beneficial *not* to choose a fiber bundle decomposition of the tangent bundle, defined by the reduced row echelon form of Ψ. For example consider the unconstrained rotational motion of the free rigid body. A fiber bundle approach would entail writing the equations of motion for the Euler angle rates, since $\mathrm{rref}(\Psi) = \mathbb{I}_3$. However, if one writes the equations of motion in terms of the quasi-velocities (body axis components of the angular velocity) one obtains Euler's equations, a much simpler set of equations than the corresponding dynamical equations for the Euler angle rates. In fact, the Euler equations are just the Boltzmann-Hamel equations

that arise if one uses the body-axis components of the angular velocity vector as a set of quasi-velocities (Greenwood [58]). Quasi-velocity techniques can therefore provide a simplification even for unconstrained systems.

5.4.4 The Transpositional Relations

We now state a well known result about the relation between Lie Derivatives and covariant derivatives:

Theorem 22. *Given a connection* $\tilde{\nabla}$ *and vector fields* $X,\ Y \in TQ$*, we have*

$$\mathcal{L}_X Y = \tilde{\nabla}_X Y - \tilde{\nabla}_Y X - \tilde{T}(X, Y) \tag{5.38}$$

In this section we will show that if one applies (5.38) using the connection defined in Definition 16, one recovers the transpositional relations.

As a direct consequence of the connection definition (5.21), we state without proof the following theorem:

Theorem 23. *Consider the connection* $\tilde{\nabla}$ *defined by* (5.22) *and* (5.23)*. Let* $X,\ Y \in TQ$ *and* $\mathscr{X} = \Psi \cdot X,\ \mathscr{Y} = \Psi \cdot Y$ *their representations in the basis* $\mathscr{E}$*. Then, with respect to the basis* $\mathscr{E}$*, our covariant derivative takes the form*

$$\tilde{\nabla}_X Y = \mathscr{X}^s E_s(\mathscr{Y}^t) E_t \tag{5.39}$$

Thus the connection ∇ has the following interpretation. When expressed in the basis $\mathscr{E}$, the covariant derivative $\tilde{\nabla}_X Y$ is the ordinary directional derivative of Y in the direction X.

Remark 1. *Because of* (5.7)*, this is equivalent to*

$$\tilde{\nabla}_X Y = X^s \frac{\partial \mathscr{Y}^t}{\partial q^s} E_t = X^s \frac{\partial (\Psi^t_\kappa Y^k)}{\partial q^s} E_t$$

Thus, the E_t *component of* $\nabla_X Y$ *is the directional derivative of* $\mathscr{Y}^t$ *in the* X *direction.*

Referring to our definitions of the quasi-velocity and quasi-variations (5.10), we immediately have:

Corollary 24. $\tilde{\nabla}_{\dot{q}}\delta q = \dfrac{\partial \zeta^j}{\partial t} E_j$

Corollary 25. $\tilde{\nabla}_{\delta q} V = \dfrac{\partial u^j}{\partial s} E_j$

where V is an extension of the velocity $\dot{q}$ to the variation, and the scalar functions u^j are the quasi-velocities of the extended velocity field, i.e. $V = u^j E_j$.

Setting $X = V$ and $Y = \delta q$ and expressing (5.38) with respect to the basis $\mathscr{E}$, we obtain the transpositional relations:

Theorem 26 (The Transpositional Relations). *Let $V : [-\varepsilon, \varepsilon] \times [a, b] \to TQ$ be an extension of the velocity $\dot{q}$, and u^j its quasi-velocities relative to the basis $\mathscr{E}$. The components of* (5.38) *with respect to the basis $\mathscr{E}$ yields can be represented in any of the following different ways:*

$$
\begin{aligned}
(\mathcal{L}_V \delta q)^i \, \Psi_i^j &= \left(\frac{\partial \zeta^j}{\partial t} - \frac{\partial u^j}{\partial s} \right) - \tilde{T}^j(V, \delta q) && (5.40)\\
&= \left(\frac{\partial \zeta^j}{\partial t} - \frac{\partial u^j}{\partial s} \right) - d\omega^j(V, \delta q) && (5.41)\\
&= \left(\frac{\partial \zeta^j}{\partial t} - \frac{\partial u^j}{\partial s} \right) - \left\{ \frac{\partial \Psi_b^j}{\partial q^a} - \frac{\partial \Psi_a^j}{\partial q^b} \right\} V^a \delta q^b && (5.42)\\
\left(\frac{\partial \delta q^i}{\partial t} - \frac{\partial V^i}{\partial s} \right) \Psi_i^j &= \left(\frac{\partial \zeta^j}{\partial t} - \frac{\partial u^j}{\partial s} \right) + \gamma_{pq}^j u^p \zeta^q && (5.43)
\end{aligned}
$$

where $\tilde{T}$ is the torsion of the connection defined in (5.21)*, which is related to the constraint one-forms by* (5.26)*.*

Notice that the left hand side of (5.43) is equivalent to the left hand side of (5.40) by the definition of the Lie derivative of a vector field:

$$
\mathcal{L}_V \delta q = V \delta q - \delta q V = \left(V^j \frac{\partial \delta q^i}{\partial q^j} - \delta q^j \frac{\partial V^i}{\partial q^j} \right) \frac{\partial}{\partial q^i} = \left(\frac{\partial \delta q^i}{\partial t} - \frac{\partial V^i}{\partial s} \right) \frac{\partial}{\partial q^i}
$$

The transpositional relations (5.43) can be viewed as a commutation relation of the operators d_t and δ. Loosely speaking it says that the velocity of the infinitesimal variation in coordinates, minus the infitesimal variation of the coordinate velocity, is equal to the velocity of the quasi-infinitesimal variation, minus the infinitesimal variation of the quasi-velocity, plus a commutator term. This commutator term has a specific geometric significance; it is the negative of the torsion of the connection defined in (5.21). To elucidate this some authors (Ehlers et al. [47], Greenwood [58], Neimark and Fufaev [106], for example) have used a set of nonexistent quantities π^i called *quasi-coordinates*. If we formally define $d\pi^i = u^i$, $\delta\pi^i = \zeta^i$, and $dq^i = V^i$, the transpositional relations can be written as:

$$\left(d\delta q^i - \delta dq^i\right)\Psi_i^j = \left(d\delta\pi^j - \delta d\pi^j\right) + \gamma_{pq}^j d\pi^p \delta\pi^q$$

Or, alternatively, as:

$$\left([d,\delta]q^i\right)\Psi_i^j = [d,\delta]\pi^j + \gamma_{pq}^j d\pi^p \delta\pi^q$$

Note that the Hamel coefficients γ_{pq}^j are related to the torsion $\tilde{T}$ by (5.24) and to the constraint one-forms by (5.26). We would like to point out that this notation is purely formal since the quantities π^j do not exist and the one forms $d\pi^j$ are *not* exact differential forms. This notation is however nice as it helps us to vivify the statement that the transpositional relations are a set of commutation relations for the operators d_t and δ.

Our final Corollary from Theorem 23 is

Corollary 27 (The Lie Derivative). *Suppose the vector fields $X,\ Y \in TQ$ are kinematically admissible, i.e. they satisfy the m constraint equations:*

$$\omega^\sigma(X) = \omega^\sigma(Y) = 0$$

Then

$$\mathcal{L}_X Y = 0 \qquad \textit{iff} \qquad d\omega^i(X,Y) = 0$$

Therefore, since the constraints are nonholonomic, we will generally have $\mathcal{L}_X Y \neq 0$.

Proof. From (5.38), we have

$$\mathcal{L}_X Y = \tilde{\nabla}_X Y - \tilde{\nabla}_Y X - \tilde{T}(X,Y)$$

Since $\mathscr{Y}^\sigma = \Psi_i^\sigma Y^i \equiv 0$ and $\mathscr{X}^\sigma = \Psi_i^\sigma X^i \equiv 0$ for $j = n-m+1, \ldots n$, i.e. X and Y satisfy the constraints, if $\tilde{T}^\sigma(X,Y) = d\omega^\sigma(X,Y) \neq 0$, at least m components of $\mathcal{L}_X Y$ will be nonzero. □

5.5 Suslov's Principle

In this section we will discuss Suslov's Principle. We will begin by carefully considering how one applies the nonholonomic constraints when using the second transitivity choice.

5.5.1 Application of the Constraints

In this section we will closely consider how one applies the constraints to a variation $q(s,\tau)$ of a kinematically admissible path $c(t)$, specifically for the (T2) or Suslov case. When using the second transitivity choice, the extended velocity field is defined by the relations (5.18) and (5.19). The variation $\mathcal{V} = \text{image}(q)$ is a two-dimensional submanifold of Q, and its tangent space $T\mathcal{V} \subset TQ$ is a two-dimensional subbundle of TQ. When choosing contemporaneous variations, the extended velocity vectors always lie tangent to the variation. In contrast, when one takes the second transitivity choice, the extended velocity vectors are choosen so that only the independent or base variables will lie tangent to the variation, while the dependent or fiber components are choosen as to satisfy the nonholonomic constraints. One can see this in Fig.

5.3. The varied velocity $V + \delta V$ itself is kinematically admissible. Even though the differential quadrilaterals are not closed in the full space, the variations are closed when projected onto the base space.

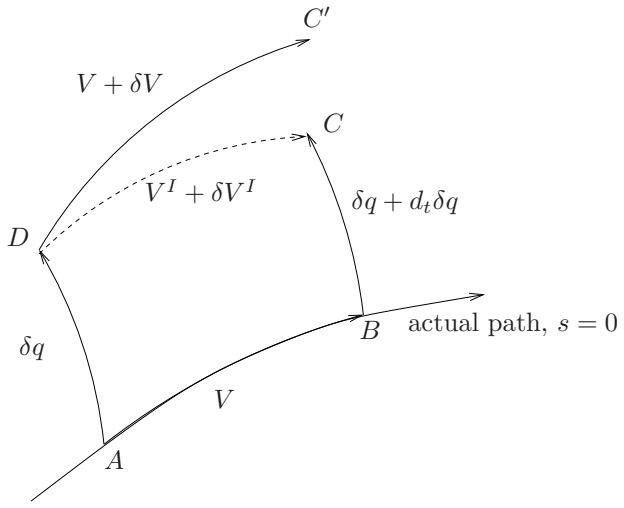

Figure 5.3: The Extended Velocity Field for the (T2) Case. All quantities are evaluated at A.

In determining the correct dynamical equations of motion, one must further apply the Principle of Virtual Work, which states that the infinitesimal variations $\delta q(t)$ must satisfy the nonholonomic constraints. The nonholonomic constraints must be applied to the infinitesimal variations at both A and B, and also to the extended velocity field at both A and D.

At the point A, the actual and virtual constraints are, respectively:

$$\dot{q}^\sigma = -A^\sigma_I \dot{q}^I \tag{5.44}$$

$$\delta q^\sigma = -A^\sigma_I \delta q^I \tag{5.45}$$

The second of these enforces the Principle of Virtual Work at A. The virtual constraint at B then becomes:

$$\delta q^\sigma + d\delta q^\sigma = -\left(A^\sigma_I + \frac{\partial A^\sigma_I}{\partial q^\kappa}\dot{q}^\kappa + \frac{\partial A^\sigma_I}{\partial q^J}\dot{q}^J \right) \left(\delta q^I + d\delta q^I \right)$$

where all of the quantities are evaluated at A. Since the actual and virtual constraints (5.44)-(5.45) apply at A, keeping terms to second order, the virtual constraints at B can be rewritten as:

$$d\delta q^\sigma = A^\kappa_J \frac{\partial A^\sigma_I}{\partial q^\kappa} \dot{q}^J \delta q^I - \frac{\partial A^\sigma_I}{\partial q^J} \dot{q}^J \delta q^I - A^\sigma_I d\delta q^I \tag{5.46}$$

This equation then enforces the virtual constraint at B. With its addition, the Principle of Virtual Work will hold everywhere along the solution curve γ.

To enforce the actual constraint at D, one then requires that

$$\dot{q}^\sigma + \delta V^\sigma = -\left(A^\sigma_I + \frac{\partial A^\sigma_I}{\partial q^\kappa} \delta q^\kappa + \frac{\partial A^\sigma_I}{\partial q^J} \delta q^J \right) (\dot{q}^I + \delta V^i)$$

Each of these quantities is evaluated at A, where the actual and virtual constraints (5.44)-(5.45) apply. The actual constraint at D therefore simplifies to, keeping only terms to second order:

$$\delta V^\sigma = A^\kappa_J \frac{\partial A^\sigma_I}{\partial q^\kappa} \dot{q}^I \delta q^J - \frac{\partial A^\sigma_I}{\partial q^J} \dot{q}^I \delta q^J - A^\sigma_I \delta V^I \tag{5.47}$$

Equation (5.46) enforces the virtual constraint at B (hence everywhere along $q(0,\tau) = c(\tau)$) whereas equation (5.47) enforces the actual constraint at C (hence $\dot{q}(s,\tau)$ satisfies the constraint). Enforcing both these constraints simultaneously yields:

$$\frac{d}{dt}(\delta q^\sigma) - \delta V^\sigma = B^\sigma_{IJ} \dot{q}^I \delta q^J + A^\sigma_I \left(\delta V^I - \frac{d}{dt}\left(\delta q^I\right) \right) \tag{5.48}$$

where B^σ_{IJ} is the curvature of the Ehresmann connection:

$$B^\sigma_{IJ} = \left(\frac{\partial A^\sigma_I}{\partial q^J} - \frac{\partial A^\sigma_J}{\partial q^I} + A^\kappa_I \frac{\partial A^\sigma_J}{\partial q^\kappa} - A^\kappa_J \frac{\partial A^\sigma_I}{\partial q^\kappa} \right) \dot{q}^I \delta q^J$$

Notice that (5.48) only holds in conjunction with the second transitivity choice (T2). If one chooses (T1) *and* the Principle of Virtual Work, one *cannot* enforce the actual constraint at D; thus both (5.47) and (5.48) would be invalid.

Due to the continuity of the variations and the definition of the independent components of the extended velocity field (5.18), we have that $\delta V^I - d\delta q^I = 0$. It therefore follows that:

$$\frac{d}{dt}\left(\delta q^\sigma\right) - \delta \dot{q}^\sigma = B^\sigma_{IJ}\dot{q}^I \delta q^J \tag{5.49}$$

5.5.2 Suslov's Principle

An alternative method for writing the nonholonomic mechanical equations of motion (5.56) can be obtained by using Suslov's Principle. For a mechanical system with a fiber bundle decomposition (choice of dependent and independent variables) and Lagrangian $L(q, \dot{q}^\sigma, \dot{q}^I) : TQ \to \mathbb{R}$, the *constrained Lagrangian* L^* is formed by writing

$$L^*(q, \dot{q}^I) = L(q, V^\sigma(q, \dot{q}^I), \dot{q}^I)$$

This is the constrained Lagrangian since the extended velocity field was choosen to satisfy the constraints. Suslov's Principle is a variational principle for the constrained Lagrangian.

We know from §5.3.3 that taking variations of the unconstrained Lagrangian yields:

$$\delta L = \frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\delta q^i = \frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial \dot{q}^\sigma}\frac{d}{dt}\delta q^\sigma + \frac{\partial L}{\partial \dot{q}^I}\frac{d}{dt}\delta q^I$$

In taking variations of L^* now, one must take variations of L treating the $\dot{q}^\sigma$'s not as independent variables but rather as generic functions $\dot{q}^\sigma = V^\sigma(q, \dot{q}^I)$. This yields:

$$\delta L^* = \frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial \dot{q}^\sigma}\delta V^\sigma + \frac{\partial L}{\partial \dot{q}^I}\frac{d}{dt}\dot{q}^I$$

Therefore

$$\delta L = \delta L^* + \frac{\partial L}{\partial \dot{q}^\sigma}\left[\frac{d}{dt}\left(\delta q^\sigma\right) - \delta V^\sigma\right]$$

From the Lagrange-D'Alembert Principle, $\int \delta L \; dt = 0$, one obtains:

$$\int \left\{\delta L^* + \frac{\partial L}{\partial \dot{q}^\sigma}\left[\frac{d}{dt}\left(\delta q^\sigma\right) - \delta V^\sigma\right]\right\} dt = 0 \tag{5.50}$$

This is *Suslov's Principle*. One then obtains the dynamical equations of motion by writing this in terms of the independent δq^I's, given by the relation (5.49).

Enforcing the Principle of Virtual Work (5.45), variations of the constrained Lagrangian are given by:

$$\delta L^* = -\frac{\partial L^*}{\partial q^\sigma} A_I^\sigma \delta q^I + \frac{\partial L^*}{\partial q^I}\delta q^I + \frac{\partial L^*}{\partial \dot{q}^I}\frac{d}{dt}\delta q^I$$

Using this, (5.48), (T2), integration by parts, and Suslov's Principle (5.50), we have

$$\int \left(-\frac{\partial L^*}{\partial q^\sigma} A_I^\sigma + \frac{\partial L^*}{\partial q^I} - \frac{d}{dt}\frac{\partial L^*}{\partial \dot{q}^I} + \frac{\partial L}{\partial \dot{q}^\sigma} B_{JI}^\sigma \dot{q}^J \right) \delta q^I \; dt = 0$$

Since the $(n-m)$ variations δq^I are independent, and due to the fact $B_{IJ}^\sigma = -B_{JI}^\sigma$, we have the following dynamical equations of motion:

$$\frac{d}{dt}\frac{\partial L^*}{\partial \dot{q}^I} - \frac{\partial L^*}{\partial q^I} + A_I^\sigma \frac{\partial L^*}{\partial q^\sigma} = -\frac{\partial L}{\partial \dot{q}^\sigma} B_{IJ}^\sigma \dot{q}^J$$

which agrees with (5.2.8) in Bloch [16] and (2.1.6) in Bloch, Krishnaprasad, Marsden, & Murray [20].

5.6 The Variational Principles

In this section, we will fully exploit the transpositional relations and clearly distinguish the variational principles behind nonholonomic and vakonomic systems. We will emphasize that both the dynamical and vakonomic equations of motion are deriveable in a similar fashion, and that the difference is not *when* you apply the constraints (they are applied at the same time in both cases), it is a matter of *how* you apply the constraints (i.e. how you enforce the constraints on your variations, with the Principle of Virtual Work or the Vakonomic Principle).

5.6.1 Nonholonomic versus Vakonomic Variational Principles

Corollary 27 demonstrates nonclosure of the configuration space due to the nonholonomic constraints at the differential level. If X and Y are both kinematically

admissible vector fields, one cannot trace out a differential quadrilateral using only X and Y. See Figure 5.4.

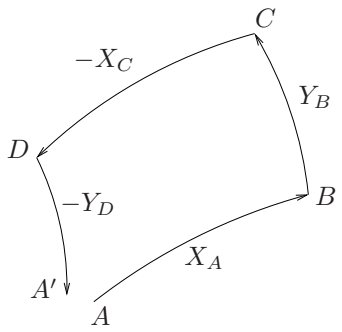

Figure 5.4: Nonclosure of q-space at the differential level due to nonholonomic constraints

Consider again Figure 5.2. By definition of our variation field, we require $\mathcal{L}_{\dot{q}}\delta q \equiv 0$, so that the variations are continuous. Due to (5.43), one *cannot* choose *both* auxillary conditions

$$\frac{\partial \zeta^\sigma}{\partial t} \equiv 0 \qquad \text{and} \qquad \frac{\partial u^\sigma}{\partial s} \equiv 0$$

This choice will result in different equations of motion. One must carefully choose how one applies the nonholonomic constraints on the variations. Incorrectly applying the constraints will lead to incorrect equations of motion.

The Principle of Virtual Work

The correct *dynamical* equations of motion satisfy the *Principle of Virtual Work*:

Definition 28 (The Principle of Virtual Work). *The constraint forces do no work on the particle under virtual displacements.*

In order to satisfy the Principle of Virtual Work the variations must satisfy:

$$\Psi_i^\sigma \delta q^i = \zeta^\sigma(0,t) \equiv 0$$

In this case the infinitesimal variations themselves are kinematically admissible, whereas the varied paths, due to the transposition relations, will not be. A direct

consequence of the Principle of Virtual Work is:

$$\frac{\partial \zeta^\sigma(0,t)}{\partial t} \equiv 0 \tag{5.51}$$

In this case, the first m transpositional relations reduce to

$$\frac{\partial u^\sigma}{\partial s} = -\left\{ \frac{\partial \Psi_j^\sigma}{\partial q^i} - \frac{\partial \Psi_i^\sigma}{\partial q^j} \right\} \dot{q}^i \delta q^j \tag{5.52}$$

Figure 5.5 depicts variations that satisfy the Principle of Virtual Work. Since the actual path of motion must be kinematically admissible, we have $u^\sigma = 0$ along the curve c. (5.52) states that the derivative of u^σ along the transversal direction must be nonzero, and hence the varied paths cannot be kinematically admissible. Meanwhile enforcing the Principle of Virtual Work through the condition $\zeta^\sigma = 0$ guarentees that the infinitesimal variations δq will be kinematically admissible.

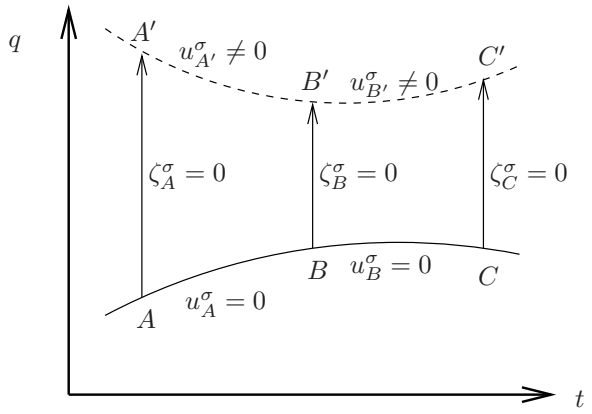

Figure 5.5: Principle of Virtual Work

The Vakonomic Principle

On the other hand, if we wish to enforce the condition that the varied paths be kinematically admissible, we require that $u^\sigma(s,t) \equiv 0$. The resulting equations

of motion will not represent the dynamical evolution of the system, but rather an evolution that is called the vakonomic dynamics.

Definition 29 (The Vakonomic Principle). *The constraint forces do no work on the particle as it traverses any of its varied paths.*

If one chooses this principle, one requires

$$u^\sigma(s,t) \equiv 0$$

If one selects the Vakonomic Principle, the varied paths are kinematically admissible, but the infinitesimal variations themselves are not. A direct consequence of this choice yields:

$$\frac{\partial u^\sigma(0,t)}{\partial s} \equiv 0 \tag{5.53}$$

In this case, the vakonomic case, the first m transpositional relations reduce to

$$\frac{\partial \zeta^\sigma}{\partial t} = \left\{ \frac{\partial \Psi^\sigma_j}{\partial q^i} - \frac{\partial \Psi^\sigma_i}{\partial q^j} \right\} \dot{q}^i \delta q^j \tag{5.54}$$

Figure 5.6 depicts variations when one selects the Vakonomic Principle. Since we are now taking the varied paths to be kinematically admissible, (5.54) implies that it is not possible for the infinitesimal variations themselves to satisfy the nonholonomic constraints. The variations δq therefore do not satisfy the Principle of Virtual Work, and the resulting equations of motion will not produce the correct dynamical equations of motion.

5.6.2 Hamilton's Principle

In geometric mechanics, one typically has a Lagrangian function $L : TQ \to \mathbb{R}$ defined on the tangent bundle of the configuration space. In this setting, in the absence of nonholonomic constraints, the correct dynamical equations of motion satisfy Hamilton's Principle:

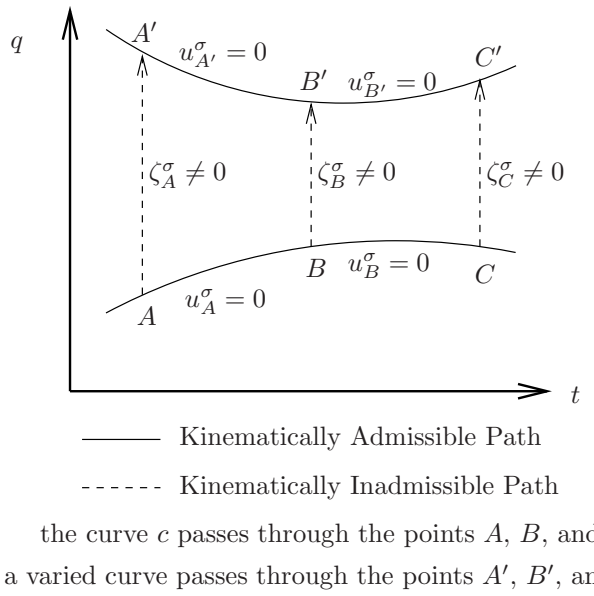

Figure 5.6: The Vakonomic Principle

Definition 30 (Hamilton's Principle). *A curve $c : [a, b] \to Q$, expressed in local coordinates as $c = q^i(t)$, linking the points $c(a), c(b) \in Q$, is said to satisfy* ***Hamilton's Principle****, if it is an extermal curve of the following functional:*

$$I[c] = \int_a^b L(q, \dot{q}, t)\ dt \tag{5.55}$$

where the domain is taken to be the set of all kinematically admissible curves connecting $q(a)$ and $q(b)$.

In the absence of nonholonomic constraints, the dynamical paths of the particle satisfy Hamilton's Principle. However, when nonholonomic constraints are added, Hamilton's Principle produces the *incorrect* dynamical equations of motion. In order to minimize the action amongst the set of all *kinematically admissible* variations, one must apply the Vakonomic Principle when taking the variations, i.e. it is the *varied curves* that are taken to be kinematically admissible, not the infinitesimal variations. Thus, in the nonholonomic setting, Hamilton's Principle produces what we call the vakonomic equations of motion. The correct *dynamical* equations of motion can be

derived from a similar variational principle, except it is the *infinitesimal variations* and not the *varied paths* that must be taken to be kinematically admissible, i.e. the Principle of Virtual Work must be chosen over the Vakonomic Principle if the correct *dynamical* equations of motion are to be obtained. We say that the dynamical trajectories instead satisfy the Lagrange-D'Alembert Principle:

Definition 31 (Lagrange-D'Alembert Principle). *A curve is said to satisfy the Lagrange-D'Alembert Principle if it minimizes the action*

$$I[c] = \int_a^b L(q, \dot{q}, t)\ dt$$

where the variations are chosen to satisfy the Principle of Virtual Work.

In the absence of nonholonomic constraints, the particle simultaneously satisfies *both* Hamilton's Principle and the Lagrange-D'Alembert Principle. Due to the transpositional relations, this cannot be the case when nonholonomic constraints are present.

5.6.3 The Dynamical Nonholonomic Equations of Motion

By applying the Principle of Virtual Work to the variations, we will show that the correct dynamical equations of motion are given by the following Theorem:

Theorem 32 (Nonholonomic Dynamics). *The correct dynamical equations of motion are given by the fundamental nonholonomic form of Lagrange's equations:*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = \lambda_\sigma a_i^\sigma \tag{5.56}$$

$$a_k^\sigma \dot{q}^k = 0$$

Where we call the constraint matrix: $a_i^\sigma = \Psi_i^\sigma$.

Proof. To obtain the correct dynamical equations of motion, we first take variations of the action (5.55) and then integrate by parts, recalling the variations vanish at the endpoints.

$$\begin{aligned} \delta I &= \int_a^b \left(\frac{\partial L}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{q}^i} \frac{d}{dt} \left(\delta q^i \right) \right) \, dt \\ &= \int_a^b \left\{ \frac{\partial L}{\partial q^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) \right\} \delta q^i \, dt \end{aligned}$$

Here the variations δq^i cannot be choosen arbitrarily, but rather must be choosen so as to satisfy the Principle of Virtual Work. If one has, in addition, nonconservative, generalized applied forces Q_i, one must satisfy the equation

$$\left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} + Q_i \right) \delta q^i = 0$$

where the constraints δq^i satisfiy the Principle of Virtual Work. This is Lagrange's Principle. The constraint forces do not enter into this equation.

We will employ the use of the following shorthand, the so-called "variational derivative:"

$$\frac{\delta L}{\delta q^i} = \frac{\partial L}{\partial q^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right)$$

We now wish to enforce the Principle of Virtual Work by applying the constraints (5.51). We can add m Lagrange Multipliers $\mu_1, \ldots, \mu_m$, to the integrand as follows:

$$\begin{aligned} \delta I &= \int_a^b \left(\frac{\delta L}{\delta q^i} \delta q^i + \mu_\sigma \frac{\partial \zeta^\sigma}{\partial t} \right) dt \\ &= \int_a^b \left(\frac{\delta L}{\delta q^i} \delta q^i + \mu_\sigma \frac{\partial (\Psi_i^\sigma \delta q^i)}{\partial t} \right) dt \\ &= \int_a^b \left\{ \frac{\delta L}{\delta q^i} - \dot{\mu}_\sigma \Psi_i^\sigma \right\} \delta q^i \, dt \end{aligned}$$

We see that the variations of $I[\gamma]$ vanish exactly when:

$$\left\{ \frac{\delta L}{\delta q^i} - \dot{\mu}_\sigma \Psi_i^\sigma \right\} \delta q^i = 0 \tag{5.57}$$

We can now choose the multipliers μ_σ so that the infinitesimal variations δq^i are independent. Setting their respective coefficients to zero, while identifying $\lambda_\sigma = -\dot{\mu}_\sigma$ yields the result. □

5.6.4 The Vakonomic Equations of Motion

To obtain the Vakonomic Equations of Motion, one varies the action while applying the Vakonomic Principle to the variations. The equations of motion are given here.

Theorem 33 (Vakonomic Dynamics). *The correct vakonomic equations of motion are given by:*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = \mu_\sigma \left(\frac{\partial a_k^\sigma}{\partial q^i} - \frac{\partial a_i^\sigma}{\partial q^k}\right)\dot{q}^k - \dot{\mu}_\sigma a_i^\sigma \tag{5.58}$$

$$a_k^\sigma \dot{q}^k = 0$$

Proof. Varying the action (5.55), we obtain the usual

$$\delta I = \int_a^b \frac{\delta L}{\delta q^i}\delta q^i \; dt$$

We can now apply Lagrange Multipliers to enforce the Vakonomic Principle (5.53), to obtain:

$$\delta I = \int_a^b \left(\frac{\delta L}{\delta q^i}\delta q^i + \mu_\sigma \frac{\partial u^\sigma}{\partial s}\right) \; dt$$

Since our variations are continuous, i.e. $\mathcal{L}_{\dot{q}}\delta q \equiv 0$, the Transpositional Relations (Theorem 26) yield:

$$\frac{\partial u^\sigma}{\partial s} = \frac{\partial \zeta^\sigma}{\partial t} - \left(\frac{\partial \Psi_i^\sigma}{\partial q^k} - \frac{\partial \Psi_k^\sigma}{\partial q^i}\right)\dot{q}^k \delta q^i \tag{5.59}$$

Applying this, and integrating the first term by parts (as in the proof of Theorem 32), we obtain:

$$\delta I = \int_a^b \left\{\frac{\delta L}{\delta q^i} - \dot{\mu}_\sigma \Psi_i^\sigma - \mu_\sigma \left(\frac{\partial \Psi_i^\sigma}{\partial q^k} - \frac{\partial \Psi_k^\sigma}{\partial q^i}\right)\dot{q}^k\right\}\delta q^i \; dt$$

The variations of $I[\gamma]$ vanish exactly when:

$$\left\{ \frac{\delta L}{\delta q^i} - \dot{\mu}_\sigma \Psi_i^\sigma - \mu_\sigma \left(\frac{\partial \Psi_i^\sigma}{\partial q^k} - \frac{\partial \Psi_k^\sigma}{\partial q^i} \right) \dot{q}^k \right\} \delta q^i = 0 \tag{5.60}$$

We can choose the multipliers μ_σ so that the infinitesimal variations δq^i are independent. Setting their coefficients to zero yields the result. □

5.7 The Gyroscopic Principle of Vakonomic Mechanics

The Gyroscopic Principle, which we propose here, will allow us to extend Maggi's equations and the Boltzmann-Hamel equations to Vakonomic Mechanics in the following sections.

5.7.1 Dynamical Systems with External Forces

For a mechanical system with nonholonomic constraints and external forces $F_i dq^i$, the dynamical equations of motion can be derived in a similar fashion as in Theorem 32. Including the external forces, (5.57) is replaced with:

$$\left\{ \frac{\delta L}{\delta q^i} - \dot{\mu}_\sigma \Psi_i^\sigma + F_i \right\} \delta q^i = 0 \tag{5.61}$$

where the μ_σ's can be choosen so that the infinitesimal variations δq^i are independent. The correct dynamical equations of motion are therefore given by:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) - \frac{\partial L}{\partial q^i} = -\dot{\mu}_\sigma a_i^\sigma + F_i$$

$$a_i^\sigma \dot{q}^i = 0$$

Setting $\lambda_\sigma = -\dot{\mu}_\sigma$, we recognize the term

$$C_i = \lambda_\sigma a_i^\sigma \tag{5.62}$$

as the constraint force acting on the system.

5.7.2 The Gyroscopic Principle

The following discussion is motivated by a comparison of (5.60) and (5.61), keeping in mind that the variations for both cases are now arbitrary.

For a mechanical system, define a set of m *Gyroscopic Controls* by

$$Q_i^\sigma = -\left(\frac{\partial \Psi_i^\sigma}{\partial q^k} - \frac{\partial \Psi_k^\sigma}{\partial q^i}\right)\dot{q}^k$$

Define the *Gyroscopic Control Coefficients* from the constraint force coefficients (5.62) by:

$$\mu_\sigma(t) = M_\sigma - \int_a^t \lambda_\sigma(\tilde{t})\, d\tilde{t}$$

where the initial conditions M_σ can be chosen arbitrarily.

By direct comparison between (5.60) and (5.61), one sees that one can obtain the vakonomic solutions by applying the following control forces to the nonholonomic dynamical system:

$$F_i = \mu_\sigma Q_i^\sigma \tag{5.63}$$

Since these control forces are gyroscopic, they do no work on the system. We thus have the following.

Definition 34 (The Gyroscopic Principle). *A nonholonomic dynamical system with external applied gyroscopic forces* (5.63) *produces the same set of differential equations of motion as the corresponding vakonomic system (with the same Lagrangian).*

This has the following interpretations. A nonholonomically constrained dynamical system can be forced to satisfy Hamilton's Principle with the appropriate gyroscopic control forces. This observation is useful in optimal control. Also, a vakonomic system can be thought of as a forced dynamical system. Under this interpretation,

the appropriate constraints are the ones given by the Princple of Virtual Work, and not the Vakonomic Principle. We can thus trade the Vakonomic Principle for the Principle of Virtual Work *and* the appropriate external gyroscopic control forces, and still obtain the same set of differential equations. This observation is crucial in extending quasi-velocity techniques, such as Maggi's equation and Boltzmann-Hamel equation, to vakonomic mechanics.

5.8 Maggi's Equations

5.8.1 Nonholonomic Mechanics with External Forces

For a dynamical system with nonholonomic constraints and external forces, the dynamics are given by (5.61), where the δq^i's are independent. An alternate approach to the Lagrange Multiplier method is Maggi's Equation. We start with Lagrange's Principle:

$$\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} - F_i\right] \delta q^i = 0 \tag{5.64}$$

which comes from taking variations of the action $I[\gamma]$. This is identical to (5.61) before the addition of the Lagrange Multipliers. In this equation, the δq^i's are taken to satisfy the Principle of Virtual Work, and are therefore linearly dependent.

By the Gyroscopic Principle, the vakonomic equations are a special case of this, where the forces are chosen by the relations (5.63).

In order to enforce the Principle of Virtual Work, one can transform (5.64) into the basis $\mathscr{E}$ as follows:

$$\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} - F_i\right] \Phi^i_j \zeta^j = 0 \tag{5.65}$$

Enforcing the Principle of Virtual Work, $\zeta^\sigma = 0$, we can take the remaining $n-m$ ζ^A's to be independent, thus we obtain a set of $n-m$ second order differential

equations known as Maggi's Equations:

$$\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} - F_i\right] \Phi^i_A = 0 \tag{5.66}$$

Coupled with the m constraint equations (5.1), this gives a minimal set, equivalent to $2n - m$ first order differential equations of motion. In the vakonomic case, one requires an *additional* m differential equations to solve for the Lagrange Multipliers, as the gyroscopic control force (5.63) depends on the history of the constraint force coefficients.

5.8.2 A Justification of the Lagrange Multiplier Method

We can separate the summation (5.65) into two parts, thereby expressing Lagrange's Principle as follows:

$$\left[\frac{\delta L}{\delta q^i} + F_i\right] \Phi^i_\sigma \zeta^\sigma + \left[\frac{\delta L}{\delta q^i} + F_i\right] \Phi^i_A \zeta^A = 0$$

Now, the Principle of Virtual Work mandates that $\zeta^\sigma = 0$, so we can add an arbitrary multiplier to these coefficients without affecting the total sum. One obtains:

$$\left\{\left[\frac{\delta L}{\delta q^i} + F_i\right] \Phi^i_\sigma + \lambda_\sigma\right\} \zeta^\sigma + \left[\frac{\delta L}{\delta q^i} + F_i\right] \Phi^i_A \zeta^A = 0$$

Choosing each λ_σ so that the *coefficient* to ζ^σ is also zero, we can now let ζ^σ be arbitrary, without affecting the total sum (5.65). Converting back to the regular basis and making the usual identification $\lambda_\sigma = -\dot{\mu}_\sigma$, we recover the fundamental nonholonomic form of Lagrange's Equations (5.61), where the δq^i's are independent.

5.8.3 Maggi's Equations for Vakonomic Mechanics

According to the Gyroscopic Principle, we can apply the forces (5.63) to the dynamical system to obtain the correct equations of motion. In a similar fashion as in the preceeding subsection, we can transform (5.61) into the basis $\mathscr{E}$ as follows:

$$\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i}\right] \Phi^i_R = -\mu_\sigma \left[\frac{\partial \Psi^\sigma_i}{\partial q^k} - \frac{\partial \Psi^\sigma_k}{\partial q^i}\right] \dot{q}^k \Phi^i_R \tag{5.67}$$

$$\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i}\right]\Phi^i_\alpha = -\dot{\mu}_\alpha - \mu_\sigma\left[\frac{\partial \Psi^\sigma_i}{\partial q^k} - \frac{\partial \Psi^\sigma_k}{\partial q^i}\right]\dot{q}^k\Phi^i_\alpha \tag{5.68}$$

One notices that with the choice (5.63), that Maggi's equations (5.66) become exactly (5.67). However, the gyroscopic forces depend upon the coefficients μ_σ, hence additional equations are necessary to solve for their time evolution. These additional equations are exactly (5.68). By transforming to the basis $\mathscr{E}$, we separate the $\dot{\mu}_\sigma$ terms so that they are confined to only m of the equations. The remaining $n-m$ second order differential equations (5.66), coupled with the constraint equations, provide the $2n-m$ first order equations of motion for the dynamics, coupled with the m equations (5.67), from which one solves for the constraint coefficients. A much more efficient way of producing such a separation is given in the next section.

5.9 The Boltzmann-Hamel Equations

The Boltzmann-Hamel equation the counterpart of Lagrange's equation when the Lagrangian is expressed in terms of the quasi-velocities.

5.9.1 The Boltzmann-Hamel Equations for unconstrained systems

We will first consider an *unconstrained* system with external forces. The generalization to the dynamic nonholonomic and the vakonomic cases will be straightforward. The Boltzmann-Hamel equations are a form of the Euler-Lagrange equations which apply when the Lagrangian is written in terms of the quasi-velocities. Greenwood [58] §4.3 takes a more algebraic approach. We will derive the Boltzmann-Hamel equations using variational principles. Define

$$\mathscr{L}(q,u) = L(q,\dot{q})$$

where $\mathscr{L}$ is simply L expressed in terms of the quasi-velocities. Taking variations of (5.55) with $\mathscr{L}(q,u)$ substituted for $L(q,\dot{q})$, we obtain:

$$\delta I = \int_a^b \left(\frac{\partial \mathscr{L}}{\partial q^i}\delta q^i + \frac{\partial \mathscr{L}}{\partial u^i}\delta u^i + F_i \delta q^i \right) dt \tag{5.69}$$

Here F_i is the external force. Recall that

$$\frac{\partial \mathscr{L}}{\partial q^i}\delta q^i = \frac{\partial \mathscr{L}}{\partial q^i}\Psi^i_r \zeta^r$$

Taking continuous variations, so that $\mathcal{L}_{\dot{q}}\delta q \equiv 0$, the Transpositional Relations of Theorem 26 can be written:

$$\delta u^i = \frac{\partial u^i}{\partial s} = \frac{\partial \zeta^i}{\partial t} + \gamma^i_{sr} u^s \zeta^r$$

where γ^i_{sr} are the Hamel coefficients of (5.25). Integrating by parts, (5.69) becomes

$$\delta I = \int_a^b \left\{ \frac{\partial \mathscr{L}}{\partial q^i}\Phi^i_r - \frac{d}{dt}\left(\frac{\partial \mathscr{L}}{\partial u^r}\right) + \frac{\partial \mathscr{L}}{\partial u^i}\gamma^i_{sr}u^s + F_i\Phi^i_r \right\} \zeta^r \, dt$$

Since we have written δI for the unconstrained system, the set of variations ζ^r are independent. We thus obtain the Boltzmann-Hamel equations:

$$\frac{d}{dt}\left(\frac{\partial \mathscr{L}}{\partial u^r}\right) - \frac{\partial \mathscr{L}}{\partial q^i}\Phi^i_r - \frac{\partial \mathscr{L}}{\partial u^j}\gamma^j_{sr}u^s = F_i\Phi^i_r \tag{5.70}$$

5.9.2 The Boltzmann-Hamel Equations for constrained dynamical systems

We've seen that the nonholonomic dynamical systems can be viewed as an unconstrained system with the appropriate constraint forces (5.62). Setting $F_i = C_i + R_i$, where R_i is the applied force, (5.70) becomes:

$$\frac{d}{dt}\left(\frac{\partial \mathscr{L}^*}{\partial u^A}\right) - \frac{\partial \mathscr{L}^*}{\partial q^i}\Phi^i_A - \frac{\partial \mathscr{L}}{\partial u^j}\gamma^j_{SA}u^S = R_i\Phi^i_A \tag{5.71}$$

$$\frac{d}{dt}\left(\frac{\partial \mathscr{L}}{\partial u^\sigma}\right) - \frac{\partial \mathscr{L}^*}{\partial q^i}\Phi^i_\sigma - \frac{\partial \mathscr{L}}{\partial u^j}\gamma^j_{S\sigma}u^S = R_i\Phi^i_\sigma - \dot{\mu}_\sigma \tag{5.72}$$

where $\mathscr{L}^*$ can be taken to be either the constrained *or* the unconstrained Lagrangian. For these terms, setting $u^\sigma = 0$ can occur before or after taking the indicated derivatives. This is an important observation for later.

The $\lambda_\sigma = -\dot{\mu}_\sigma$ are whatever they need to be so that (5.72) holds. Hence, these m equations can be ignored. Thus, the Boltzmann Hamel equations for forced nonholonomic dynamical systems are simply the $n - m$ first order equations (5.71). Constraining $u^\sigma = 0$, (5.5) provides an additional set of n kinematic differential equations:

$$\dot{q}^i = \Phi^i_A u^A \tag{5.73}$$

Together, (5.71) and (5.73) provide $2n - m$ differential equations, from which one can solve for $q^i(t)$ and $u^A(t)$. These $2n - m$ equations are known as the *Boltzmann Hamel equations*. The m equations contained in (5.72) determine only the dynamics of the multipliers; however this approach isolates the influence of the multipliers, so that they affect only these same m equations. They can therefore be ignored.

We would like to again stress that the Lagrangian must be written in terms of the unconstrained variables, except where noted. It is the summation $\frac{\partial \mathscr{L}}{\partial w^j}\gamma^j_{SA}$ over all $j = 1, \ldots, n$ in (5.71) that contains terms that would otherwise be missing if one began with the constrained Lagrangian.

5.9.3 The Boltzmann-Hamel Equations for Vakonomic Systems

According to the Gyroscopic Principle, we can obtain the dynamical equations of motion with the selection of the forces (5.63). The m equations (5.72) are now important, because one must solve for the constraint coefficients μ_σ, and these equations give exactly that.

Moreover, we recognize

$$F_i \Phi^i_j = \mu_\sigma \gamma^\sigma_{sj} u^s \tag{5.74}$$

for the Gyroscopic forces F_i.

Let us make the ansatz that the multipliers take the form:

$$\mu_\sigma = -\frac{\partial \mathscr{L}}{\partial u^\sigma} + \nu_\sigma \tag{5.75}$$

where $\nu_\sigma(t)$ are unknown functions of time. The vakonomic equations of motion, derived from (5.71) and (5.72), now simplify to:

$$\frac{d}{dt}\left(\frac{\partial \mathscr{L}^*}{\partial u^A}\right) - \frac{\partial \mathscr{L}^*}{\partial q^i}\Phi^i_A - \frac{\partial \mathscr{L}^*}{\partial u^J}\gamma^J_{BA} u^B = \nu_\sigma \gamma^\sigma_{BA} u^B \tag{5.76}$$

$$\frac{\partial \mathscr{L}^*}{\partial q^i}\Phi^i_\sigma + \frac{\partial \mathscr{L}^*}{\partial u^J}\gamma^J_{B\sigma} u^B = -\nu_\tau \gamma^\tau_{B\sigma} u^B + \dot{\nu}_\sigma \tag{5.77}$$

In many cases, $\gamma^J_{ij} = 0$ since this is the torsion in the unconstrained dimensions. This results in a further simplification in the vakonomic Boltzmann-Hamel equations of motion. Moreover, in any case, in the vakonomic form of the Boltzmann-Hamel equations, we notice that one can use the constrained Lagrangian $\mathscr{L}^*(q^i, u^A) = \mathscr{L}(q^i, u^\sigma = 0, u^A)$, and still obtain the correct vakonomic equations of motion. This is due to the fact that the derivatives

$$\frac{\partial \mathscr{L}}{\partial u^\sigma}$$

do not appear in the vakonomic Boltzmann-Hamel equations. Recall that these terms *do* appear in the dynamic Boltzmann-Hamel equations (5.71), forcing the added complexity of the necessity of writing the Lagrangian for the unconstrained system. (Of course, if the components $\gamma^A_{BC} = 0$, one can also write the dynamical Boltzmann-Hamel equations from the constrained Lagrangian).

The Vakonomic Equations of Motion, expressed in Boltzmann-Hamel form, are given by the $n-m$ first order differential equations (5.76) for the u^A's, the m first order

differential equations (5.77) for the ν_σ's, and the n first order differential equations (5.73) for the q's. The ν_σ's are related to the standard Lagrange Multipliers via (5.75) (in terms of the *unconstrained* Lagrangian). This gives a minimal set of $2n$ first order differential equations that produce the vakonomic motion.

5.9.4 Equivalence of Dynamics

An interesting question is whether the dynamical equations of motion can be thought of as a special case of the vakonomic ones, with appropriate choice of the initial values of the Lagrange Multipliers. Examples have been found where the dynamical motion is a special case of the vakonomic motion, and where it is not. Much research on this subject has been done in recent years, see for example Cortes et al [34], Ehlers et al [47], Favretti [48], and Fernandez and Bloch [49]. We propose a new test for equivalence that is based on the vakonomic Boltzmann-Hamel equations presented earlier in this paper.

The vakonomic equations of motion are obtained if one substitutes (5.74) for $R_i\Phi^i_j$ into the dynamic forced Boltzmann-Hamel equations (5.71) and (5.72). The dynamical equations of motion are found by the substitution $R_i = 0$. Carefully studying these equations, one sees that the dynamical motion is a special case of the vakonomic, if the following two conditions hold:

1. The m Lagrange Multipliers can be choosen such that the $n-m$ equations hold:

$$\mu_\sigma \gamma^\sigma_{BA} u^B \equiv 0 \tag{5.78}$$

2. The restriction placed on the μ_σ's by (5.78) does not interfere with their ability to satisfy:

$$\frac{d}{dt}\left(\frac{\partial \mathscr{L}}{\partial u^\sigma}\right) - \frac{\partial \mathscr{L}^*}{\partial q^i}\Phi^i_\sigma - \frac{\partial \mathscr{L}}{\partial u^j}\gamma^j_{S\sigma}u^S = \mu_\tau \gamma^\tau_{B\sigma}u^B - \dot{\mu}_\sigma \tag{5.79}$$

The condition (5.78) states that the components of the gyroscopic control force in the unconstrained quasi-directions vanish identically. In this case, (5.71) reduces to its nonholonomic dynamical form with $F_i = 0$. This, alone, is not enough. One must further check that the m equations (5.72) can still be satisfied. (5.78) places a linear dependency relation on the multipliers, so, in general, one will have less than m multiplier degrees of freedom to satisfy the m equations (5.72)/(5.79) with, but it can be done.

5.9.5 Application to Rigid Body Dynamics

In this section we will use the Boltzmann-Hamel equations to derive the dynamical equations of motion for a free rigid body with moment of inertia tensor $\mathbb{I}$. We shall take as generalized coordinates the Type-I Euler angles (ψ, θ, ϕ). As quasi-velocities, choose the body-fixed components of the angular momentum:

$$\begin{aligned} u_1 &= \omega_x = -\dot{\psi}\sin\theta + \dot{\phi} \\ u_2 &= \omega_y = \dot{\psi}\cos\theta\sin\phi + \dot{\theta}\cos\phi \\ u_3 &= \omega_z = \dot{\psi}\cos\theta\cos\phi - \dot{\theta}\sin\phi \end{aligned}$$

We will further assume that the body-fixed frame is a principal axis frame, so that $\mathbb{I} = \mathrm{diag}(I_{xx}, I_{yy}, I_{zz})$. The transformation matrices are given by:

$$\Psi = \begin{bmatrix} -\sin\theta & 0 & 1 \\ \cos\theta\sin\phi & \cos\phi & 0 \\ \cos\theta\cos\phi & -\sin\phi & 0 \end{bmatrix}$$

$$\Phi = \begin{bmatrix} 0 & \sec\theta\sin\phi & \sec\theta\cos\phi \\ 0 & \cos\phi & -\sin\phi \\ 1 & \tan\theta\sin\phi & \tan\theta\cos\phi \end{bmatrix}$$

The Lagrangian, when expressed in terms of the quasi-velocities, is given by:

$$\mathscr{L}(q,u) = \frac{1}{2}(I_{xx}u_1^2 + I_{yy}u_2^2 + I_{zz}u_3^2)$$

The nonzero Hamel coefficients are:

$$\gamma_{23}^1 = 1 \qquad \gamma_{13}^2 = -1 \qquad \gamma_{12}^3 = 1$$

$$\gamma_{32}^1 = -1 \qquad \gamma_{31}^2 = 1 \qquad \gamma_{21}^3 = -1$$

It is interesting to note that if you view the Hamel coefficients as the components of a linear mapping $\gamma : TQ \times TQ \to TQ$, then, for this special example, the above Hamel coefficients yield $\gamma(X,Y) = X \times Y$, for any $X, Y \in \mathbb{R}^3$. The Boltzmann-Hamel equations, (5.71) and (5.73), produce the following set of dynamical equations of motion for this system, which we immediately recognize as the classical Euler equations:

$$I_{xx}\dot{u}_1 + (I_{zz} - I_{yy})u_2u_3 = M_x \tag{5.80}$$

$$I_{yy}\dot{u}_2 + (I_{xx} - I_{zz})u_1u_3 = M_y \tag{5.81}$$

$$I_{zz}\dot{u}_3 + (I_{yy} - I_{xx})u_1u_2 = M_z \tag{5.82}$$

where M_x, M_y, and M_z are the components of the net applied control torque with respect to the body fixed principal axis frame. We will return to rigid body dynamics again in Chapter VII, where we will discuss the dynamics from a Lie group perspective and further show how Euler's equations can also be derived from the Euler-Poincaré equations. The dynamical optimal control of the free rigid body is discussed in §6.5.4.

5.10 Conclusion

In this chapter we discussed the variational principles of nonholonomic and vakonomic motion, and how one may use quasi-velocity techniques in obtaining a reduced

set of differential equations of motion. We introduced a new connection related to one's choice of quasi-velocities for the system. With respect to this connection the quasi-velocities are covariantly constant. We used this connection to derive the transpositional relations. We then discussed the transitivity choice that one makes when defining variations, and the implication of this choice on carrying out the variational principles. We showed how the (T1) case introduced by Hamel then leads to an additional choice, so that one may obtain either the correct dynamical equations of motion or the vakonomic equations of motion. We presented a parallel discussion for the transitivity choice (T2), and discussed its relation to Suslov's principle and fiber bundles. We also explored mechanics on Lie groups and demonstrated that the Euler-Poincaré equations are simply a special case of the Boltzmann-Hamel equations that are suitable in handling dynamics on Lie groups.

CHAPTER VI

The Boltzmann-Hamel Equations for Optimal Control

6.1 Introduction

In this chapter we extend the quasi-velocity techniques discussed in Chapter V to the optimal control setting. We introduce a fourth order version of the Boltzmann-Hamel equations, which yields a reduced set of equations for the kinematic and dynamic optimal control problems for mechanical systems with nonholonomic constraints. In particular, we will show the dynamic optimal control problem can be written as a minimal set of $4n - 2m$ first order differential equations of motion.

6.1.1 Background

In this paper we extend the classical Boltzmann-Hamel equations to kinematic and dynamic optimal control problems. In the analysis of nonholonomic systems, a number of eloquent formalisms have emerged based on a set of quantities known as *quasi-velocities*. For an n degree of freedom system with m nonholonomic constraints, one defines the first m quasi-velocities $u^\sigma, \sigma = 1, \ldots, m$ in a way such that they span the constraint distribution. In this way the constraints reduce to the relations $u^\sigma = 0$, and one only need solve for the remaining $n - m$ independent quasi-velocities. In addition the n kinematic relations which define the constraints can be integrated to produce the curve of motion. One requires in total $2n - m$ differential equations

of motion, as opposed to the $2n + m$ equations necessary using the fundamental nonholonomic form of Lagrange's equations, Greenwood [56, 58].

Given a mechanical system with nonholonomic constraints, one may reexpres the Lagrangian in terms of the generalized coordinates and the quasi-velocities. This new Lagrangian will not, in general, satisfy the Euler-Lagrange equations. Instead it satisfies a similar set of equations known as the Boltzmann-Hamel equations.

The Boltzmann-Hamel equations have also been used for the analysis of unconstrained systems with symmetry. For instance, in rigid body mechanics the body-axis components of the angular velocity constitute a set of quasi-velocities. The Boltzmann-Hamel equations for this set of quasi-velocities produce Euler's equations for rigid body dynamics, see Greenwood [58]. As a generalization of Euler's equations, the Euler-Poincaré equations are a set of dynamical equations of motion for mechanical systems with symmetry, i.e. systems whose underlying configuration manifold is a Lie group (Bloch et al. [21] and Marsden [83]). As we will discuss in Chapter VII, the Euler-Poincaré equations are a special case of the Boltzmann-Hamel equations, under the conditions that 1) the underlying configuration manifold is a Lie group, 2) the quasi-velocities are taken to be the pullback of the velocity vector to the Lie algebra by the left translation map, and 3) the Lagrangian is left-invariant. The Euler-Poincaré equations have been generalized to a set of Euler-Poincaré optimal control equations, see for instance Cortes et al. [35], Koon and Marsden [71], and Martinez et al. [87]. In this way, the Boltzmann-Hamel optimal control equations presented here can be thought of as a generalization of the Euler-Poincaré optimal control equations.

Under certain conditions, the optimal control problem can be reformulated as a vakonomic problem (Bloch and Crouch [18]). This makes sense, as optimal control

problems are variational in nature. One can further analyze optimal control problems with Pontryagin's Maximum Principle, see Betts [13], Bloch [16], Bullo and Lewis [27], or Agrachev and Sachkov [3]. Nonholonomic control systems on manifolds is also discussed in Bloch and Crouch [19], and the geometry of underactuated control systems is discussed in Bloch [16] and Montgomery [101].

In the optimal control problem one has a cost function that one seeks to minimize, usually taken to be an integral over the optimal path. In kinematic or dynamic optimal control, the path itself is subject to a certain set of kinematic or dynamic equations of motion, respectively. If one uses Pontryagin's Maximum Principle, each of these equations is enforced by a corresponding Lagrange Multiplier. The formalism we present here has the additional feature that it circumvents the necessity of this set of Lagrange Multipliers. m Lagrange Multipliers are still required to enforce the nonholonomic constraints, an inescapable feature of the vakonomic problem. In the kinematic optimal control problem, the set of n differential equations for the Lagrange multipliers used to enforce the kinematic equations of motion are replaced with $n - m$ differential equations for the quasi-velocities. In the dynamic optimal control problem, the set of differential equations for the $2n$ Lagrange multipliers used to enforce the dynamical equations of motion are replaced with $2n - 2m$ differential equations for the quasi-accelerations and quasi-jerks; the n differential equations for the velocities are further replaced by $n - m$ differential equations for the quasi-velocities.

6.1.2 Chapter Overview

In §6.2 we will add to the notions of nonholonomic constraints and quasi-velocities that were presented in §5.2 and §5.3 by further introducing quasi-accelerations and quasi-jerks. We will further present a higher order set of transpositional relations,

called the second transpositional relations, that compliment the original set derived in §5.4.4. The first and second transpositional relations play a crucial role in correctly applying variational methods to nonholonomically constrained systems. In §6.3 we will review the classical Boltzmann-Hamel equations already presented in §5.9. In §6.4 we will generalize the Boltzmann-Hamel equations for use with kinematic optimal control problems, and write the corresponding solutions as a system of $2n$ first order differential equations of motion. We will show how our technique applies to the kinematically controlled Heisenberg system, vertical rolling disc, and falling rolling disc. In §6.5 we will derive a form of the Boltzmann-Hamel equations applicable to dynamically controlled systems, obtaining a set of $4n - 2m$ first order differential equations. We will apply our technique to the dynamically controlled Chaplygin Sleigh, vertical rolling disc, free rigid body. We will use our techniques to reproduce the result of Noakes et al. [110], for the optimal control equations of a free rigid sphere with control torques about a body fixed axis. We will show these equations are a special case for the more general set of equations which yield the optimal reorientation of a free rigid body. Finally we produce a higher order version of the Euler-Poincaré equations, discussed in §7.3, that is applicable to optimal control problems on Lie groups.

6.1.3 Summation Convention

To aid in notation, we will invoke the same summation convention as used in the previous chapter. The indices that will be in use throughout this chapter are the following. The Greek letters σ and τ are run over the constrained dimensions $1, \ldots, m$. The capital letters I, J, K, and S run over the unconstrained dimensions $m+1, \ldots, n$. Finally the lower case letters a, b, i, j, k, and s run over all dimensions $1, \ldots, n$.

letter type	indices	summation over
Greek	σ, τ	$1, \ldots, m$
capital	I, J, K, S	$m+1, \ldots, n$
lower case	a, b, i, j, k, s	$1, \ldots, n$

Table 6.1: Summation Convention for Chapter VI

6.2 The Second Transpositional Relations

In this chapter we will continue our discussion of a n degree of freedom system with m nonholonomic constraints. We will follow the definitions and notation for constraints, quasi-velocities, and variations as they were set in §5.2 and §5.3. In §5.4.4 we first introduced the transpositional relations, which, intuitively, represent a set of commutator relations for the operators $\partial/\partial t$ and $\partial/\partial s$ as they act on the variations. In this section we will derive a higher order set of transpositional relations, called the second transpositional relations, which will be used in generalizing the Boltzmann-Hamel equations to the optimal control setting. Recall that the first transpositional relations are given by Cor. 26 as follows: are

$$\left(\frac{\partial \delta q^i}{\partial t} - \delta \dot{q}^i\right) \Psi_i^j = \left(\frac{\partial \zeta^j}{\partial t} - \delta u^j\right) + \gamma_{ab}^j u^a \zeta^b \tag{6.1}$$

Assuming the first transitivity choice (T1), equation (5.17), discussed in §5.3.4, the transpositional relations reduce to the form:

$$\delta u^j = \dot{\zeta}^j + \gamma_{ab}^j u^a \zeta^b \tag{6.2}$$

This form assumes full closure of the differential quadrilaterals in q-space. Setting either $\dot{\zeta}^\sigma = 0$ or $\delta u^\sigma = 0$ then leads to the Principle of Virtual Work and the correct dynamical equations of motion or the Vakonomic Principle and the correct Vakonomic "equations of motion," respectively. This was discussed in §5.6 in detail.

As optimal control problems are fourth order in nature, we will make use of the

notion of quasi-accelerations and quasi-jerks. The quasi-acceleration is defined as the components of the covariant acceleration with respect to the quasi-basis:

Definition 35. *The j-th quasi-acceleration, a^j, is the projection of the covariant acceleration on the j-th quasi-basis vector:*

$$a^j = \omega^j \left(\tilde{\nabla}_{\dot{q}} \dot{q} \right)$$

where the ω^j are the differential forms given by (5.3) *and the connection $\tilde{\nabla}$ is the one associated with the quasi-basis $\{E_i\}$, given in Def. 16.*

We note that this definition is equivalent to the expression:

$$\tilde{\nabla}_{\dot{q}} \dot{q} = a^j E_j \tag{6.3}$$

Proposition 1. *The n-th quasi-accelerations is the time derivative of the n-th quasi-velocity, i.e.*

$$a^j = \dot{u}^j$$

Proof. A direct calculation using (5.5), (5.7), and Theorem 23 yields:

$$\tilde{\nabla}_{\dot{q}} \dot{q} = u^s E_s(u^j) E_j = \dot{q}^i \Psi_i^s \Phi_s^k \frac{\partial u^j}{\partial q^k} E_j = \dot{q}^k \frac{\partial u^j}{\partial q^k} E_j = \dot{u}^j E_j$$

□

Similarly we define:

Definition 36. *The j-th quasi-jerk, $\jmath^j$, is the projection of the covariant jerk onto the j-th quasi-basis direction:*

$$\jmath^j = \omega^j \left(\tilde{\nabla}_{\dot{q}} \left(\tilde{\nabla}_{\dot{q}} \dot{q} \right) \right)$$

where the ω^j are the differential forms given by (5.3) *and the connection $\tilde{\nabla}$ is the one associated with the quasi-basis $\{E_i\}$, given in Def. 16.*

Equivalently:

$$\tilde{\nabla}_{\dot{q}}\left(\tilde{\nabla}_{\dot{q}}\dot{q}\right) = \jmath^j E_j$$

In direct analogy with the result above, we also have:

Proposition 2. *The n-th quasi-jerk is the time derivative of the n-th quasi-acceleration, i.e.*

$$\jmath^j = \dot{a}^j$$

Proof. By definition

$$\omega^j\left(\tilde{\nabla}_{\dot{q}}\dot{q}\right) = a^j$$

Hence, by applying (5.5), (5.7), and Theorem 23, one obtains:

$$\tilde{\nabla}_{\dot{q}}\left(\tilde{\nabla}_{\dot{q}}\dot{q}\right) = u^s E_s(a^j)E_j = \dot{q}^i \Psi_i^s \Phi_s^k \frac{\partial a^j}{\partial q^k} E_j = \dot{q}^k \frac{\partial a^j}{\partial q^k} E_j = \dot{a}^j E_j$$

□

The first transpositional relations show how we can replace the variation of the quasi-velocity with the time derivative of the quasi-infinitesimal variation ζ^i plus a commutator term (6.2). The second transpositional relations state that you can replace the variation of the quasi-acceleration with the time derivative of the variation of the quasi-velocity. Hence the operators $\partial/\partial t$ and $\partial/\partial s$ commute when applied to quasi-velocities. We state this result in the following theorem:

Theorem 37 (Second Transpositional Relation). *For continuous variations, we have:*

$$\delta\frac{\partial u^i}{\partial t} = \frac{\partial \delta u^i}{\partial t} \tag{6.4}$$

or, equivalently, $\delta a^i = \partial(\delta u^i)/\partial t$. *(Recall the infinitesimal variation operator* δ *is equivalent to* $\partial/\partial s$*).*

Proof. By Theorem 19, the curvature $\tilde{R}$ associated with $\tilde{\nabla}$ vanishes. By using (6.3) and the results of Corollaries 24 and 25, while recalling $[\dot{q}, \delta q] = 0$, we have:

$$\begin{aligned}
\tilde{R}(\dot{q}, \delta q, \dot{q}) &= \tilde{\nabla}_{\dot{q}} \tilde{\nabla}_{\delta q} \dot{q} - \tilde{\nabla}_{\delta q} \tilde{\nabla}_{\dot{q}} \dot{q} - \tilde{\nabla}_{[\dot{q}, \delta q]} \dot{q} \\
&= \tilde{\nabla}_{\dot{q}} \left(\frac{\partial u^j}{\partial s} E_j \right) - \tilde{\nabla}_{\delta q} \left(a^j E_j \right) \\
&= \left(\frac{\partial^2 u^j}{\partial t \partial s} - \frac{\partial^2 u^j}{\partial s \partial t} \right) E_j
\end{aligned}$$

Since the curvature of this connection vanishes identically, we obtain our result. □

6.3 The Dynamical Boltzmann-Hamel Equations

We discussed the Boltzmann-Hamel Equations in §5.9 for unconstrained and constrained nonholonomic and vakonomic systems. We restate the result here for convenience. Recall $\mathscr{L}(q, u) = L(q, \dot{q}(q, u))$ is the re-expression of the *unconstrained* Lagrangian in terms of the quasi-velocities. Taking variations of the action and using the first transpositional relations (6.2), one obtains:

$$\begin{aligned}
\delta I &= \int_a^b \left(\frac{\partial \mathscr{L}}{\partial q^i} \delta q^i + \frac{\partial \mathscr{L}}{\partial u^i} \delta u^i + F_i \delta q^i \right) \, dt \\
&= \int_a^b \left(\frac{\partial \mathscr{L}}{\partial q^s} \Phi_i^s - \frac{d}{dt} \frac{\partial \mathscr{L}}{\partial u^i} + \frac{\partial \mathscr{L}}{\partial u^j} \gamma_{ki}^j u^k + Q_i \right) \zeta^i \, dt
\end{aligned}$$

where F_i is the external applied force and we have defined $Q_i = \Phi_i^j F_j$. After applying the Principle of Virtual Work, $\zeta^\sigma \equiv 0$, the remaining $n-m$ variations ζ^I can be taken to be independent, and we obtain the Boltzmann-Hamel equations for nonholonomic mechanics:

$$\frac{d}{dt} \frac{\partial \mathscr{L}}{\partial u^I} - \frac{\partial \mathscr{L}}{\partial q^s} \Phi_I^s - \frac{\partial \mathscr{L}}{\partial u^j} \gamma_{KI}^j u^K = Q_I \tag{6.5}$$

$$\dot{q}^i = \Phi_J^i u^J \tag{6.6}$$

One must use the *unconstrained* Lagrangian for these equations. After the partial derivatives are taken, one then applies the constraints $u^\sigma = 0$. The Boltzmann-Hamel

equations (6.5)- (6.6) are a *minimal* set of $2n-m$ first order differential equations for the n q^i's and the $n-m$ u^I's. We will come back to this set of equations during our discussion of optimal dynamical control problems in §6.5.

6.4 Kinematic Optimal Control

In this section we present a quasi-velocity based method for kinematic optimal control problems, where one has direct controls over the velocities. To illustrate the theory we will consider three examples: kinematic control of the Heisenberg system, the vertical rolling disc, and the falling rolling disc.

6.4.1 Theory

Given a mechanical system with a set of nonholonomic constraints, define the following distribution on the tangent bundle:

$$\mathcal{K} = \{(q,v) \in TQ : \omega^\sigma(v) = 0,\ \sigma = 1,\dots,m\}$$

where the differential forms ω^σ are the constraint one forms. The distribution $\mathcal{K}$ is simply the tangent subbundle of kinematically admissible velocity vectors. We now define constrained affine kinematic control systems (e.g. Bullo and Lewis [27] or Nijmeijer and van der Schaft [107]).

Definition 38. *A **constrained affine kinematic control system** is a triple $(Q, \{\omega^\sigma\}, \{X_I\})$, where Q is a configuration manifold, $\{\omega^\sigma\}$ is a set of m nonholonomic constraint one forms $\omega^\sigma \in T^*Q$ $(\sigma = 1,\dots,m)$ and $\{X_I\}$ is a set of $(n-m)$ independent and kinematically admissible vector fields $X_I \in \mathcal{K}$ $(I = m+1,\dots,n)$ called the **control vector fields**.*

Given a constrained affine kinematic control system, the *controls* are a set of $(n-m)$ scalar functions $w^I : [t_1,t_2] \to \mathbb{R}$ $(I = m+1,\dots,n)$. Given a set of controls

and an initial condition $q_0 \in Q$, the *resultant trajectory* is a curve $c : [t_1, t_2] \to Q$ that satisfies the differential equations:

$$\dot{q}^i(t) = X_I^i(q(t))w^I(t), \qquad q^i(0) = q_0^i \tag{6.7}$$

where X_I^i is the i-th component of the I-th control vector field.

Definition 39. *Given a constrained affine kinematic control system $(Q, \{\omega^\sigma\}, \{X_I\})$, two fixed endpoints $q_1, q_2 \in Q$, and a cost functional*

$$I[c] = \int_{t_1}^{t_2} g(c(t), w(t))\ dt, \tag{6.8}$$

where $g : Q \times \mathbb{R}^{n-m} \to \mathbb{R}$, then the associated ***kinematic optimal control problem*** *is to determine the control functions $w^I : [t_1, t_2] \to \mathbb{R}$ such that the resultant trajectory c, i.e. the solution of* (6.7)*, and the controls w, minimize the functional* (6.8)*.*

We will now solve this problem using an appropriately defined set of quasi-velocities. To begin we note that, without loss of generality, one can invert the relation (6.7) in the following way. Let $\mathbf{X} : Q \to \mathbb{R}^{n\times(n-m)}$ be an $n \times (n-m)$ matrix valued function on the configuration manifold Q defined such that its (i, I)-th entry is $X_I^i(q)$, i.e. the columns of $\mathbf{X}$ are the control vector fields X_I. We can define now the pseudoinverse of $\mathbf{X}$ (see Campbell and Meyer [28]) as the matrix valued function $\mathbf{b} : Q \to \mathbb{R}^{(n-m)\times n}$ by the relation:

$$\mathbf{b}(q) = \left(\mathbf{X}^T(q) \cdot \mathbf{X}(q)\right)^{-1} \cdot \mathbf{X}^T(q) \in \mathbb{R}^{(n-m)\times n} \tag{6.9}$$

where $\mathbf{X}^T$ is the transpose of the matrix $\mathbf{X}$. This is well-defined since the rank of $\mathbf{X}$ is $(n-m)$. Notice that $\mathbf{b}(q) \cdot \mathbf{X}(q)$ is the $(n-m) \times (n-m)$ identity matrix, for all $q \in Q$. Now let $b_i^I(q)$ be the (I, i)-th component of $\mathbf{b}(q)$ at q. Premultiplying (6.7) by $\mathbf{b}$ we obtain:

$$b_i^J(q(t))\dot{q}^i(t) = b_i^J(q(t))X_I^i(q(t))w^I(t) = \delta_I^J w^I(t) = w^J(t)$$

This relation motivates the following definition. We define a set of $(n-m)$ functions $\varpi^I : TQ \to \mathbb{R}$ by the relation

$$\varpi^I(q, v) = b_i^I(q)v^i$$

The significance of these functions is as follows. Suppose the curve $c = q^i(t)$ is a solution of the equation (6.7). Then the controls which generate that curve are given by:

$$w^I(t) = \varpi^I(q(t), \dot{q}(t)) = b_i^I(q(t))\dot{q}^i$$

The ϖ^I, when evaluated along a kinematically admissible curve, are literally the controls which generate that curve. They also, coincidentally, constitute a set of $n-m$ quasi-velocities that are linearly independent from the quasi-velocities which arise due to the constraints. We take as quasi-velocities the following set:

$$u^\sigma = a_i^\sigma(q)\dot{q}^i = 0 \tag{6.10}$$

$$u^I = b_i^I(q)\dot{q}^i = \varpi^I(q, \dot{q}) \tag{6.11}$$

The quasi-velocities u^σ are literally the constraints, and the quasi-velocities given by u^I are literally the controls. It follows that the transformation matrix Ψ is defined by the relations $\Psi_i^\sigma(q) = a_i^\sigma(q)$ and $\Psi_i^I(q) = b_i^I(q)$, i.e. the first m rows of Ψ form the constraint coefficient matrix and the remaining $n-m$ rows of Ψ form the pseudo-inverse of the matrix of control vector fields.

We may now rewrite the integrand of the cost function (6.8) in terms of the quasi-velocities in the following way:

$$C(q, u^I) = g(q, \varpi^I)$$

Notice that $C(q, u^I)$ only depends on the $(n-m)$ unconstrained quasi-velocities $u^I = \varpi^I$.

We now wish to minimize the curve c which minimizes the cost function

$$I[c] = \int_{t_1}^{t_2} C(q,u)\ dt$$

In order to enforce the nonholonomic constraints (6.10), we must choose the vakonomic principle, i.e. we wish to enforce the condition that $\delta u^\sigma = 0$ along the curve. Taking variations and adding a set of Lagrange Multipliers to enforce the vakonomic condition, we have:

$$\delta I = \int_a^b \left(\frac{\partial C}{\partial q^s}\Phi_i^s \zeta^i + \frac{\partial C}{\partial u^I}\delta u^I + \mu_\sigma \delta u^\sigma \right)\ dt \tag{6.12}$$

We may now use the transpositional relations (6.2) and integrate the resulting equation by parts, yielding:

$$\delta I = \int_a^b \left[\left(\frac{\partial C}{\partial q^s}\Phi_i^s + \frac{\partial C}{\partial u^I}\gamma_{si}^I u^s + \mu_\sigma \gamma_{si}^\sigma u^s \right)\zeta^i - \frac{d}{dt}\frac{\partial C}{\partial u^I}\zeta^I - \dot{\mu}_\sigma \zeta^\sigma \right]\ dt$$

The variations ζ^i are now independent, and therefore the optimal trajectory is found by setting their coefficients to zero. We have therefore proved:

Theorem 40. *Let $(Q, \{\omega^\sigma\}, \{X_I\})$ and the cost functional $I[c]$ defined in* (6.8) *be a kinematic optimal control problem. Let b_i^I be the coefficients of the matrix valued function defined in* (6.9)*. Let a_i^σ be the coefficients of the constraint matrix, so that $\omega^\sigma = a_i^\sigma dq^i$ and let $\Psi : Q \to \mathbb{R}^{n\times n}$ be defined by $\Psi_i^\sigma = a_i^\sigma$ and $\Psi_i^I = b_i^I$. Let $\Phi(q) = \Psi^{-1}(q)$, $u^i = \Psi_j^i \dot{q}^j$, $C(q,u) = g(q^i, \Psi_j^I \dot{q}^j)$, and let γ_{jk}^i be the Hamel coefficients. Then the* ***Boltzmann-Hamel equations for the kinematic optimal control problem*** *are:*

$$\frac{d}{dt}\frac{\partial C}{\partial u^I} - \frac{\partial C}{\partial q^j}\Phi_I^j - \frac{\partial C}{\partial u^J}\gamma_{KI}^J u^K = \mu_\tau \gamma_{KI}^\tau u^K \tag{6.13}$$

$$-\frac{\partial C}{\partial q^j}\Phi_\sigma^j - \frac{\partial C}{\partial u^J}\gamma_{K\sigma}^J u^K = -\dot{\mu}_\sigma + \mu_\tau \gamma_{K\sigma}^\tau u^K \tag{6.14}$$

$$\dot{q}^i = \Phi_K^i u^K \tag{6.15}$$

The solution to these equations is the resultant trajector of the optimal control problem, and the controls that produce the optimal trajectory are given by

$$w^I(t) = \varpi^I(q(t), \dot{q}(t)) = \Psi_i^I(q(t))\dot{q}^i(t)$$

These equations represent a minimal set of $2n$ first order differential equations: the $n-m$ equations (6.13) for the unconstrained u^I's, the m equations (6.14) for the multipliers μ_σ's, and n kinematic relations (6.15) for the q^i's.

We would like to point out that by applying the Lagrange Multipliers *before* taking the variations, we were implicitly selecting the Vakonomic Principle over the Principle of Virtual Work. Recall that the Vakonomic Principle can be enforced by the constraints $\delta u^\sigma = 0$, so that the *varied paths* satisfy the constraints, whereas the infinitesimal variations δq are *kinematically inadmissible*. Notice that this constraint $\delta u^\sigma = 0$ is precisely the third term of the integrand in (6.12). The fourth term of this integrand, $u^\sigma \delta \mu_\sigma$, is just a restatement of the system's constraints. Thus the the Boltzmann-Hamel equations for the kinematic optimal control problem are derived in direct analogy to the derivation of the Vakonomic equations of motion presented in §5.6.4, except that the mechanical Lagrangian is replaced with the integrand of the Cost function. Additionally, if this cost function integrand $C(q, u)$, when expressed in terms of the quasi-velocities, is identical to the *constrained* mechanical Lagrangian, then these equations produce the vakonomic motion associated with the system. See Bloch and Crouch [18] for additional discussion on the coincidence of the vakonomic motion (Lagrange's Problem) and the optimal control problem.

6.4.2 Optimal Control of the Heisenberg System

The optimal control of the Heisenberg system, discussed in Bloch [16, 24], is a classical underactuated kinematic control problem. Local coordinates are given by

$q = \langle x, y, z \rangle$ and motion is subject to the nonholonomic constraint:

$$\dot{z} = y\dot{x} - x\dot{y} \tag{6.16}$$

The control velocity field is given by:

$$\dot{q} = X_1 w^1 + X_2 w^2 \tag{6.17}$$

where $X_1 = \langle 1, 0, y \rangle^T$ and $X_2 = \langle 0, 1, -x \rangle^T$. Also, $w_1(t)$ and $w_2(t)$ are the unknown control functions. Using these controls, one seeks to steer the particle from the point $\langle 0, 0, 0 \rangle$ at time $t = 0$ to the point $\langle 0, 0, a \rangle$ at time $T > 0$, while minimizing the functional

$$I = \frac{1}{2} \int_0^T \left(w_1^2 + w_2^2 \right) dt$$

Using the pseudoinverse matrix (6.9) we can invert (6.17) to yield the following:

$$\begin{aligned} \varpi_1 &= \frac{(1 + x^2)\dot{x} + xy\dot{y} + y\dot{z}}{1 + x^2 + y^2} \\ \varpi_2 &= \frac{xy\dot{x} + (1 + y^2)\dot{y} - x\dot{z}}{1 + x^2 + y^2} \end{aligned}$$

One can eliminate $\dot{z}$ by using the constraint equation (6.16). The functions ϖ_i then simplify to:

$$\varpi_1 = \dot{x} \qquad \text{and} \qquad \varpi_2 = \dot{y}$$

We will now use the optimal control Boltzmann-Hamel equations to determine the differential equations which yield the optimal control. We take as quasi-velocities:

$$u_1 = y\dot{x} - x\dot{y} - \dot{z} \qquad u_2 = \dot{x} \qquad u_3 = \dot{y}$$

Notice the quasi-velocities u_2 and u_3 coincide with the functions ϖ_1 and ϖ_2, which themselves are equivalent to the control functions when evaluated along the optimal

trajectory of motion. The transformation matrices Ψ and Φ are given by:

$$\Psi = \begin{bmatrix} y & -x & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \Phi = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & y & -x \end{bmatrix}$$

The nonzero Hamel coefficients are

$$\gamma_{23}^1 = -\gamma_{32}^1 = 2$$

Expressing the integrand of the cost function in terms of quasi-velocities yields:

$$C = \frac{1}{2}\left(u_2^2 + u_3^2\right)$$

The kinematic optimal control Boltzmann-Hamel equations (6.13)-(6.15) immediately produce the following set of first order differential equations:

$$\begin{aligned} \dot{u}_2 &= -2\mu u_3 & \dot{x} &= u_2 \\ \dot{u}_3 &= 2\mu u_2 & \dot{y} &= u_3 \\ \dot{\mu} &= 0 & \dot{z} &= -u_2 + y u_2 - x u_3 \end{aligned}$$

The optimal path then satisfies this set of differential equations as well as the prescribed boundary conditions $\langle x(0), y(0), z(0)\rangle = \langle 0, 0, 0\rangle$ and $\langle x(T), y(T), z(T)\rangle = \langle 0, 0, a\rangle$.

6.4.3 Optimal Control of the Vertical Rolling Disc

The generalized coordinates of the vertical rolling disc are given by $q = \langle x, y, \theta, \phi\rangle$, where (x, y) is the contact point of the disc and the $x - y$ plane, ϕ is the angle the disc makes with the x-axis, and ϕ is the angle a reference point on the disc makes with the vertical. The motion is subject to the nonholonomic constraints:

$$\begin{aligned} \dot{x} - \cos(\phi)\dot{\theta} &= 0 \\ \dot{y} - \sin(\phi)\dot{\theta} &= 0 \end{aligned}$$

We shall consider the control vector field:

$$\dot{q} = X_1 w^1 + X_2 w^2$$

where $X_1 = \langle \cos\phi, \sin\phi, 1, 0\rangle^T$ and $X_2 = \langle 0, 0, 0, 1\rangle^T$. Using the relation for the pseudoinverse (6.9) we can define the functions ϖ_i as:

$$\begin{aligned} \varpi_1 &= \frac{\cos\phi\dot{x} + \sin\phi\dot{y} + \dot{\theta}}{2} \\ \varpi_2 &= \dot{\phi} \end{aligned}$$

Upon substituting in the nonholonomic constraints, the variables $\dot{x}$ and $\dot{y}$ may be eliminated to obtain:

$$\varpi_1 = \dot{\theta} \qquad \text{and} \qquad \varpi_2 = \dot{\phi}$$

We now consider the following optimal control problem. We wish to determine controls which steer the disc, starting from $\langle x(0), y(0), \theta(0), \phi(0)\rangle$ and stopping at $\langle x(T), y(T), \theta(T), \phi(T)\rangle$, along the path that minimizes the cost function:

$$\frac{1}{2}\int_0^T (w_1^2 + w_2^2)\, dt$$

We choose quasi-velocities:

$$\begin{aligned} u_1 &= \dot{x} - \cos(\phi)\dot{\theta} & u_3 &= \dot{\theta} \\ u_2 &= \dot{y} - \sin(\phi)\dot{\theta} & u_4 &= \dot{\phi} \end{aligned}$$

so that the transformation matrices Ψ and Φ are given by:

$$\Psi = \begin{bmatrix} 1 & 0 & -\cos\phi & 0 \\ 0 & 1 & -\sin\phi & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \Phi = \begin{bmatrix} 1 & 0 & \cos\phi & 0 \\ 0 & 1 & \sin\phi & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The nonzero Hamel coefficients are:

$$\gamma^1_{34} = \sin\phi = -\gamma^1_{43} \qquad \gamma^2_{34} = -\cos\phi = -\gamma^2_{43}$$

In terms of the quasi-velocities, the integrand of the cost function becomes $C(q,u) = \frac{1}{2}u_3^2 + \frac{1}{2}u_4^2$. The Boltzmann-Hamel equations (6.13)-(6.15) then produce the following set of first order differential equations:

$$\begin{aligned} \dot{u}_3 &= (\mu_2\cos\phi - \mu_1\sin\phi)u_4 \\ \dot{u}_4 &= (\mu_1\sin\phi - \mu_2\cos\phi)u_3 \end{aligned}$$

$$\begin{aligned} &\dot{\mu}_1 = 0 \qquad && \dot{x} = \cos(\phi)u_3 \qquad && \dot{\theta} = u_3 \\ &\dot{\mu}_2 = 0 \qquad && \dot{y} = \sin(\phi)u_3 \qquad && \dot{\phi} = u_4 \end{aligned}$$

6.4.4 Kinematic Optimal Control of the Falling Rolling Disc

The falling rolling disc can be described by the contact point (x,y) and Classical Euler angles (ϕ,θ,ψ), as shown in Figure 6.1. We will take the coordinate ordering (ϕ,θ,ψ,x,y). The system is subject to the following nonholonomic constraints

$$\dot{x} + r\dot{\psi}\cos\phi = 0 \qquad \text{and} \qquad \dot{y} + r\dot{\psi}\sin\phi = 0$$

Let the control velocity fields be given by $\dot{q} = X_1w_1 + X_2w_2 + X_3w_3$, where:

$$X_1 = \begin{bmatrix} \csc\theta \\ 0 \\ -\cot\theta \\ r\cos\phi\cot\theta \\ r\sin\phi\cot\theta \end{bmatrix}, \qquad X_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad X_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -r\cos\phi \\ -r\sin\phi \end{bmatrix}$$

Upon inverting these relations, one finds that the controls are exactly given by the functions:

$$\varpi_1 = \omega_d = \dot{\phi}\sin\theta \qquad \varpi_2 = \dot{\theta} \qquad \varpi_3 = \Omega = \dot{\phi}\cos\theta + \dot{\psi}$$

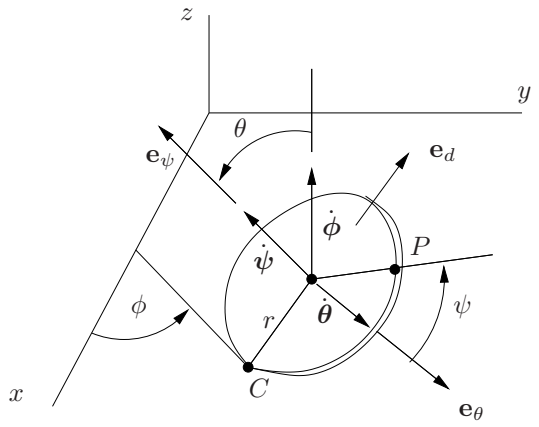

Figure 6.1: Euler Angles of the Falling Rolling Disc

In other words, one has direct control over the body-axis components of angular velocity.

We now wish to steer the disc between two fixed points while minimizing the cost functional:

$$I[\gamma] = \frac{1}{2}\int_a^b \left(w_1^2 + w_2^2 + w_3^2\right)\ dt$$

We will choose as quasi-velocities:

$$u_1 = \dot{\phi}\sin\theta \quad u_2 = \dot{\theta} \quad u_3 = \dot{\phi}\cos\theta + \dot{\psi}$$

$$u_4 = \dot{x} + r\dot{\psi}\cos\phi \qquad\qquad u_5 = \dot{y} + r\dot{\psi}\sin\phi$$

The quasi-velocities $(u_1, u_2, u_3) = (\omega_d, \dot{\theta}, \Omega)$ represent the angular velocity expressed in the body-fixed frame, and are coincident with the kinematic controls. These are not true velocities (like the Euler Angle Rates), as they are non-integrable. The nonholonomic constraints in terms of these variables are $u_4 = u_5 = 0$.

The transformation matrices are

$$\Psi = \begin{bmatrix} \sin\theta & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \cos\theta & 0 & 1 & 0 & 0 \\ 0 & 0 & r\cos\phi & 1 & 0 \\ 0 & 0 & r\sin\phi & 0 & 1 \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \csc\theta & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -\cot\theta & 0 & 1 & 0 & 0 \\ r\cos\phi\cot\theta & 0 & -r\cos\phi & 1 & 0 \\ r\sin\phi\cot\theta & 0 & -r\sin\phi & 0 & 1 \end{bmatrix}$$

The nonzero Hamel-coefficients are

$$\gamma^1_{21} = -\cot\theta = -\gamma^1_{12} \qquad \gamma^3_{21} = 1 = -\gamma^3_{12}$$

$$\gamma^4_{13} = r\sin\phi\csc\theta = -\gamma^4_{31}$$

$$\gamma^5_{13} = -r\cos\phi\csc\theta = -\gamma^5_{31}$$

Written in terms of the quasi-velocities, the integrand of the cost function is $C(q,u) = \frac{1}{2}(u_1^2 + u_2^2 + u_3^2)$.

The kinematic optimal control Boltzmann-Hamel equations (6.13)-(6.15) give us a minimal set of 10 first order differential equations:

$$\begin{aligned} \dot{u}_1 &= u_2u_3 - u_1u_2\cot\theta - r(\mu_4\sin\phi - \mu_5\cos\phi)\csc\theta u_3 \\ \dot{u}_2 &= u_1^2\cot\theta - u_1u_3 \\ \dot{u}_3 &= r(\mu_4\sin\phi - \mu_5\cos\phi)\csc\theta u_1 \end{aligned}$$

$$\dot{\mu}_4 = 0, \quad \dot{\mu}_5 = 0, \quad \dot{\phi} = \csc\theta u_1, \quad \dot{\theta} = u_2, \quad \dot{\psi} = -\cot\theta u_1 + u_3$$

$$\dot{x} = r\cos\phi\cos\theta u_1 - r\cos\phi u_3$$

$$\dot{y} = r\sin\phi\cot\theta u_1 - r\sin\phi u_3$$

6.5 Dynamic Optimal Control

In this section, we will derive a set of Boltzmann-Hamel equations for the dynamic optimal control problem, which is normally a fourth order system. We will present a minimal set of $4n - 2m$ first order differential equations that produces the optimal control. As an example, we will derive the optimal dynamic control equations for the vertical rolling disc, Chaplygin Sleigh, and free rigid body.

6.5.1 Boltzmann-Hamel Equations for Optimal Dynamic Control

Given a nonholonomic mechanical system with $n - m$ independent acceleration controls, it can be recast into the form given by the dynamical Boltzmann-Hamel equations (6.5)-(6.6). The dynamical optimal control problem is then interested in finding solution curves between two fixed points $\langle q(a), \dot{q}(a)\rangle$ and $\langle q(b), \dot{q}(b)\rangle$ that minimize the cost function

$$I = \int_a^b g(q, \dot{q}, Q)\ dt$$

Utilizing (6.5) and (6.6), we can rewrite the integrand as an explicit function of the coordinates, quasi-velocities, and quasi-accelerations:

$$C(q, u, a) = g(q, \dot{q}(q, u), Q(q, u, a))$$

Since the Boltzmann-Hamel equations no longer depend on the constrained quasi-velocities and quasi-accelerations, $C(q, u, a)$ is also independent of u^σ and a^σ. Taking variations yields:

$$\delta I = \int \left(\frac{\partial C}{\partial q^i}\delta q^i + \frac{\partial C}{\partial u^J}\delta u^J + \frac{\partial C}{\partial a^J}\delta a^J \right) dt$$

Using the second transpositional relations (6.4) for δa^J and then integrating by parts we obtain:

$$\delta I = \int \left[\frac{\partial C}{\partial q^i} \delta q^i + \left(\frac{\partial C}{\partial u^J} - \frac{d}{dt} \frac{\partial C}{\partial a^J} \right) \delta u^J \right] dt$$

Defining the parameters

$$\kappa_J = \frac{\partial C}{\partial u^J} - \frac{d}{dt} \frac{\partial C}{\partial a^J} \tag{6.18}$$

and using the first Transpositional relations (6.2) we obtain:

$$\delta I = \int \left[\left(\frac{\partial C}{\partial q^s} \Phi_i^s + \kappa_J \gamma_{si}^J u^s \right) \zeta^i - \dot{\kappa}_J \zeta^J \right] \; dt$$

These variations are not free, but subject to the nonholonomic constraints $a_i^\sigma \dot{q}^i = 0$. Since we are in the optimal control setting, we wish to find an extremum of I out of the class of kinematically admissible curves. We therefore must enforce the *vakonomic constraints* $\delta u^\sigma = 0$. To do so, append these to the integrand with a set of Lagrange multipliers μ_σ:

$$\delta I = \int \left[\left(\frac{\partial C}{\partial q^s} \Phi_i^s + \kappa_J \gamma_{si}^J u^s \right) \zeta^i - \dot{\kappa}_J \zeta^J + \mu_\sigma \delta u^\sigma \right] \; dt$$

After using the Transpositional relations (6.2) and integrating by parts, this becomes:

$$\delta I = \int \left[\left(\frac{\partial C}{\partial q^s} \Phi_i^s + \kappa_J \gamma_{si}^J u^s + \mu_\sigma \gamma_{si}^\sigma u^s \right) \zeta^i - \dot{\kappa}_J \zeta^J - \dot{\mu}_\sigma \zeta^\sigma \right] \; dt$$

where the variations are now taken to be unconstrained. Notice the multipliers μ_σ are *not* the *mechanical* multipliers, but a multiplier on the cost function that enforces Hamilton's Principle.

We thus have the following generalized Boltzmann-Hamel equations for optimal dynamic control:

$$-\frac{\partial C}{\partial q^s} \Phi_A^s + \dot{\kappa}_A - \kappa_J \gamma_{SA}^J u^S = \mu_\tau \gamma_{SA}^\tau u^S \tag{6.19}$$

and

$$-\frac{\partial C}{\partial q^s}\Phi^s_\sigma - \kappa_J \gamma^J_{S\sigma} u^S = \mu_\tau \gamma^\tau_{S\sigma} u^S - \dot{\mu}_\sigma \tag{6.20}$$

The optimal control system can therefore be given by a minimal set of $4n-2m$ first order differential equations as follows. We have n kinematic relations (6.6), $2n-2m$ relations:

$$\dot{u}^A = a^A \qquad \text{and} \qquad \dot{a}^A = \jmath^A$$

$n-m$ equations for j^A (given by inserting (6.18) into (6.19)), and, finally, m relations for the reduced multipliers $\dot{\nu}_\sigma$ (6.20). Once the resulting optimal control dynamics are determined, the control forces which produce the optimal trajectory are then given by the $n-m$ algebraic equations (6.5). The solution is then found by solving the related boundary value problem, with $4n-2m$ prescribed boundary conditions: $q^i(0),\ u^A(0),\ q^i(T),\ u^A(T)$.

6.5.2 Dynamic Optimal Control of the Vertical Rolling Disc

Consider the vertical rolling disc of §6.4.3 with control torques in the θ and ϕ directions. The corresponding dynamical equations of motion (see Bloch [16]) are:

$$\frac{3}{2}\ddot{\theta} = w_3 \qquad \frac{1}{4}\ddot{\phi} = w_4 \qquad \dot{x} = \dot{\theta}\cos\phi \qquad \dot{y} = \dot{\theta}\sin\phi$$

This is equivalent to a minimal set of 6 first order differential equations (the number obtained by using the Boltzmann-Hamel equations (6.5) and (6.6).

We now wish to choose the control forces so as to minimize the cost function $\int \frac{1}{2}(w_3^2 + w_4^2)\ dt$. Solving for the controls in terms of the quasi-accelerations $w_3 = \frac{3}{2}\ddot{\theta} = \frac{3}{2}a_3$ and $w_4 = \frac{1}{4}\ddot{\phi} = \frac{1}{4}\ddot{a}_4$, this is equivalent to minimizing the action

$$\int \left(\frac{9}{8}a_3^2 + \frac{1}{32}a_4^2\right)\ dt$$

subject to the nonholonomic constraints. Using the dynamic optimal control Boltzmann-Hamel equations (6.19) and (6.20), coupled with the dynamical equations of motion above, and eliminating the controls, we have a minimal system of 12 first order differential equations:

$$\dot{x} = \cos\phi\ u_3 \qquad\qquad \dot{\jmath}_3 = \frac{4}{9}(\mu_1 \sin\phi - \mu_2 \cos\phi)u_4$$
$$\dot{y} = \sin\phi\ u_3 \qquad\qquad \dot{\jmath}_4 = 16(-\mu_1 \sin\phi + \mu_2 \cos\phi)u_3$$

$$\dot{\theta} = u_3 \qquad \dot{u}_3 = a_3 \qquad \dot{a}_3 = \jmath_3 \qquad \dot{\mu}_1 = 0$$
$$\dot{\phi} = u_4 \qquad \dot{u}_4 = a_4 \qquad \dot{a}_4 = \jmath_4 \qquad \dot{\mu}_2 = 0$$

By use of quasi-velocities, quasi-accelerations, and quasi-jerks, we have made the following simplifications: $u_1 = u_2 = a_1 = a_2 = \jmath_1 = \jmath_2 = 0$, thereby eliminating the necessity of 6 of the 18 first order differential equations necessary in the standard approach. The solution to this system of differential equations yields the optimal dynamic control equations of the vertical rolling disc. It is equivalent to the following reduced system:

$$\dot{x} = \cos\phi\ \dot{\theta} \qquad\qquad \dddot{\theta} = \frac{4}{9}(\mu_1 \sin\phi - \mu_2 \cos\phi)\dot{\phi}$$
$$\dot{y} = \sin\phi\ \dot{\theta} \qquad\qquad \dddot{\phi} - 16(-\mu_1 \sin\phi + \mu_2 \cos\phi)\dot{\theta}$$

where $\mu_1,\ \mu_2$ are constants.

6.5.3 Dynamic Optimal Control of the Chaplygin Sleigh

Consider the Chaplygin Sleigh, discussed in Bloch [16] and the references therein. The generalized coordinates can be taken to be the (x, y) coordinate position of the point A, and the angle θ. A knife-edge constraint is given at A, whereas the other contact points can be considered as castered wheels that can roll in any direction, see Figure 6.2.

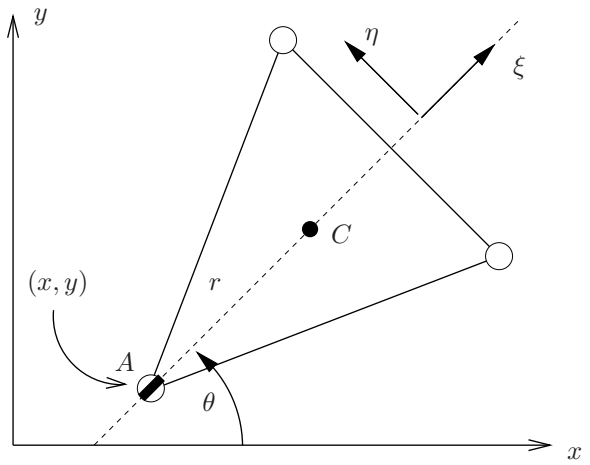

Figure 6.2: Chaplygin Sleigh

Suppose the sleigh has mass m and moment of inertia J with respect to its center of mass C. Let r be the length AC and $I := J + mr^2$.

Define quasi-velocities as follows:

$$u_1 = -\dot{x}\sin\theta + \dot{y}\cos\theta \qquad u_2 = \dot{x}\cos\theta + \dot{y}\sin\theta \qquad u_3 = \dot{\theta}$$

with transformation matrices:

$$\Psi = \begin{bmatrix} -\sin\theta & \cos\theta & 0 \\ \cos\theta & \sin\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \Phi = \begin{bmatrix} -\sin\theta & \cos\theta & 0 \\ \cos\theta & \sin\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The first quasi-direction $\frac{\partial}{\partial\theta^1} = \eta$ is the constraint direction, and the second $\frac{\partial}{\partial\theta^2} = \xi$ is the direction of motion. A control force Q_2 is given in the direction ξ and a control torque about the central axis Q_3. The nonholonomic constraint is exactly $u_1 = 0$.

The nonzero Hamel coefficients are computed to be:

$$\gamma^1_{32} = 1 \qquad \gamma^1_{23} = -1 \qquad \gamma^2_{31} = -1 \qquad \gamma^2_{13} = 1$$

The unconstrained kinetic energy for the system is:

$$\mathscr{T}(q,u) = \frac{1}{2}mu_1^2 + \frac{1}{2}mu_2^2 + \frac{1}{2}Iu_3^2 + mau_1u_3$$

The Boltzmann-Hamel equations (6.5) and the kinematic relations (6.6) work out to be:

$$Q_2 = m\dot{u}_2 - mru_3^2 \qquad Q_3 = I\dot{u}_3 + mru_2u_3 \tag{6.21}$$

$$\dot{x} = u_2\cos\theta \qquad \dot{y} = u_2\sin\theta \qquad \dot{\theta} = u_3 \tag{6.22}$$

Suppose now one wishes to find the extremal trajectories which minimize the control effort $I = \frac{1}{2}\int(Q_2^2 + Q_3^2)\,dt$. In terms of quasi-variables, the integrand can be rewritten

$$C(q,u,a) = \frac{1}{2}(m^2a_2^2 + m^2r^2u_3^4 - 2m^2ru_3^2a_2 + I^2a_3^2 \\ +m^2r^2u_2^2u_3^2 + 2Imru_2u_3a_3)$$

The variables κ_J (6.18) associated with this cost Lagrangian work out to be:

$$\begin{aligned}\kappa_2 &= m^2r^2u_2u_3^2 + Imru_3a_3 - m^2\jmath_2 + 2m^2ru_3a_3 \\ \kappa_3 &= 2m^2r^2u_3^3 - 2m^2ru_3a_2 + m^2r^2u_2^2u_3 \\ &\quad +Imru_2a_3 - I^2\jmath_3 - Imru_2a_3 - Imra_2u_3\end{aligned}$$

where $\jmath_2 = \dot{a}_2,\ \jmath_3 = \dot{a}_3$ are the quasi-jerks. The optimal control Boltzmann-Hamel equations (6.19)-(6.20) work out to be:

$$\dot{\kappa}_2 = \mu u_3 \qquad \dot{\kappa}_3 = -\mu u_2 \qquad \dot{\mu} = -\kappa_2 u_3$$

These are three first order equations for j_2, j_3, and $\dot{\mu}$. In addition, we have the kinematic relations (6.22), $\dot{u}_2 = a_2,\ \dot{u}_3 = a_3,\ \dot{a}_2 = \jmath_2$, and $\dot{a}_3 = \jmath_3$. This totals 10 first order differential equations that determine the optimal control trajectories. Once solved, the optimal controls are then given by the algebraic relations (6.21).

6.5.4 Dynamic Optimal Control of the Free Rigid Body

In this section we will discuss the optimal reorientation problem for the free rigid body. A recent real life instance of a more complicated version of this problem is

the optimal reorientation of the International Space Station using Control Moment Gyros. This optimal control problem was analyzed using pseudospectral methods, see Bedrossian et al. [9] and Kang and Bedrossian [69] for an overview. The optimal control equations themselves are laid out in Bedrossian and Bhatt [8] and Bhatt [14].

The dynamics of the free rigid body under the influence of an applied torque $\mathbf{M}$ was described in §5.9.5. We will again take the Type-I Euler angles to be our generalized coordinates and the components of the angular velocity vector with respect to a principal body fixed frame as quasi-velocities. The resulting Boltzmann-Hamel equations for this system are given in (5.80)-(5.82), and are recognized to be the Euler equations for rigid body dynamics. For notational convenience, we define the parameters:

$$\eta_{32} = I_{zz} - I_{yy} \qquad \eta_{13} = I_{xx} - I_{zz} \qquad \eta_{21} = I_{yy} - I_{xx}$$

We now wish to find the controls $\mathbf{M}(t)$ that will reorient the rigid body from one orientation into a second, while minimizing the cost function:

$$\frac{1}{2}\int (M_x^2 + M_y^2 + M_z^2)dt$$

The integrand of this cost function, when expressed in terms of quasi-variables, is given by:

$$\begin{aligned} C &= \frac{1}{2}\{I_{xx}^2 a_1^2 + I_{yy} a_2^2 + I_{zz} a_3^2 + 2I_{xx}\eta_{32} a_1 u_2 u_3 + 2I_{yy}\eta_{13} u_1 a_2 u_3 \\ &\qquad + 2I_{zz}\eta_{21} u_1 u_2 a_3 + \eta_{32}^2 u_2^2 u_3^2 + \eta_{13}^2 u_1^2 u_3^2 + \eta_{21}^2 u_1^2 u_2^2\} \end{aligned}$$

The κ's (6.18) are given by:

$$\begin{aligned}\kappa_1 &= I_{yy}\eta_{13}a_2u_3 + I_{zz}\eta_{21}u_2a_3 + \eta_{13}^2u_1u_3^2 \qquad (6.23)\\ &\quad +\eta_{21}^2u_1u_2^2 - I_{xx}\jmath_1 - I_{xx}\eta_{32}u_2a_3 - I_{xx}\eta_{32}a_2u_3\\ \kappa_2 &= I_{xx}\eta_{32}a_1u_3 + I_{zz}\eta_{21}u_1a_3 + \eta_{32}^2u_2u_3^2 \qquad (6.24)\\ &\quad +\eta_{21}^2u_1^2u_2 - I_{yy}\jmath_2 - \eta_{13}I_{yy}u_1a_3 - \eta_{13}I_{yy}a_1u_3\\ \kappa_3 &= I_{xx}\eta_{32}a_1u_2 + I_{yy}\eta_{13}u_1a_2 + \eta_{32}^2u_2^2u_3 \qquad (6.25)\\ &\quad +\eta_{13}^2u_1^2u_3 - I_{zz}\jmath_3 - \eta_{21}I_{zz}u_1a_2 - \eta_{21}I_{zz}a_1u_2\end{aligned}$$

The optimal control Boltzmann-Hamel equations (6.19) then work out to be:

$$\dot{\kappa}_1 - \kappa_2u_3 + \kappa_3u_2 = 0 \qquad (6.26)$$

$$\dot{\kappa}_2 - \kappa_3u_1 + \kappa_1u_3 = 0 \qquad (6.27)$$

$$\dot{\kappa}_3 - \kappa_1u_2 + \kappa_2u_1 = 0 \qquad (6.28)$$

These provide 3 differential equations for the $\jmath$'s. Let $\mathbb{I}$ be the moment inertia tensor with respect to the principal axes basis $\hat{\mathbf{e}}_x$, $\hat{\mathbf{e}}_y$, $\hat{\mathbf{e}}_z$, so that, in dyadic notation, $\mathbb{I} = I_{xx}\hat{\mathbf{e}}_x\hat{\mathbf{e}}_x + I_{yy}\hat{\mathbf{e}}_y\hat{\mathbf{e}}_y + I_{zz}\hat{\mathbf{e}}_z\hat{\mathbf{e}}_z$. Let $\boldsymbol{\Pi} := \mathbb{I}\cdot\boldsymbol{\omega}$ be the body axis angular momentum, and $\boldsymbol{\kappa} = \langle\kappa_1, \kappa_2, \kappa_3\rangle$. Then (6.23)-(6.25) can alternatively be re-expressed as:

$$\begin{aligned}\boldsymbol{\kappa} &= \boldsymbol{\Pi}\times\dot{\boldsymbol{\Pi}} + \boldsymbol{\Pi}\times(\boldsymbol{\omega}\times\boldsymbol{\Pi}) - \ddot{\boldsymbol{\Pi}} \qquad (6.29)\\ &\quad -\mathbb{I}\cdot\Big\{2\boldsymbol{\omega}\times\dot{\boldsymbol{\Pi}} + \dot{\boldsymbol{\omega}}\times\boldsymbol{\Pi} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\boldsymbol{\Pi})\Big\}\end{aligned}$$

(6.26) - (6.28) can be rewritten as

$$\dot{\boldsymbol{\kappa}} = \boldsymbol{\kappa}\times\boldsymbol{\omega} \qquad (6.30)$$

Finally, by defining $\boldsymbol{\lambda}(\boldsymbol{\omega}, \dot{\boldsymbol{\omega}}) = \boldsymbol{\kappa} + \ddot{\boldsymbol{\Pi}}$, the dynamic optimal control equations for the free rigid body can be expressed as:

$$\dddot{\boldsymbol{\Pi}} = \dot{\boldsymbol{\lambda}} + \ddot{\boldsymbol{\Pi}}\times\boldsymbol{\omega} - \boldsymbol{\lambda}\times\boldsymbol{\omega} \qquad (6.31)$$

In addition, we have the kinematic relations

$$\dot{\psi} = \sec\theta \sin\phi u_2 + \sec\theta \cos\phi u_3 \tag{6.32}$$

$$\dot{\theta} = \cos\phi u_2 - \sin\phi u_3 \tag{6.33}$$

$$\dot{\phi} = u_1 + \tan\theta \sin\phi u_2 + \tan\theta \cos\phi u_3 \tag{6.34}$$

as well as the relations $\dot{u}_i = a_i,\ \dot{a}_i = \jmath_i$. This is a set of 12 first order differential equations. Once one solves the corresponding boundary value problem (initial, final Euler angles, angular velocities specified), the controls are determins by the algebraic relations (5.80)-(5.82).

6.5.5 Dynamic Optimal Control of a Free Sphere

The optimal reorientation problem for a free rigid sphere is a special case of the formalism presented in the previous paragraph that has achieved some acclaim in the mathematics community recently, under the synonym "cubic splines on $SO(3)$." See, for example, the work of Noakes [108, 109, 110], Crouch and Leite [38], and the references contained therein. For this case, we have that $I_{xx} = I_{yy} = I_{zz} = I$. One therefore sees from (6.29) that $\boldsymbol{\kappa} = -\ddot{\boldsymbol{\Pi}}$ and $\boldsymbol{\lambda} = \mathbf{0}$. The Boltzmann-Hamel equations for the optimal dynamic control of the free rigid body (6.31) then reduce to:

$$\dddot{\boldsymbol{\omega}} = \ddot{\boldsymbol{\omega}} \times \boldsymbol{\omega}$$

When coupled with the kinematic relations (6.32)-(6.34) and the algebraic relations (5.80)-(5.82), the optimal control trajectories of the free rigid sphere are produced. Integrating once yields the second order system:

$$\ddot{\boldsymbol{\omega}} = \mathbf{c} + \dot{\boldsymbol{\omega}} \times \boldsymbol{\omega}. \tag{6.35}$$

which coincides with the result of Noakes, et al. [110]. See also Crouch and Leite [38] and Noakes [108, 109].

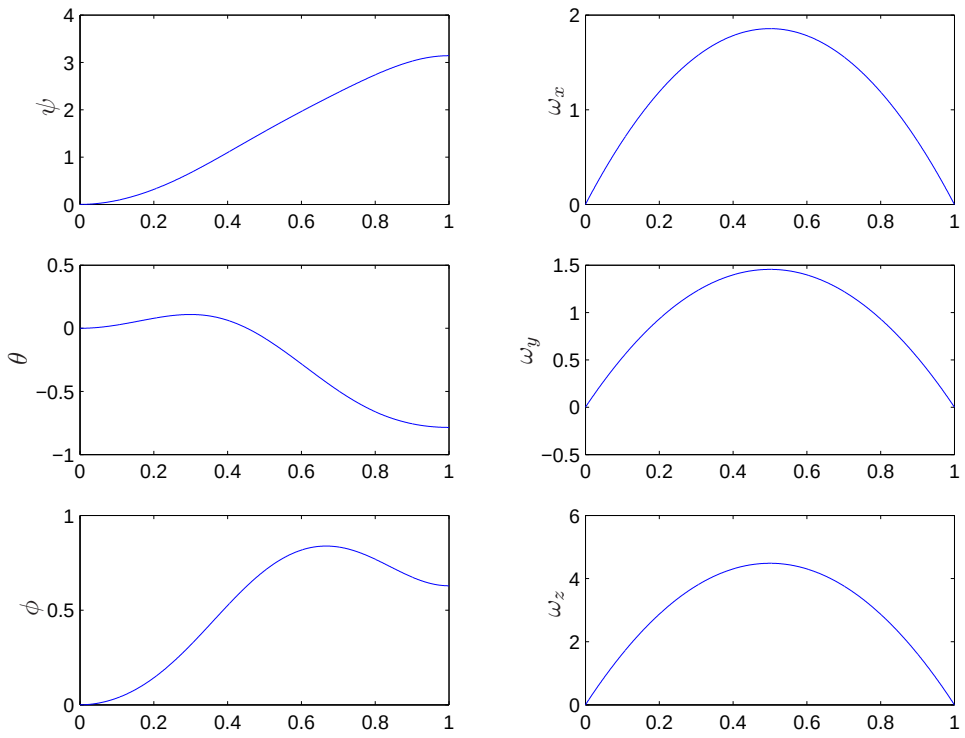

Figure 6.3: Optimal Dynamic Control of Free Sphere: Euler Angles and Body Fixed Angular Velocity with respect to time.

It is interesting to note that a particular solution of (6.35) is $\boldsymbol{\omega}(t) = \left(\frac{1}{2}t^2 + \alpha t + \beta\right)\mathbf{c}$, where α and β are constants. An analysis of these particular solutions in relationship to the general solutions will be a subject of further study.

The optimal solution trajectory of the reorientation of the rigid sphere from $\mathbf{q}(0) = \langle 0, 0, 0\rangle$, $\boldsymbol{\omega}(0) = \langle 0, 0, 0\rangle$ to the point $\mathbf{q}(1) = \langle \pi, -\pi/4, \pi/5\rangle$, $\boldsymbol{\omega}(1) = \langle 0, 0, 0\rangle$ is plotted in Fig. 6.3.

6.6 Conclusion

In this chapter we showed how one can extend quasi-velocity techniques to kinematic and optimal control problems. For kinematic optimal control problems, one gains a saving of m first order differential equations, as one need not integrate the constraint quasi-velocities: $u^\sigma \equiv 0$. For dynamic optimal control problems, the sav-

ing increases to $3m$ first order equations, as one no longer need integrate the m constrained quasi-velocities, quasi-accelerations, and quasi-jerks, $u^\sigma \equiv 0$, $a^\sigma \equiv 0$, $\jmath^\sigma \equiv 0$, respectively. Initial and final conditions are then enforced by solving the resulting system of differential equations as a two point boundary value problem.

CHAPTER VII

Mechanics and Control on Lie Groups

In this chapter we consider the special case where the underlying configuration manifold for our system is a Lie group G. For this case all of the dynamics can be pulled back to the Lie algebra $\mathfrak{g}$, which is simply the tangent space to the identity $\mathfrak{g} \equiv T_eG$. The pullback of the velocity vector to the Lie algebra is actually tantamount to defining a set of quasi-velocities for the flow; therefore the quasi-velocity techniques described in Chapters V and VI are suitable for providing geometric insight into these systems, albeit they are unconstrained systems. The equations of motion that govern the resulting dynamics are set of equations known as the Euler-Poincaré equations. We will show that these are really a Lie group version of the Boltzmann-Hamel equations discussed in Chapter V and further show that, analogously to the discussion in Chapter VI, they can be extended to a set of Lie group optimal control equations.

7.1 Lie Groups

In this section we will present some of the mathematical background on Lie groups and Lie algebras in preparation for our subsequent discussion of mechanics on Lie groups. For more details on the theory of Lie groups and their application to mechanics and control, see Bloch [16], Bullo and Lewis [27], and Marsden and Ratiu [83]. We will pay particular attention throughout this chapter to rigid body mechanics

from a Lie group perspective. We will further pay attention to the case of a non-left-invariant Lagrangian and discuss in some analytic detail the use of exponential coordinates on Lie groups. We will illustrate this towards the end of this chapter with a discussion of the heavy top.

7.1.1 Preliminary Definitions

We begin with some preliminary definitions.

Definition 41. *A **Lie group** G is an n-dimensional differential manifold endowed with a group structure, i.e. the underlying topological set G has a binary operator $\star : (a, b) \in G \times G \to a \star b \in G$ that satisfies the following properties:*

(i) $a \star (b \star c) = (a \star b) \star c$ *for all* $a, b, c \in G$.

(ii) there exists $e \in G$ *such stat* $a \star e = e \star a = a$ *for all* $a \in G$.

(iii) for each $a \in G$, *there exists* $a^{-1} \in G$ *such that* $a \star a^{-1} = a^{-1} \star a = e$.

Further the maps $(a, b) \to a \star b$ *and* $a \to a^{-1}$ *must be smooth.*

We will refer to $g \star h$ simply as gh when there is no confusion. Throughout this chapter we will consider only *matrix Lie groups*, i.e. groups whose elements are real matrices, but will from time to time use notation that is more general.

Definition 42. *A **Lie algebra** V is a real vector space endowed with a bilinear operation $[\cdot, \cdot] : V \times V \to V$ called the **bracket** satisfying:*

(i) $[\xi, \eta] = -[\eta, \xi]$, *for all* $\xi, \eta \in V$ *(anti-commutativity).*

(ii) $[\xi, [\eta, \zeta]] + [\eta, [\zeta, \xi]] + [\zeta, [\xi, \eta]] = 0$ *for all* $\xi, \eta, \zeta \in V$ *(the Jacobi identity).*

For a Lie group G and fixed $g \in G$ we define the *left translation map* $L_g : G \to G$ by $L_g h = gh$. Since L_g takes the identity element e to the point g, the differential

$d\left(L_g\right)_e$ is a natural isomorphism between the tangent space at the identity, T_eG, and the tangent space at g, T_gG. Similarly, the map $L_{g^{-1}}$ takes the point g to the identity, thus the vector space isomorphism $d\left(L_{g^{-1}}\right)_g : T_gG \to T_eG$ takes tangent vectors from the tangent space at g and maps them into the Lie algebra $T_eG = \mathfrak{g}$. See Fig. 7.1 for an illustration of this.

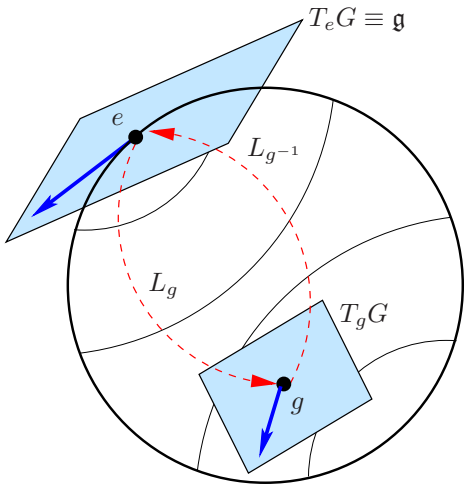

Figure 7.1: The Lie group G. The blue tangent vectors are $v \in T_eG$ and $dL_g \cdot v \in T_gG$.

A vector field X on G is *left-invariant* if $X(gh) = d(L_g)_h(X(h))$ for all $g, h \in G$. Any left-invariant vector field can therefore be identified by its value at the the identity, since $X(g) = d(L_g)_e(X(e))$. For any $\xi \in T_eG$ one can therefore define an associated left-invariant vector field by $\xi_L(g) = d(L_g)_e(\xi)$. Every Lie group has with it an associated Lie algebra:

Definition 43. *The* ***Lie algebra*** $\mathfrak{g}$ ***of a Lie group*** $\mathbf{G}$ *is the tangent space at the identity* T_eG *with bracket given by* $[\xi, \eta] = [\xi_L, \eta_L](e)$. *Note: we will use* $\mathfrak{g} \equiv T_eG$ *indistinguishably.*

7.1.2 Spatial Velocities, Body Velocities, and the Adjoint Map

In addition to the left-translation map, defined by $L_g h = gh$, one could alternatively define the right-translation map $R_g h = hg$. The right-translation maps provide an alternative vector space isomorphism $d(R_g)_e : \mathfrak{g} \to T_g G$. The pullback of a vector to the Lie algebra can therefore be accomplished either by the left- or right-translation maps; however, a single vector $X \in T_g G$ will pullback as two distinct vectors $d(L_{g^{-1}})_g(X) \in \mathfrak{g}$ and $d(R_{g^{-1}})_g(X) \in \mathfrak{g}$, with $d(L_{g^{-1}})_g(X) \neq d(R_{g^{-1}})_g(X)$, under the two distinct maps.

Definition 44. *For a curve $g(t) : I \subset \mathbb{R} \to G$, define:*

(i) the ***material velocity*** *of $g(t)$ as $\dot{g}(t) \in T_{g(t)}G$,*

(ii) the ***body velocity*** *of $g(t)$ as $d(L_{g(t)^{-1}})_{g(t)}(\dot{g}(t)) \in \mathfrak{g}$, and*

(iii) the ***spatial velocity*** *of $g(t)$ as $d(R_{g(t)^{-1}})_{g(t)}(\dot{g}(t)) \in \mathfrak{g}$.*

Figure 7.2 shows how $X \in T_g G$ maps to the Lie algebra under the two different maps $d(L_{g^{-1}})_g$ and $d(R_{g^{-1}})_g$. For a fixed $g \in G$, the *adjoint map* $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ takes the body representation of a vector to its spatial representation, i.e. it takes the pullback by the left-translation map of a vector $X \in T_g G$ to the Lie algebra to the corresponding pullback by the right-translation map.

For $g \in G$ and $\xi \in \mathfrak{g}$, the adjoint map is defined by:

$$\mathrm{Ad}_g \xi = d(R_{g^{-1}})_g \left(d(L_g)_e(\xi) \right)$$

For matrix Lie groups, this expression simplifies to

$$\mathrm{Ad}_g \xi = g \cdot \xi \cdot g^{-1}$$

For vectors $\xi, \eta \in \mathfrak{g}$, differentiating the adjoint map $\mathrm{Ad}_g \eta$ with respect to g at $g = e$ in the direction ξ produces the *adjoint operator* $\mathrm{ad}_\xi : \mathfrak{g} \to \mathfrak{g}$. It is well-known that

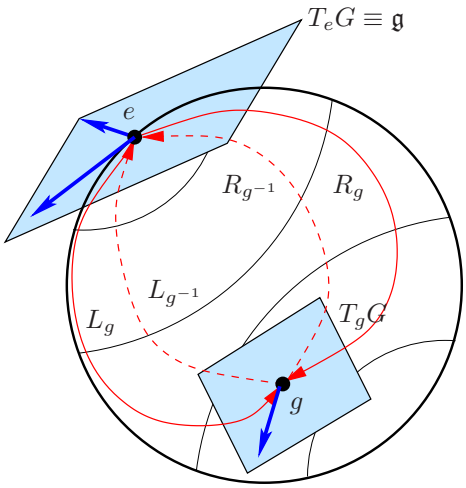

Figure 7.2: The left- and right-translation maps and the Adjoint map

this is identical to the bracket operator of the Lie algebra:

$$\mathrm{ad}_\xi \eta = [\xi, \eta]$$

Note that without reference to a fixed g or ξ, respectively the maps Ad_g and ad_ξ can be thought of as binary operations $\mathrm{Ad} : G \times \mathfrak{g} \to \mathfrak{g}$ and $\mathrm{ad} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, respectively.

7.1.3 The Exponential Map

Another important concept that will be important for our later discussion of mechanics on Lie groups is that of the exponential map. For a fixed $\xi \in \mathfrak{g}$, we have seen how one can define a corresponding left-invariant vector field $\xi_L \in TG$ by the relation $\xi_L(g) = d(L_g)_e(\xi)$. One can similarly define a corresponding right-invariant vector field $\xi_R \in TG$ by the relation $\xi_R(g) = d(R_g)_e(\xi)$. For matrix Lie groups, these relations simplify to $\xi_L(g) = g \cdot \xi \in T_gG$ and $\xi_R(g) = \xi \cdot g \in T_gG$. The left- and right-invariant vector fields generated by an element $\xi \in \mathfrak{g}$ are important vector fields in mechanics, and will arise in our definition of the exponential map. Before we introduce this, we need one further definition:

Definition 45. *Let $X : G \times \mathbb{R} \to TG$ be a (possibly time-varying) vector field on G*

so that $X(g;t) \in T_gG$. Consider the initial value problem:

$$\frac{dg}{dt} = X(g;t) \qquad g(0) = g_0 \in G \tag{7.1}$$

The solution of this differential equation is a one-parameter subgroup of G known as ***the flow of*** *g* ***along*** *X, and is denoted by $g(t) = \phi(t; g_0; X)$.*

In particular, given a vector $\xi \in \mathfrak{g}$, one can flow along the left-invariant vector field that is generated by ξ. We now define the exponential map as:

Definition 46. *Given an element $\xi \in \mathfrak{g}$ and its associated left-invariant vector field $\xi_L(g) = d(L_g)_e(\xi)$, we define the* ***exponential map*** $\exp : \mathfrak{g} \to G$ *as the flow of g along ξ_L for a unit interval of time:*

$$\exp(\xi) = \phi(1; e; \xi_L)$$

The exponential map enjoys the following property. We refer the reader to Warner [132] for the proof.

Theorem 47. *Let G be a Lie group and $\xi \in \mathfrak{g}$ an element of the Lie algebra. Let $\xi_L(g)$ and $\xi_R(g)$ be the left- and right-invariant vector fields on TG generated by ξ, respectively. Then:*

$$\phi(t; g; \xi_L) = L_g \circ \exp(\xi t) \tag{7.2}$$

$$\phi(t; g; \xi_R) = R_g \circ \exp(\xi t) \tag{7.3}$$

where $\phi(t; g; X)$ is the solution of the differential equation (7.1) *defined in Def. 45. Left- and right-invariant vector fields are therefore complete.*

Turning back to our specialization of mechanics on *matrix* Lie groups, consider now the following definition for the matrix exponential:

Definition 48. *If A is a matrix, the* ***matrix exponential*** *is defined by:*

$$\mathrm{expm}(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

The matrix-exponential is an absolutely convergent power series. It is well-known that for matrix Lie groups the map $\exp : \mathfrak{g} \to G$ coincides with the matrix exponential, i.e. for $\xi \in \mathfrak{g}$, $\mathrm{expm}(\xi) \in G$ and further $\mathrm{expm}(\xi) = \exp(\xi)$. Therefore, in our discussion of matrix Lie groups and matrix Lie algebras, we will exclusively use the notation $\exp(\xi)$ to refer to both definitions.

The exponential map is also useful in defining a set of coordinates on G. Suppose $\{E_1, \ldots, E_n\}$ is a basis of the Lie algebra $\mathfrak{g}$. Then one can define a set of coordinates $\{\theta^1, \ldots, \theta^n\}$ for G, with $\theta^i = 0$ corresponding to the identity, by the relation $g(\theta^1, \ldots, \theta^n) = \exp(\theta^1 E_1) \star \cdots \star \exp(\theta^n E_n)$. This will be useful in our discussion of the generalized Euler-Poincaré equations in §7.6.

7.2 Quasi-Velocities on Lie Groups

As we shall see, a natural set of quasi-velocities exist on any Lie Group. It is the goal of this paragraph to discuss how these quasi-velocities arise and further write down their corresponding transpositional relations. In the following section we will then show how the Lagrange-D'Alembert Principle can be written for Lie groups and from this determine the resulting dynamical equations of motion.

Suppose the n vectors $E_i \in T_eG$ form a basis for the Lie algebra. They can then be thought of as a quasi-basis for the tangent bundle TG, and the components of the pullback of the material velocity to the Lie algebra via the left-translation map with respect to this basis form a set of quasi-velocities for the system. To see how this comes about, consider a basis $\{X_i\}_{i=1}^n$ for T_gG and let the transformation $\mathbf{\Psi}$ be defined by the linear transformation $d\left(L_{g^{-1}}\right)_g$ with respect to these bases, given by

the relation:

$$d\left(L_{g^{-1}}\right)_g(X_i) = \mathbf{\Psi}_i^j E_j \in T_eG$$

Note that we have not made any reference to coordinates on G, we only require a basis for T_gG. This is all that is required for systems with left-invariant Lagrangians. Treatment of coordinates on G and mechanical systems whose Lagrangian is not left-invariant shall be postponed until §7.6. The most common basis for T_gG in the absence of coordinates $\{q^i\}_{i=1}^n$ on G (which induce the basis $\{\partial/\partial q^i\}_{i=1}^n$) is simply the pushforwards of the basis of the Lie algebra by the left-translation map, i.e.:

$$T_gG = \text{span}\left(d(L_g)_e(E_1), \ldots, d(L_g)_e(E_n)\right) = \text{span}\left(gE_1, \ldots, gE_n\right)$$

where the second equality holds if G is a matrix Lie group. A velocity vector $\dot{g} = v^i X_i \in T_gG$ for the curve $g(t) \in G$ therefore transforms as:

$$d\left(L_{g^{-1}}\right)_g\left(v^i X_i\right) = \mathbf{\Psi}_i^j v^i E_j \in T_eG$$

The *body velocity* of this curve is defined as:

$$\xi(t) = d\left(L_{g(t)^{-1}}\right)_{g(t)} \dot{g}(t)$$

The components of $\xi(t)$ with respect to the basis of the Lie algebra are therefore given by $\xi^i(t) = E^i(\xi) = \Psi_j^i v^j$ and further constitute a set of quasi-velocities for the system, where the covectors $E^i \in \mathfrak{g}^*$ form the dual basis. In our present discussion, we will restrict our attention to the special case when G is a matrix Lie group. In this case $\mathbf{\Psi} = g(t)^{-1}$, and thus:

$$\xi(t) = g(t)^{-1}\dot{g}(t)$$

In the case where $G = SO(3)$, the components of $\xi(t)$ with respect to the basis for $T_ISO(3)$ are simply the components of the angular velocity vector for a rigid body expressed relative to the body fixed reference frame.

Taking variations of the curve $g(t)$, we define the *body infinitesimal variation* as the pullback of the infinitesimal variation to the Lie algebra:

$$\eta(t) = d\left(L_{g(t)^{-1}}\right)_{g(t)} \delta g(t) = g(t)^{-1}\delta g(t)$$

where $\delta = \frac{\partial}{\partial \varepsilon}$ denotes a derivative in the direction of the variation. As above, the components of $\eta(t)$ with respect to the basis of the Lie algebra are the corresponding quasi-infinitesimal variations.

There is an analogous formulation of the transpositional relations (Cor. 26) for Lie groups. We present it as the following Lemma:

Lemma 49 (Lie Group Transpositional Relations). *Let $g : U \subset \mathbb{R}^2 \to G$ be a proper variation in a matrix Lie group and denote the pullback of its partial derivatives to the Lie algebra by:*

$$\xi(t,\varepsilon) = g^{-1}\frac{\partial g}{\partial t} \qquad \eta(t,\varepsilon) = g^{-1}\frac{\partial g}{\partial \varepsilon}$$

Then the following Lie group transpositional relations hold:

$$g^{-1}\left(\frac{d\delta g}{dt} - \delta \dot{g}\right) = (\dot{\eta} - \delta\xi) + [\xi, \eta] \tag{7.4}$$

Proof. The proof is straightforward:

$$\begin{aligned}\frac{\partial \xi}{\partial \varepsilon} - \frac{\partial \eta}{\partial t} &= -g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial g}{\partial t} + g^{-1}\frac{\partial^2 g}{\partial \varepsilon \partial t} + g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial g}{\partial \varepsilon} - g^{-1}\frac{\partial^2 g}{\partial t \partial \varepsilon} \\ &= g^{-1}\left(\delta\dot{g} - \frac{d\delta g}{dt}\right) + \xi\eta - \eta\xi = g^{-1}\left(\delta\dot{g} - \frac{d\delta g}{dt}\right) + [\xi,\eta]\end{aligned}$$

Rearranging the terms produces the result. □

Notice that this is in the same form as the standard transpositional relations given in (5.43), since, for matrix Lie groups, the transformation matrix $\mathbf{\Psi}$ is identified

with g^{-1}. Also notice that the Hamel coefficients corresponding to this set of quasi-velocities, which can be thought of as a components of a mapping $\gamma : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, must be such that:

$$\gamma(X, Y) = [X, Y], \qquad \forall X, Y \in \mathfrak{g}$$

Notice for an unconstrained system or for a system whose variations are chosen to satisfy the first transitivity relation (T1), that one can take the variations to be smooth in the sense that $\partial_s \partial_t g = \partial_t \partial_s g$. In this case, the Lie group transpositional relations (7.4) reduce to the form they were presented as in Bloch, et al. [21]:

$$\delta\xi = \dot{\eta} + [\xi, \eta] \tag{7.5}$$

As was the case in §5.6 and in §5.9, a correct understanding of the transpositional relations was key to determining the correct dynamical equations of motion. With the Lie group transpositional relations in hand, we are now in a position to derive the dynamical equations of motion for Lie groups.

7.3 The Euler-Poincaré Dynamical Equations

The topic of this section will be the Euler-Poincaré equations for left-invariant Lagrangians, which describe the dynamical evolution of mechanical systems when the underlying configuration manifold is a Lie group. These equations are well studied, and the derivation presented in this section is taken from Bloch [16], Marsden and Ratiu [83], and Bloch, et al. [21]. We will generalize these to the case where the Lagrangian is not left-invariant in §7.6.

Theorem 50. *Let G be a matrix Lie group and $L : TG \to \mathbb{R}$ a left invariant Lagrangian. Let $l : \mathfrak{g} \to \mathbb{R}$ be its restriction to the tangent space at the identity. For a curve $g(t) \in G$, let $\xi(t) = g(t)^{-1}\dot{g}(t)$. Then the following are equivalent:*

1. $g(t)$ satisfies the Euler-Lagrange equations for L on G:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{g}^i} - \frac{\partial L}{\partial g^i} = 0$$

2. The variational principle

$$\delta \int_a^b L(g(t), \dot{g}(t))\ dt = 0$$

holds, for proper variations of g.

3. The Euler-Poincaré equations hold:

$$\frac{d}{dt}\frac{\delta l}{\delta \xi} = \mathrm{ad}^*_\xi \frac{\delta l}{\delta \xi}$$

4. The variational principle

$$\delta \int_a^b l(\xi(t))\ dt = 0$$

holds on $\mathfrak{g}$, using variations of the form

$$\delta\xi = \dot{\eta} + [\xi, \eta]$$

where η vanishes at the endpoints.

Proof. The equivalence of 1. and 2. holds for any configuration manifold.

To show equivalence of 2. and 4., we must simply show that all variations $\delta g(t) \in TG$ of $g(t)$ induce variations $\delta\xi$ of $\xi(t) \in \mathfrak{g}$ of the form $\delta\xi = \dot{\eta} + [\xi, \eta]$, where $\eta(t)$ vanishes at the endpoints. This, however, is the content of the Lemma 49.

To show equivalence of 3. and 4., we explicitly compute variations:

$$\begin{aligned}
\delta \int l(\xi(t))\ dt &= \int \frac{\delta l}{\delta \xi}\delta\xi\ dt \\
&= \int \frac{\delta l}{\delta \xi}\left(\dot{\eta} + \mathrm{ad}_\xi \eta\right)\ dt \\
&= \int \left[-\frac{d}{dt}\left(\frac{\delta l}{\delta \xi}\right) + \mathrm{ad}^*_\xi \frac{\delta l}{\delta \xi}\right] \eta\ dt
\end{aligned}$$

which vanishes iff the Euler-Poincaré equations hold. □

We can add forces to our mechanical system as follows. This is a direct generalization of the previous theorem and is discussed in Bloch, et al. [21]. Most of the proof follows in analogue to Theorem 50; however there are a few subtle differences which we shall emphasize.

Theorem 51. *Let G, L, l, g, and ξ be as in Theorem 1. Let $F \in T^*G$ be a force field equivariant relative to the canonical left actions of G on TG and T^*G, and let $f \in \mathfrak{g}^*$ be its restriction to the Lie algebra, obtained by $f = Fg$. Then the following are equivalent:*

1. *$g(t)$ satisfies the Euler-Lagrange equations with force field F.*

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{g}^i} - \frac{\partial L}{\partial g^i} = F_i \tag{7.6}$$

2. *The Lagrange-d'Alembert principle holds:*

$$\delta \int_a^b L(g(t), \dot{g}(t))\ dt + \int_a^b F(g(t), \dot{g}(t)) \cdot \delta g(t)\ dt = 0 \tag{7.7}$$

   holds for all proper variations $\delta g(t)$.

3. *The Euler-Poincaré equations hold:*

$$\frac{d}{dt}\frac{\delta l}{\delta \xi} - \mathrm{ad}^*_\xi \frac{\delta l}{\delta \xi} = f \tag{7.8}$$

4. *The variational principle*

$$\delta \int_a^b l(\xi(t))\ dt + \int_a^b f(\xi(t)) \cdot \eta(t)\ dt = 0 \tag{7.9}$$

   holds on $\mathfrak{g}$, using variations of the form $\delta\xi = \dot{\eta} + [\xi, \eta]$, where η vanish at the endpoints.

Proof. The equivalence of 1. and 2. hold for any configuration manifold.

We have already showed equivalence of the first term of (7.7) and the first term of (7.9) in our proof of Theorem 50. To demonstrate equivalence of the second term of these equations, consider the following. The pullback of δg to the Lie algebra is given by the mapping

$$d\left(L_{g^{-1}}\right)_g : T_gG \to T_eG$$

so that $\eta = d\left(L_{g^{-1}}\right)_g \delta g = g^{-1}\delta g \in \mathfrak{g}$. The pullback of F to the dual of the Lie algebra is given by the dual of the above mapping:

$$d^*\left(L_{g^{-1}}\right)_g : T_g^*G \to T_e^*G$$

so that $f = d^*\left(L_{g^{-1}}\right)_g F = Fg \in \mathfrak{g}^*$. We therefore find that:

$$f \cdot \eta = (Fg) \cdot \left(g^{-1}\delta g\right) = F \cdot \delta g$$

The equivalence of (7.8) and (7.9) can be demonstrated by following the steps in the proof of Theorem 50. □

7.4 The Rigid Body from a Lie Group Perspective

The Euler equations for rigid body motion were derived in §5.9.5 from the Boltzmann-Hamel equations. We return now again to the study of these equations from a Lie group perspective, for the purpose of understanding the rotational motion of rigid bodies from a more geometric perspective and also of illustrating the theory developed above on a practical example.

7.4.1 Rotation Matrices of $SO(3)$

The configuration manifold for the free rigid body can be taken to be the Lie group $SO(3) = \{A \in \mathbb{R}^{n\times n} : A^TA = AA^T = I \text{ and } \det A = 1\}$. An element $A \in SO(3)$ is identified with the transformation matrix relative to a given coordinate system

that maps the inertial coordinates into the coordinates of the body fixed frame. The components of this transformation matrix are related to any choice of Euler angles. For instance, as is shown in Greenwood [58], in terms of the Type II Euler angles (ϕ, θ, ψ) considered in §5.9.5, the corresponding transformation matrix would be given by:

$$A = \begin{bmatrix} (\mathrm{c}\phi\mathrm{c}\psi - \mathrm{s}\phi\mathrm{c}\theta\mathrm{s}\psi) & (-\mathrm{s}\phi\mathrm{c}\psi - \mathrm{c}\phi\mathrm{c}\theta\mathrm{s}\psi) & \mathrm{s}\theta\mathrm{s}\psi \\ (\mathrm{c}\phi\mathrm{s}\psi + \mathrm{s}\phi\mathrm{c}\theta\mathrm{c}\psi) & (-\mathrm{s}\phi\mathrm{s}\psi + \mathrm{c}\phi\mathrm{c}\theta\mathrm{c}\psi) & -\mathrm{s}\theta\mathrm{c}\psi \\ \mathrm{s}\phi\mathrm{s}\theta & \mathrm{c}\phi\mathrm{s}\theta & \mathrm{c}\theta \end{bmatrix} \tag{7.10}$$

where we have abbreviated sin and cos with s and c, respectively. This matrix is obtained simply by taking a rotation of ϕ about the inertial z axis, a rotation of θ about the resulting x axis, and a final rotation of ψ about the resulting z axis.

7.4.2 The Lie Algebra: $\mathfrak{so}(3)$

The identity element of the group $SO(3)$ is the 3×3 identity matrix. Consider the one-parameter rotation groups about the identity transformation:

$$R_1(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(t) & -\sin(t) \\ 0 & \sin(t) & \cos(t) \end{bmatrix} \qquad R_2(t) = \begin{bmatrix} \cos(t) & 0 & \sin(t) \\ 0 & 1 & 0 \\ -\sin(t) & 0 & \cos(t) \end{bmatrix}$$

$$R_3(t) = \begin{bmatrix} \cos(t) & -\sin(t) & 0 \\ \sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

These curves represent rotations about the three principal axes (assuming the principal axes initially coincide with the inertial reference system). By differentiating with respect to t and evaluating at $t = 0$, one finds that the Lie algebra $\mathfrak{so}(3)$ is generated

by the following three basis vectors:

$$E_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \qquad E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \qquad E_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{7.11}$$

The Lie algebra $\mathfrak{so}(3)$ therefore consists of the set of 3×3 skew-symmetric matrices. Moreover, the above rotation curves are the matrix exponentials of the basis vectors, i.e.:

$$R_1(t) = \exp(tE_1), \qquad R_2(t) = \exp(tE_2), \qquad R_3(t) = \exp(tE_3)$$

Since the transformation matrix $A \in SO(3)$ (7.10) is formed by a rotation of ϕ about the z-axis, followed by a rotation of θ about the resulting x-axis, followed by a final rotation of ψ about the resulting z-axis; we could alternatively express it as $A = R_3(\psi) \cdot R_1(\theta) \cdot R_3(\phi)$, or as:

$$A = \exp(\psi E_3) \cdot \exp(\theta E_1) \cdot \exp(\phi E_3) \tag{7.12}$$

We can identify $\mathbb{R}^3$ with $\mathfrak{so}(3)$ by the hat mapping $\widehat{(\ )} : \mathbb{R}^3 \to \mathfrak{so}(3)$ given by:

$$\hat{A} = \begin{pmatrix} 0 & -A_z & A_y \\ A_z & 0 & -A_x \\ -A_y & A_x & 0 \end{pmatrix}$$

for $A = \langle A_x, A_y, A_z \rangle^T \in \mathbb{R}^3$. One can easily show that the Lie bracket for $\mathfrak{so}(3)$ yields, for vectors $A, B \in \mathbb{R}^3$:

$$\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A} = \widehat{A \times B}$$

We therefore see that the Lie algebra $\mathfrak{so}(3)$ is isomorphic to $\mathbb{R}^3$ with the cross product operator, i.e. $\mathfrak{so}(3) \cong (\mathbb{R}^3, \times)$.

7.4.3 Euler's Equations

Suppose the body fixed angular velocity of the rigid body is given by $\Omega \in \mathbb{R}^3$. An equivalent expression of the body-fixed angular velocity is given by $\tilde{\Omega} \in \mathfrak{so}(3)$. Using the above Lie algebra isomorphism, the restriction of the Lagrangian to the Lie algebra can be written as follows:

$$l(\Omega) = \frac{1}{2}\Omega^T \cdot \mathbb{I} \cdot \Omega$$

where $\mathbb{I}$ is the moment of inertia tensor of the rigid body. If one chooses a principal axis frame, this inertia tensor can be taken to be of the form $\mathbb{I} = \text{diag}(I_{xx}, I_{yy}, I_{zz})$. For any $X \in \mathbb{R}^3$, we have:

$$\left\langle \text{ad}^*_\Omega \frac{\delta l}{\delta \Omega}, X \right\rangle = \left\langle \frac{\delta l}{\delta \Omega}, \text{ad}_\Omega X \right\rangle$$

Since the bracket $[\cdot, \cdot]$ on $\mathfrak{so}(3)$ is identified with the cross product, this above equation is identical to:

$$\frac{\delta l}{\delta \Omega} \cdot (\Omega \times X) = \left(\frac{\delta l}{\delta \Omega} \times \Omega \right) \cdot X$$

We therefore have that, for the Lie algebra $\mathfrak{so}(3)$, the following:

$$\text{ad}^*_\Omega \frac{\delta l}{\delta \Omega} = \frac{\delta l}{\delta \Omega} \times \Omega$$

It is further clear that $\delta l / \delta \Omega = \mathbb{I}\Omega$. Defining the body axis angular momentum as $\Pi = \mathbb{I} \cdot \Omega$ and using the above results, we find that the Euler-Poincaré equations (7.8) produce the following set of dynamical equations for the free rigid body:

$$\dot{\Pi} - \Pi \times \Omega = M \tag{7.13}$$

Here M is the applied torque expressed with respect to the body fixed principal axis. This agrees precisely with the set of equations (5.80)-(5.82), obtained at the end of §5.9.5 from the Boltzmann-Hamel equations.

7.5 The Relation between Euler-Poincaré and Boltzmann-Hamel

Throughout our discussion of Lie groups we have strived to relate each topic to the formalism developed earlier this chapter for quasi-velocities. In particular, we recognized the equations from Lemma 49 as a form of the transpositional relations (5.43) suitable for Lie groups. As it turns out, the Euler-Poincaré equations themselves are a special case of the Boltzmann-Hamel equations.

To see this, we begin by noting that the Euler-Poincaré equations (7.8) can be written in coordinate notation as follows:

$$\frac{d}{dt}\frac{\partial l}{\partial \xi^i} - \gamma_{ki}^j \frac{\partial l}{\partial \xi^j}\xi^k = f_i \tag{7.14}$$

where γ_{ki}^j are the structure constants of the Lie algebra $\mathfrak{g}$. Compare now the coordinate form of the Euler-Poincaré equations, (7.14), with the Boltzmann-Hamel equations for unconstrained systems, (5.70). Recall that $f = Fg$ and that we have the identification of $g = \mathbf{\Phi}$, as discussed in §7.2, so that the right hand side of each equation agrees. Since $l(\xi)$ is independent of the position $g(t)$, we conclude the two sets of equations must be identical, and further that the Hamel coefficients are simply the structure coefficients of the Lie algebra.

7.6 The Generalized Euler-Poincaré Equations

The goal of this section is to generalize the Euler-Poincaré dynamical equations to the case where the Lagrangian function is *not* left-invariant. This situation arised naturally in systems that are acted upon by conservative force fields. This situation is discussed in terms of the general theory of reduction and semi-direct products in Marsden, Ratiu, and Weinstein [84, 85]. One can naturally handle the situation with the Boltzmann-Hamel equations as well. Our approach will be to generalize the

Euler-Poincaré equations further to their full term. In this case they will coincide with the Boltzmann-Hamel equations, including the $\partial \mathscr{L}/\partial q$ term. We will discuss how this term arises in terms of mechanics on Lie groups. Since we will be handling the case of a non-left-invariant Lagrangian, defining a set of coordinates on the Lie group will play an important role. We will begin with a discussion of exponential coordinates and relate the variations of these coordinates to the material and body variations. A brief discussion similar to this can be found in Hairer et al. [59]. We will also relate the coordinate velocity to the material and body velocities.

7.6.1 Exponential Coordinates on Lie Groups and Their Variations

In studying the mechanics on Lie groups induced by a non-left-invariant Lagrangian function one must assign a set of generalized coordinates $\{q^i\}_{i=1}^n$ on G. The standard choice is a set of exponential coordinates, which we will treat in detail momentarily. One then defines a coordinate map $\phi : \mathbb{R}^n \to G$ such that $g(q) = \phi(q)$ for $q \in U \subset \mathbb{R}^n$. The vectors $\partial/\partial q^i$ then form a basis of T_gG. As we have seen in the derivation of the classical Euler-Poincaré equations, variations are ultimately expressed relative to the Lie algebra ($\eta = d(L_{g^{-1}})_g \delta g$). The Lagrangian will ultimately be expressed as $l(q, \xi)$, where ξ is the pullback of the velocity to the Lie algebra. One must then be able to write the term

$$\frac{\partial l}{\partial q} \delta q$$

in terms of the quasi-variation $\eta \in \mathfrak{g}$. Consider the map $\Psi : \mathbb{R}^n \to (\mathbb{R}^n, \diamond)$ that takes the *coordinate representation* of the tangent space T_gG in terms of the basis $\{\partial/\partial q^i\}_{i=1}^n$ to the *coordinate representation* of the Lie algebra $\mathfrak{g}$ in terms of the basis $\{E_i\}_{i=1}^n$. (The operator $\diamond : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined by the relation:

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = \widehat{A \diamond B} \qquad \text{for all } A, B \in \mathbb{R}^n).$$

We have already seen that for the case $\mathfrak{g} = \mathfrak{so}(3)$ the operator $\diamond$ corresponds to the cross product on $\mathbb{R}^3$. By examination of Fig. 7.3, we see that:

$$\begin{aligned} \Psi &= \vee \circ d(L_{g^{-1}})_g \circ d\varphi \\ \Phi &= \Psi^{-1} = (d\varphi)^{-1} \circ d(L_g)_e \circ \wedge \end{aligned}$$

Since $\mathfrak{g} \cong (\mathbb{R}^n, \diamond)$, one can think of the maps Ψ and Φ either with or without the elevator maps $\vee$ and $\wedge$, depending on whether one is using matrices or vectors to describe elements of the Lie algebra. In terms of the Boltzmann-Hamel formulation, the elevator maps are included. The chart in Fig. 7.3 also clarifies the distinction between the transformation $\mathbf{\Psi}$, discussed in §7.2, which takes the material velocity *matrix* to the *matrix* element of the Lie algebra, and the transformation Ψ which takes the *coordinate velocity* to the Lie algebra.

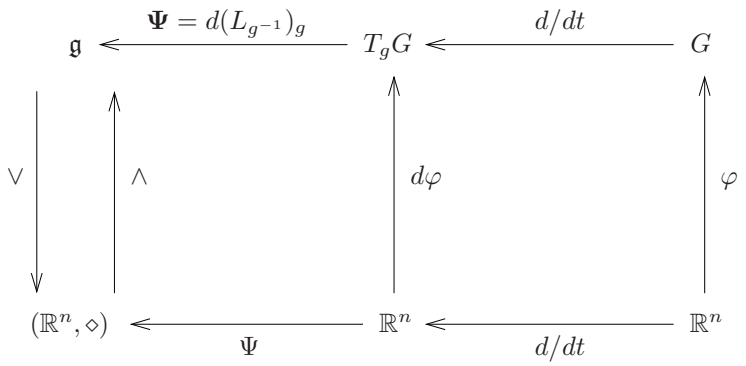

Figure 7.3: Coordinates of G

As we discussed in §7.1.3, a group element $g \in G$ can be written in terms of the real parameters $\{\theta^1, \ldots, \theta^N\}$ (known as *exponential coordinates*) as:

$$g = \varphi(\theta) = \exp(\theta^N E_N) \star \exp(\theta^{N-1} E_{N-1}) \star \cdots \star \exp(\theta^2 E_2) \star \exp(\theta^1 E_1) \in G$$

where the set of matrices $\{E_1, \ldots, E_N\}$ form a basis of the Lie algebra $\mathfrak{g}$. Note that

E_j commutes with $\exp(\theta^j E_j)$, so that

$$\frac{\partial \exp(\theta^j E_j)}{\partial \theta^j} = E_j \exp(\theta^j E_j) = \exp(\theta^j E_j) E_j$$

In computing the map Ψ we first need the differential $d\varphi : \mathbb{R}^n \to T_R G$ given by

$$d\varphi = \frac{\partial R}{\partial \theta^i} d\theta^i$$

where $d\theta^i \in T_R^* G$ form a basis of the cotangent space. Let us first compute:

$$\frac{\partial g}{\partial \theta^j} = \exp(\theta^N E_N) \star \cdots \star \exp(\theta^j E_j) \star E_j \star \exp(\theta^{j-1} E_{j-1}) \star \cdots \star \exp(\theta^1 E_1)$$

$$= g \star \exp(-\theta^1 E_1) \star \cdots \star \exp(-\theta^{j-1} E_{j-1}) \star E_j \star \exp(\theta^{j-1} E_{j-1}) \star \cdots \star \exp(\theta^1 E_1)$$

Defining

$$\Lambda_j = \exp(-\theta^1 E_1) \star \cdots \star \exp(-\theta^{j-1} E_{j-1}) \in G \tag{7.15}$$

one can rewrite this as:

$$\frac{\partial g}{\partial \theta^j} = g \star \Lambda_j \star E_j \star \Lambda_j^{-1} = g \star \mathrm{Ad}_{\Lambda_j} E_j$$

Since $\mathrm{Ad}_{\Lambda_j} E_j \in \mathfrak{g}$, there exists functions $\Psi_j^i(\theta) : G \to \mathbb{R}$ such that:

$$\mathrm{Ad}_{\Lambda_j} E_j = \Psi_j^i(\theta) E_i$$

(summation over i is understood). Finally we have:

$$\frac{\partial g}{\partial \theta^j} = \Psi_j^i(\theta) g E_i \in T_R G$$

(Note that indeed the set of matrices $\{gE_1, \ldots, gE_N\}$ is a basis for $T_g G$). The pullback of this vector to the Lie algebra by the left-translation map is therefore given by

$$d(L_{g^{-1}})_g \left(\frac{\partial g}{\partial \theta^j} \right) = \Psi_j^i(\theta) E_i \in \mathfrak{g}$$

It follows that the material and body velocities can be expressed in terms of coordinates as follows:

$$\dot{g} = \frac{\partial g}{\partial \theta^j}\dot{\theta}^j = \Psi^i_j(\theta)\dot{\theta}^j g E_i \in T_g G \qquad \text{and} \qquad g^{-1}\dot{g} = \Psi^i_j(\theta)\dot{\theta}^j E_i \in \mathfrak{g},$$

respectively. Similarly the material and body infinitesimal variations can be expressed in terms of coordinates as:

$$\delta g = \frac{\partial g}{\partial \theta^j}\delta\theta^j = \Psi^i_j(\theta)\delta\theta^j g E_i \in T_g G \qquad \text{and} \qquad g^{-1}\delta g = \Psi^i_j(\theta)\delta\theta^j E_i \in \mathfrak{g},$$

respectively. We therefore see that the matrix $\Psi(\theta)$ transforms the *coordinate representation* of the tangent space T_gG to the components of the corresponding quasi-vector in the Lie algebra with respect to a given basis $\{E_1, \ldots, E_N\}$ of the Lie algebra. We will work through an example of this in terms of rigid body mechanics in §7.6.3.

7.6.2 The Generalized Euler-Poincaré Equations

There is an important insight that can be learned by the comparison made in §7.5. Theorem 51 requires the Lagrangian to be *left invariant*. This means that $L_g^* L = L$. In other words, the pullback of $L(g, \dot{g})$ to the Lie algebra does not depend on g itself. The physical implication of this is that L must be independent of the state, and therefore must not contain any potential energy function. If the Lagrangian depends on g, we have the following generalization of the Euler-Poincaré equations:

Theorem 52. *Let G be a matrix Lie group with exponential coordinates $\{\theta^i\}_{i=1}^n$ with respect to the basis $\{E_i\}_{i=1}^n$ of the Lie algebra, $\varphi : \mathbb{R}^n \to G$ given by $\varphi(\theta) = g \in G$, $L(\theta, \dot{g})$ be a general Lagrangian, and $l(\theta, \xi)$ its restriction to the Lie algebra. Let $F \in T^*G$ be a force field equivariant relative to the canonical left actions of G on TG and T^*G, and let $f \in \mathfrak{g}^*$ be its restriction to the Lie algebra, obtained by $f = Fg$. Then the following are equivalent:*

1. *$g(t)$ satisfies the Euler-Lagrange equations with respect to the coordinates θ.*

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}^i} - \frac{\partial L}{\partial \theta^i} = F_i$$

2. *The Lagrange-d'Alembert principle holds:*

$$\delta \int_a^b L(\theta(t), \dot{\theta}(t))\ dt + \int_a^b F(\theta(t), \dot{\theta}(t)) \cdot \delta\theta\ dt = 0$$

holds for all proper variations $\delta\theta$.

3. *The generalized Euler-Poincaré equations hold:*

$$\frac{d}{dt}\frac{\partial l}{\partial \xi} - \frac{\partial l}{\partial \theta}\Phi - \mathrm{ad}^*_\xi \frac{\partial l}{\partial \xi} = f \tag{7.16}$$

where $\Phi : \mathfrak{g} \to \mathbb{R}^N$ is the inverse of the map $\Psi : \mathbb{R}^N \to \mathfrak{g}$ defined by $\Psi(u) = u^j \mathrm{Ad}_{\Lambda_j} E_j$ for all $u \in \mathbb{R}^N$, where $\Lambda_j \in G$ was defined in (7.15)*. Alternatively $\Phi = (d\phi)^{-1} \circ d(L_g)_e$.*

4. *The variational principle*

$$\delta \int_a^b l(\theta(t), \xi(t))\ dt + \int_a^b f(\xi(t)) \cdot \eta(t)\ dt = 0 \tag{7.17}$$

holds on $\mathfrak{g}$, using variations of the form $\delta\xi = \dot{\eta} + [\xi, \eta]$, where η vanish at the endpoints, and of the form $\delta\theta = \Phi \cdot \eta$.

Note that each term of the generalized Euler-Poincaré equation is an element of $\mathfrak{g}^*$. The second term in (7.16) makes sense, since $\Phi : \mathfrak{g} \to \mathbb{R}^N$. However $\partial l/\partial\theta \in (\mathbb{R}^N)^*$. Hence $\langle \partial l/\partial\theta, \Phi\rangle : \mathfrak{g} \to \mathbb{R}$ and is thus itself an element of the dual Lie algebra $\mathfrak{g}^*$.

Proof. Taking variations of (7.17) one obtains

$$\delta \int l(g, \xi)\ dt = \int \left(\frac{\partial l}{\partial \theta}\delta\theta + \frac{\partial l}{\partial \xi}\delta\xi \right) dt$$

With the second term you proceed as normal. With the first term you recognize $\delta\theta = \Phi\eta$, as was discussed in detail in the §7.6.1. In the end you correctly obtain (7.16). Notice that (7.16) are completely identical to the Boltzmann-Hamel equations (5.70). □

The Euler-Poincaré equations have the advantage that they are an elegant and geometrically based set of equations that exploit the symmetries of the system.

7.6.3 The Heavy Top

Consider the case of the heavy top, as seen in Fig. 7.4. The heavy top is an axially-symmetric rigid body with a fixed base point O, and is situated in a constant gravitational field with gravitational acceleration g. We will take the Boltzmann-

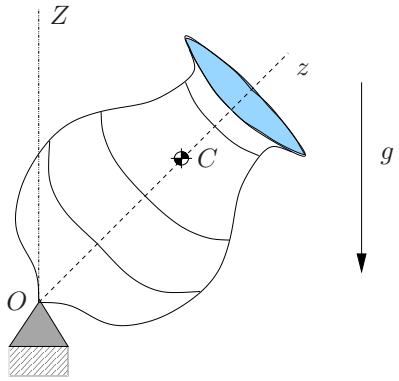

Figure 7.4: The Heavy Top.

Hamel approach, and use the Type-II Euler angles (ϕ, θ, ψ) as generalized coordinates of $SO(3)$. The corresponding transformation matrix $A \in SO(3)$ is given by (7.10). By inspection of this matrix, one sees that the body fixed z axis points in the direction $\langle \sin\theta \sin\psi, -\cos\theta\cos\psi, \cos\theta \rangle$ relative to the inertial frame. Hence the angle between the inertial Z and body fixed z axes is simply the Euler angle θ. Let d be the distance between the fixed point O and the body's center of mass C. Further suppose the body has mass m, axial moment of inertia I_a and transverse moment of inertia I_t.

The Lagrangian for the system is:

$$L = \frac{I_a}{2}\left[\dot{\theta}^2 + \dot{\psi}^2 \sin^2\theta\right] + \frac{I_t}{2}\left[\dot{\phi}^2 + 2\dot{\phi}\dot{\psi}\cos\theta + \dot{\psi}^2\cos^2\theta\right] - mgd\cos\theta$$

As was discussed above, the Lie algebra $\mathfrak{so}(3)$ is identified with $(\mathbb{R}^3, \times)$. Physically, $\boldsymbol{\omega} \in \mathfrak{so}(3)$ represents the angular velocity of the rigid body expressed in the body fixed frame. In terms of the body axis components of angular velocity, $\boldsymbol{\omega} = \langle \omega_x, \omega_y, \omega_z \rangle$, which constitute a set of quasi-velocities, the Lagrangian is:

$$l(\phi, \theta, \psi, \omega_x, \omega_y, \omega_z) = \frac{1}{2}\left(I_a\omega_x^2 + I_a\omega_y^2 + I_t\omega_z^2\right) - mgd\cos\theta \tag{7.18}$$

One can now use the Boltzmann-Hamel equations (5.70) to write out the dynamical equations of motion.

We will now follow the computation of §7.6.1 to determine the transformation matrix Ψ^i_j that takes the Euler angles to the body angular velocity. A rotation matrix $R \in SO(3)$ can be expressed in terms of Type-II Euler angles as given in (7.10). Defining the matrices

$$R_1 = \exp(\phi E_3), \qquad R_2 = \exp(\theta E_1), \qquad R_3 = \exp(\psi E_3)$$

the matrix R is equivalent to

$$R = R_3 R_2 R_1 = \exp(\psi E_3)\exp(\theta E_1)\exp(\phi E_3)$$

where $E_1,\ E_2,\ E_3$ are the basis vectors of $\mathfrak{so}(3)$ as defined in (7.11). From (7.15) we have $\Lambda_1 = \mathrm{diag}(1,1,1)$, $\Lambda_2 = R_1^{-1}$, and $\Lambda_3 = R_1^{-1}R_2^{-1}$ We therefore have:

$$\begin{aligned}
\frac{\partial R}{\partial \phi} &= RE_3 \\
\frac{\partial R}{\partial \theta} &= R_3R_2E_1R_1 = R\left(R_1^{-1}E_1R_1\right) = R\left(\mathrm{Ad}_{\Lambda_2}E_1\right) \\
\frac{\partial R}{\partial \psi} &= R_3E_3R_1R_1 = R\left(R_1^{-1}R_2^{-1}E_3R_2R_1\right) = R\left(\mathrm{Ad}_{\Lambda_3}E_3\right)
\end{aligned}$$

A direct calculation shows that

$$\begin{aligned}
\mathrm{Ad}_{\Lambda_2} E_1 &= \cos(\phi) E_1 - \sin(\phi) E_2 \\
\mathrm{Ad}_{\Lambda_3} E_3 &= \sin(\phi)\sin(\theta) E_1 + \cos(\phi)\sin(\theta) E_2 + \cos(\theta) E_3
\end{aligned}$$

Recall $T_R G = \mathrm{span}(RE_1, RE_2, RE_3)$. It follows that

$$\begin{aligned}
\frac{\partial R}{\partial \phi} &= RE_3 \\
\frac{\partial R}{\partial \theta} &= \cos(\phi) RE_1 - \sin(\phi) RE_2 \\
\frac{\partial R}{\partial \psi} &= \sin(\phi)\sin(\theta) RE_1 + \cos(\phi)\sin(\theta) RE_2 + \cos(\theta) RE_3
\end{aligned}$$

The material velocity is given in terms of Euler angle rates by

$$\dot{R} = \frac{\partial R}{\partial \phi}\dot{\phi} + \frac{\partial R}{\partial \theta}\dot{\theta} + \frac{\partial R}{\partial \psi}\dot{\psi}$$

The body-fixed components of the angular velocity are therefore

$$\begin{aligned}
\xi = R^{-1}\dot{R} &= \left[\cos(\phi)\dot{\theta} + \sin(\phi)\sin(\theta)\dot{\psi}\right] E_1 \\
&\quad + \left[-\sin(\phi)\dot{\theta} + \cos(\phi)\sin(\theta)\dot{\psi}\right] E_2 + \left[\dot{\phi} + \cos(\theta)\dot{\psi}\right] E_3 \\
\check{\xi} &= \langle \cos(\phi)\dot{\theta} + \sin(\phi)\sin(\theta)\dot{\psi}, -\sin(\phi)\dot{\theta} + \cos(\phi)\sin(\theta)\dot{\psi}, \dot{\phi} + \cos(\theta)\dot{\psi} \rangle
\end{aligned}$$

Notice $\xi \in \mathfrak{so}(3)$ and $\check{\xi} \in \mathbb{R}^3$. Similarly, the body-fixed components of the infinitesimal variation are given by:

$$\begin{aligned}
\eta = R^{-1}\delta R &= \left[\cos(\phi)\delta\theta + \sin(\phi)\sin(\theta)\delta\psi\right] E_1 \\
&\quad + \left[-\sin(\phi)\delta\theta + \cos(\phi)\sin(\theta)\delta\psi\right] E_2 + \left[\delta\phi + \cos(\theta)\delta\psi\right] E_3 \\
\check{\eta} &= \langle \cos(\phi)\delta\theta + \sin(\phi)\sin(\theta)\delta\psi, -\sin(\phi)\delta\theta + \cos(\phi)\sin(\theta)\delta\psi, \delta\phi + \cos(\theta)\delta\psi \rangle
\end{aligned}$$

The transformation matrices Ψ and Φ are therefore given by

$$\Psi = \begin{bmatrix} 0 & \cos(\phi) & \sin(\phi)\sin(\theta) \\ 0 & -\sin(\phi) & \cos(\phi)\sin(\theta) \\ 1 & 0 & \cos(\theta) \end{bmatrix}$$

$$\Phi = \begin{bmatrix} -\sin(\phi)\cot(\theta) & -\cos(\phi)\cot(\theta) & 1 \\ \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi)\csc(\theta) & \cos(\phi)\csc(\theta) & 0 \end{bmatrix}$$

Let $\alpha = \langle \phi, \theta, \psi \rangle^T$. This is why

$$\frac{\partial l}{\partial \alpha}\delta\alpha = \frac{\partial l}{\partial \alpha}\Phi \cdot \eta$$

in the derivation of (7.16).

Now consider the generalized Euler-Poincaré equations (7.16):

$$\frac{d}{dt}\frac{\partial l}{\partial \xi} - \frac{\partial l}{\partial \theta}\Phi - \mathrm{ad}^*_\xi \frac{\partial l}{\partial \xi} = f$$

We saw in §7.4.3 that, for $\mathfrak{g} = \mathfrak{so}(3)$:

$$\mathrm{ad}^*_A B = B \times A$$

We further have:

$$\frac{\partial l}{\partial \theta}\Phi = \langle mgd\sin\theta\cos\phi, -mgd\sin\theta\sin\phi, 0 \rangle$$

Notice this is simply the gravitational torque on the body relative to the body-fixed frame. The Euler-Poincaré equations then generate the following set of differential equations of motion for the heavy top:

$$\begin{aligned} I_a\dot{\omega}_x - mbd\sin\theta\cos\phi + (I_t - I_a)\omega_y\omega_z &= 0 \\ I_a\dot{\omega}_y + mgd\sin\theta\sin\phi + (I_a - I_t)\omega_x\omega_z &= 0 \\ I_t\dot{\omega}_z &= 0 \end{aligned}$$

One can then solve the following kinematic relations to obtain the Euler angles as a

function of time:

$$\begin{aligned}
\dot{\phi} &= -\sin\phi\cot\theta\omega_x - \cos\phi\cot\theta\omega_y + \omega_z \\
\dot{\theta} &= \cos\phi\omega_x - \sin\phi\omega_y \\
\dot{\psi} &= \sin\phi\csc\theta\omega_x + \cos\phi\csc(\theta)\omega_y
\end{aligned}$$

Alternatively, one could solve the set of differential equations for the rotation matrix instead of solving for the Euler angles:

$$\dot{R} = R\xi$$

where $\xi = \omega_x E_1 + \omega_y E_2 + \omega_z E_e \in \mathfrak{so}(3)$.

7.7 Optimal Control on Lie Groups

We continue our discussion of geometric methods for optimal control by discussing optimal control problems on Lie groups. Some related work is discussed in Bloch [16], Cortes et al. [35], Martinez et al. [87], and Koon and Marsden [71].

The dynamical equations of motion for systems that evolve on Lie groups are given by the Euler-Poincaré Dynamical Equations, which were introduced in §7.3. It turns out that these equations can be extended to the optimal control setting analogously to the way we extended the Boltzmann-Hamel equations. This generalization will be the topic of our present discussion.

We begin by stating a form of the optimal control problem applicable to Lie groups with acceleration controls.

Definition 53. *Let G be a matrix Lie group and $g(t) \in G$ a kinematic curve through the group. Then the dynamic optimal control problem for Lie Groups is the problem of finding the extremal curves for the cost functional*

$$\int_a^b C(g(t), \dot{g}(t), F(t))\ dt \tag{7.19}$$

C is left invariant cost integrand $C : TG \times \Omega \to \mathbb{R}$ and F is a control taken to lie in an admissible control set, $F(t) \in \Omega \subset T^*G$. The curve $g(t)$ is subject to fixed, predetermined endpoints and must further satisfy the Euler-Poincaré equations in Theorem 51:

$$\frac{d}{dt}\frac{\delta l}{\delta \xi} - \mathrm{ad}^*_\xi \frac{\delta l}{\delta \xi} = f \tag{7.20}$$

where $f = Fg$ is the pullback of the control to the dual of the Lie algebra, $f \in \mathfrak{g}^*$.

We recall from our discussion in §7.2 that the body velocity $\xi(t) = g(t)^{-1}\dot{g}(t)$, when expressed in terms of a basis $\{E_i\}$ of the Lie algebra, define a set of quasi-velocities for the system. Moreover, as we showed, the variational relation given in Lemma 49, $\delta\xi = \dot{\eta} + [\xi, \eta]$, is a Lie group equivalent of the transpositional relations. Additionally the Euler-Poincaré equations (7.20) are a special case of the Boltzmann-Hamel equations that apply to Lie groups. This point of view was discussed in §7.5. It is therefore reasonable to suggest the following Lemma, the original premise for which was motivated by Theorem 37, which states that the operators $\partial/\partial t$ and $\partial/\partial s$ commute when applied to quasi-velocities.

Lemma 54. *For the quasi-velocities $\xi = g(t)^{-1}\dot{g}(t)$, we have:*

$$\frac{\partial^2 \xi}{\partial \varepsilon \partial t} = \frac{\partial^2 \xi}{\partial t \partial \varepsilon}$$

i.e. $\delta\dot{\xi} = \frac{d}{dt}\delta\xi$.

Proof. A direct calculation shows:

$$\begin{aligned}
\frac{\partial \xi}{\partial t} &= -g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial g}{\partial t} + g^{-1}\frac{\partial^2 g}{\partial t^2} \\
\frac{\partial \xi}{\partial \varepsilon} &= -g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial g}{\partial t} + g^{-1}\frac{\partial^2 g}{\partial \varepsilon \partial t}
\end{aligned}$$

Thus, taking second derivatives:

$$\begin{aligned}
\frac{\partial^2 \xi}{\partial \varepsilon \partial t} &= g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial g}{\partial t} - g^{-1}\frac{\partial^2 g}{\partial \varepsilon \partial t}g^{-1}\frac{\partial g}{\partial t} \\
&\quad + g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial g}{\partial t} - g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial^2 g}{\partial \varepsilon \partial t} \\
&\quad - g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial g^2}{\partial t^2} + g^{-1}\frac{\partial^3 g}{\partial \varepsilon \partial t^2} \\
\frac{\partial^2 \xi}{\partial t \partial \varepsilon} &= g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial g}{\partial t} - g^{-1}\frac{\partial^2 g}{\partial t \partial \varepsilon}g^{-1}\frac{\partial g}{\partial t} \\
&\quad + g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial g}{\partial t} - g^{-1}\frac{\partial g}{\partial \varepsilon}g^{-1}\frac{\partial^2 g}{\partial t^2} \\
&\quad - g^{-1}\frac{\partial g}{\partial t}g^{-1}\frac{\partial g^2}{\partial \varepsilon \partial t} + g^{-1}\frac{\partial^3 g}{\partial t \partial \varepsilon \partial t}
\end{aligned}$$

A direct comparison of terms and smoothness of the variation $g(t, \varepsilon)$ yields the result.

□

We are now in a position to derive the Euler-Poincaré optimal control equations for Lie groups. We note that the derivation of the following proof folllows in parallel to the discussion on the Boltzmann-Hamel equations for optimal dynamic control presented in §6.5.1.

Theorem 55. *Let $\tilde{C}(\xi, f)$ be the restriction of the cost integrand to $\mathfrak{g}$ and let $c(\xi, \dot{\xi}) = \tilde{C}(\xi, f(\xi, \dot{\xi}))$ be the cost integrand with the controls eliminated by* (7.20)*. Define now*

$$\kappa = \frac{\delta c}{\delta \xi} - \frac{d}{dt}\frac{\delta c}{\delta \dot{\xi}} \tag{7.21}$$

Then the optimal control trajectory defined by Def. 53 satisfies the Euler-Poincaré Optimal Control Equations:

$$\dot{\kappa} = \mathrm{ad}^*_\xi \kappa \tag{7.22}$$

Furthermore, if $\tilde{C}$ is given by $\tilde{C} = \frac{1}{2} f \cdot f$, then

$$\kappa = f \cdot \frac{\delta f}{\delta \xi} - \frac{d}{dt}\left(f \cdot \frac{\delta f}{\delta \dot{\xi}} \right) \tag{7.23}$$

Proof. One can enforce the dynamics (7.20) by substituting the controls f directly into (7.19). Taking variations and utilizing Lemma 54 yields:

$$\begin{aligned}
\delta \int_a^b c(\xi, \dot{\xi})\ dt &= \int_a^b \left(\frac{\delta c}{\delta \xi} \delta \xi + \frac{\delta c}{\delta \dot{\xi}} \delta \dot{\xi} \right)\ dt \\
&= \int_a^b \left(\frac{\delta c}{\delta \xi} - \frac{d}{dt} \frac{\delta c}{\delta \dot{\xi}} \right) \delta \xi\ dt \\
&= \int_a^b \kappa(\dot{\eta} + \mathrm{ad}_\xi \eta)\ dt \\
&= \int_a^b \left(-\dot{\kappa} + \mathrm{ad}^*_\xi \kappa \right) \eta\ dt
\end{aligned}$$

Stationarity of the cost functional therefore holds iff the Euler-Poincaré optimal control equations (7.22) hold. □

In coordinates, the Euler-Poincaré optimal control equations are simply:

$$\dot{\kappa}_i = \gamma^j_{ki} \kappa_j \xi^k \tag{7.24}$$

where the γ^j_{ki} are the structure coefficients of the Lie algebra $\mathfrak{g}$. Comparing with the Boltzmann-Hamel optimal dynamic control equations (6.19), we again see that the Euler-Poincaré equations are a special case of the Boltzmann-Hamel equations. For example, consider the group $G = SO(3)$ and Lie algebra $\mathfrak{g} = \mathfrak{so}(3)$. Then the Euler-Poincaré optimal control equations (7.22) can be written:

$$\dot{\kappa} = \kappa \times \xi$$

where ξ is the body-fixed angular velocity. This equation agrees with (6.30). Also note that the definition of κ in our current discussion (7.21) is equivalent to the definition (6.18) found in our previous discussion on dynamical optimal control in Chapter VI. As one would expect, the Boltzmann-Hamel dynamical optimal control equations for the free rigid body, contained in §6.5.4 coincide with the Euler-Poincaré equations for dynamical optimal control on $SO(3)$. The Boltzmann-Hamel optimal

control equations have the advantage that they are written in an explicit coordinate formulation, they are applicable to non-left-invariant cost functions and they are applicable to configuration manifolds other than Lie groups. The Euler-Poincaré optimal control equations have the advantage of being a more geometric set of equations and they provide a generalization of the rigid body optimal control equations to arbitrary Lie groups other than $SO(3)$.

7.8 The Symplectic Group and the State Transition Matrix

In §2.2 we introduced some of the basic formalism behind Hamiltonian dynamical systems. Using the ordering convention $x = \langle p_1, q_1, \ldots, p_n, q_n \rangle$ (also used by Arnold [6]) and given a Hamiltonian H, the dynamical equations of motion for the state x and the State Transition Matrix (STM) Φ can be written:

$$\frac{dx}{dt} = J \cdot \frac{\partial H}{\partial x} \tag{7.25}$$

$$\frac{d\Phi}{dt} = J \cdot \frac{\partial^2 H}{\partial x^2} \cdot \Phi \tag{7.26}$$

where the matrix J is the $2n \times 2n$ block diagonal matrix with J_2's down the main diagonal, where J_2 is defined in (2.2). We saw how the preservation of the symplectic form ω and its various powers ω^k manifests itself in terms of constraints on the STM in §2.2-2.3. The equations of motion (7.25) represent a set of $2n$ coupled first order differential equations of motion to be integrated to obtain the solution flow as a function of time. One might naively view the STM dynamical equations (7.26) as an additional set of $4n^2$ equations of motion, and use standard methods to integrate these equations; this view, however, misses some of the important geometrical aspects of the dynamical equations (7.26) and could lead to a loss in the underlying symplectic structure of the flow, due to numerical error, in long term computations of the dynamics. It is the goal in this paragraph to discuss some of the Lie group

aspects of this flow and discuss a geometric integration scheme which preserves the underlying Lie group geometry implicitly present in these equations.

7.8.1 The Symplectic Group

The Symplectic Group is an $n(2n+1)$-dimensional matrix Lie group defined by:

$$Sp(2n,\mathbb{R}) = \{A \in GL(2n,\mathbb{R}) : A^T \cdot J \cdot A = J\}$$

The elements of $Sp(2n,\mathbb{R})$ are referred to as *symplectic matrices*. Given a curve $A(t) \in Sp(2n,\mathbb{R})$, $A(0) = \mathbb{I}$, one can differentiate the relation $A(t)^T \cdot J \cdot A(t) = J$ at $t = 0$ to determine the Lie algebra of $Sp(2n,\mathbb{R})$:

$$\mathfrak{sp}(2n,\mathbb{R}) = \{A \in GL(2n,\mathbb{R}) : J \cdot A + A^T \cdot J = 0\}$$

One can easily show that for any symmetric matrix B, the product $J \cdot B \in \mathfrak{sp}(2n,\mathbb{R})$. Define now the $n(2n+1)$ matrices Θ_{ij}, for $j \geq i$, as follows. The matrix Θ_{ij} is a $2n \times 2n$ matrix of zeros with a "1" in both the (i,j) and (j,i) position. It therefore follows that the $n(2n+1)$ matrices $E_{ij} = J \cdot \Theta_{ij}$, for $j \geq i$, form a basis of the Lie algebra $\mathfrak{sp}(2n,\mathbb{R})$.

For a Hamiltonian dynamical system, the State Transition Matrix Φ is an element of the symplectic group, $\Phi(t) \in Sp(2n,\mathbb{R})$. To see this, consider the matrix

$$\Omega(t) = J \cdot \frac{\partial^2 H}{\partial x^2} \in \mathfrak{sp}(2n,\mathbb{R}) \tag{7.27}$$

where H_{xx} is evaluated along the solution flow $x(t)$. Since H_{xx} is symmetric this matrix is an element of the Lie algebra $\mathfrak{sp}(2n,\mathbb{R})$. Notice also that, from (7.26), the matrix $\Omega(t) = \dot{\Phi}(t) \cdot \Phi(t)^{-1}$ is the *spatial velocity* of the STM. Therefore $\Phi(t)$ is just the flow through $Sp(2n,\mathbb{R})$ along the time-varying right-invariant vector field generated by $\Omega(t)$. Using the terminology of Def. 45 this fact can be expressed by

the relation $\Phi(t) = \phi(t, \mathbb{I}, \Omega_R(t))$. (Note that since $\Omega_R(t)$ is time-varying, this is *not* the exponential map).

One can see, just by counting, that if one integrates the $4n^2$ differential equations (7.26) without giving creed to the underlying geometric structure, something would be seriously missing. The underlying manifold upon which $\Phi(t)$ lives has a dimension of $n(2n+1)$. A standard integration scheme therefore utilizes $n(2n-1)$ *extra* differential equations of motion and moreover, even with higher order numerical methods, could cause $\Phi(t)$ to aberrate from the underlying manifold $Sp(2n, \mathbb{R})$ into $GL(2n, \mathbb{R})$, thereby destroying the underlying symplectic structure contained within $\Phi(t)$ which one might want to study. To remedy this we now introduce a geometric integrator that can be used in studying the dynamics of the STM.

7.8.2 A Kinematic Lie Group Integrator: Determining the STM

In recent years a number of numerical methods called variational integrators have arisen which preserve the symplectic structure of the solution flows, see Leimkuhler and Reich [77] and Marsden and West [86] for a comphrensive review of these methods. The earliest foreshadowing of a Lie group integrator goes back to the Magnus Series Expansion for the linear matrix differential equation $\dot{Y} = A(t)Y$. Magnus [82] showed that one could instead solve for a matrix valued function $\Omega(t)$ such that the solution would be given by $Y(t) = \exp(\Omega(t))Y_0$. Note that in terms of computing the State Transition Matrix, the function $\Omega(t)$ is given by $\Omega(t) = J \cdot H_{xx}$. Lie group integrators themselves aim at solving the differential equation $\dot{Y} = A(t, Y) \cdot Y$, where $Y \in G$ is now constrained to lie on the Lie group G for all time. Both Crouch and Grossman [39] and Munthe-Kaas [102, 103, 104] dealt with this problem on Lie groups. While Lie group integrators have primarily been used in studying rigid body mechanics on $SO(3)$, some recent work has applied them to the full body problem in

orbital mechanics, Lee et al. [74, 75]. Our interest in a Lie group based integrator is for the computation of the STM $\Phi(t)$ (7.26) for an arbitrary Hamiltonian system. The underlying configuration manifold is the noncompact Lie group $Sp(2n, \mathbb{R})$, and our main interest will be in preserving the symplecticity of $\Phi(t)$.

Let $X(g,t) \in T_g G$ be a time-varying vector field on the Lie group G. We wish to study the numerical solution of the differential equation

$$\frac{dg}{dt} = X(g,t), \qquad g(0) = g_0$$

The exact solution of this is denoted $g(t) = \phi(t; g_0; X)$. Suppose we'd like to approximate the solution on the interval $t \in [0, t_f]$ at the discrete points $\{t_i\}_{i=0}^N$, where $t_0 = 0$ and $t_N = t_f$. Let g_i be the approximation of $g(t_i)$. Suppose $g_1, \ldots, g_i$ are known and are all elements of the group G. We wish to approximate g_{i+1} while preserving the geometric fact that $g_{i+1} \in G$. Notice that, in general, we have:

$$g_{i+1}^{\text{candidate}} = g_i + (t_{i+1} - t_i) X(g_i; t_i) \notin G$$

In order to guarentee that $g_{i+1} \in G$ we invoke Theorem 47. In particular we can approximate g_{i+1} with either of the following:

$$\begin{aligned} g_{i+1}^L &= L_{g_i} \circ \exp\left(h_i \cdot d\left(L_{g_i^{-1}}\right)_{g_i} \cdot X(g_i; t_i) \right) \\ g_{i+1}^R &= R_{g_i} \circ \exp\left(h_i \cdot d\left(R_{g_i^{-1}}\right)_{g_i} \cdot X(g_i; t_i) \right) \end{aligned}$$

where $h_i = t_{i+1} - t_i$. In other words we pull the velocity $X(g_i; t_i) \in T_{g_i} G$ back to the Lie algebra $\mathfrak{g}$ by either the left- or right-translation map; map this body or spatial velocity vector from the Lie algebra onto the group by the exponential map, i.e. flow along the left-invariant or right-invariant vector field generated by the body or spatial velocity, respectively, for a short time h_i; and then map the resulting group element from its place in a neighborhood of e to its proper place in a neighborhood of g_i by

either the left- or right-translation map, respectively. For matrix Lie groups these relationships can be expressed as:

$$g_{i+1}^{L} = g_i \cdot \exp\left(h_i g_i^{-1} \cdot X(g_i; t_i)\right) \tag{7.28}$$

$$g_{i+1}^{R} = \exp\left(h_i X(g_i; t_i) \cdot g_i^{-1}\right) \cdot g_i \tag{7.29}$$

Since $g_i \in G$ and $X(g_i, t_i) \in T_{g_i}G$, this will preserve the underlying group structure of $g(t)$ by construction.

In determining the STM of a Hamiltonian phase flow, one can *first* solve the dynamical equations (7.25), using standard numerical methods (symplectic integrators included). One now has a discrete array of times and approximations of the true solution, $x_i \approx x(t_i)$. An approximation of the spatial velocity of the STM is now given at each of these discrete points from (7.27), $\Omega_i = J \cdot H_{xx}(x_i) \in \mathfrak{sp}(2n, \mathbb{R})$. Notice that once the approximate solution $\{x_i\}$ is given, the spatial velocity of the STM is given, *in advanced* of the numerical integration procedure to determine $\Phi(t)$, for all time. The initial condition is given $\Phi_0 = \mathbb{I}$. Given $\Phi_i \in Sp(2n, \mathbb{R})$, Φ_{i+1} is approximated by using (7.29):

$$\Phi_{i+1} = \exp\left(h_i \Omega_i\right) \cdot \Phi_i \tag{7.30}$$

where $h_i = t_{i+1} - t_i$. Alternatively, one can use the following:

$$\Phi_{i+1} = \exp\left(h_i \frac{\Omega_i + \Omega_{i+1}}{2}\right) \cdot \Phi_i \tag{7.31}$$

This integration scheme preserves the symplecticity of the solution $\{\Phi_i\}_{i=0}^{n}$ for all time, i.e. the approximation Φ_i does a much better job of remaining in $Sp(2n, \mathbb{R})$ without wandering off into $GL(2n, \mathbb{R})$. However, numerical error can still push Φ off of the underlying manifold $Sp(2n, \mathbb{R})$. The geometric integrator presented here is still superior than ordinary integrators as it is based on a system whose exact counterpart

does preserve symplecticity exactly. We will demonstrate a corrector method that will correct for some of the error in the following paragraph.

7.8.3 The Henon-Heiles System

We previously discussed the Henon-Heiles system in §2.5.5. The dynamics are determined by the Hamiltonian (2.19) and work out to be (2.20) for the state variables and (2.21) for the State Transition Matrix. The Lie algebra $\mathfrak{sp}(4,\mathbb{R})$ is 10-dimensional and is spanned by the basis vectors:

$$E_{11} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad E_{33} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad \begin{aligned} E_{22} &= -E_{11}^T \\ E_{44} &= -E_{33}^T \end{aligned}$$

$$E_{12} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad E_{13} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad E_{14} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$E_{23} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad E_{24} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad E_{34} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

With respect to this basis, the spatial velocity of the STM defined in (7.27) can be represented as:

$$\Omega = \sum_{i \leq j} H_{x^i x^j} E_{ij}$$

For the Henon-Heiles system this works out to be:

$$\Omega = E_{11} + (1+2y)E_{22} + E_{33} + (1-2y)E_{44} + 2xE_{24} \tag{7.32}$$

Each component of Ω is a *quasi-velocity* of the STM! Even though the STM lives in the 10-dimensional space $Sp(4,\mathbb{R})$, there are 5 *nonholonomic constraints* on its evolution. If $E^{ij} \in \mathfrak{sp}^*(4,\mathbb{R})$ form the dual basis to the Lie algebra, these 5 nonholonomic constraints can be written as:

$$E^{12}(\Omega) = 0, \qquad E^{13}(\Omega) = 0, \qquad E^{14}(\Omega) = 0, \qquad E^{23}(\Omega) = 0, \qquad E^{34}(\Omega) = 0$$

In §7.8.2 we discussed various approaches for integrating the equations of motion $\dot{\Phi} = \Omega \cdot \Phi$, with Ω given in (7.32). In order to study the effectiveness of these approaches in keeping the STM Φ on the manifold $Sp(4,\mathbb{R})$, we can define the following error measure:

$$\text{error}(t) = \text{norm}(\Phi(t)^T \cdot J \cdot \Phi(t) - J)$$

where $\text{norm}(A)$ is the matrix norm of A for any $A \in GL(4,\mathbb{R})$, i.e. it is the largest singular value of A. We plot $\text{error}(t)$ in Fig. 7.5. In each graph, the dynamical equations of motion (2.19) were solved using ode45. We then considered four different methods for computing $\Phi(t)$. The first method we integrated (2.21) using ode45 concurrently with the integration of the state variables. The error corresponding to this method is plotted with a black line. Next we used the method given by (7.30). The error is plotted in blue. We then used the method given by (7.31) and plotted the error in green. Finally we used (7.31) but then used a simple corrector method and plotted the error in red. We see that any of the methods discussed in §7.8.2 are superior to ode45 in the numerical integration of the STM equations. Using ode45, the STM rapidly looses symplecticity. The kinematic Lie group integrators each preserve the symplectic structure of the STM.

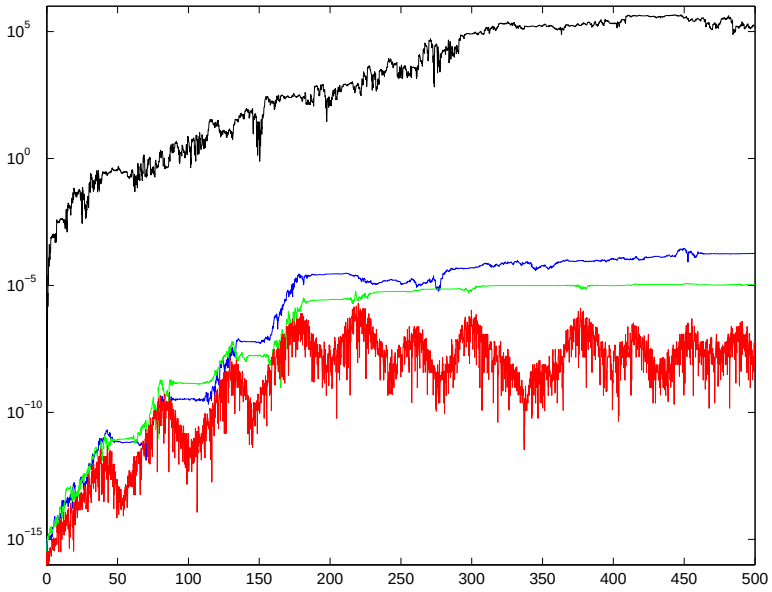

Figure 7.5: Symplectic error of the STM using different integration schemes

7.9 Conclusion

In this chapter we have explored mechanics and control systems on Lie groups. We presented the classical Euler-Poincaré equations and discussed the variational principles behind them, relating this to our discussion of quasi-velocity techniques from previous chapters. In particular we showed that the Euler-Poincaré equations can be thought of as a special case of the Boltzmann-Hamel equations. This provides additional geometric insight to the Boltzmann-Hamel equations when the underlying manifold is a Lie group and the quasi-velocities are the pullback of the velocity to the Lie algebra. We then discussed the use of coordinates on Lie groups, namely exponential coordinates, and generalized the Euler-Poincaré equations to systems with non-left-invariant Lagrangian functions. Finally we derived a set of higher order Euler-Poincaré equations that determine the optimal control dynamics on Lie groups.

CHAPTER VIII

Conclusion

In this thesis we've discussed geometric aspects of dynamical systems. This study is a graft of two main themes. The first is on the propagation of subvolumes in Hamiltonian systems. The second is on a precise geometric understanding of the manifestation of nonholonomic constraints in the variational principles of nonholonomic mechanical and control systems. In addition we discussed a new technique for describing the motion of a particle in a central force field. In this conclusion we will review the results that we have presented and indicate the directions in which this research can be continued in the future.

8.1 Subvolume Propagation

Chapters II and III discussed the theoretical aspects and practical applications of the dynamical propagation of subvolumes in Hamiltonian systems. In Chapter II we outlined some basic constraints regarding the evolution of even dimensional subvolumes in symplectic spaces. In Chapter III we took a practical turn and presented the results of our preliminary investigations of the applicatoin of a subvolume approach to the field of Space Situational Awareness (SSA). We will now discuss the main results of these chapters and future directions for this research.

Theoretical Aspects

We related the Lagrange and Poisson brackets to the standard symplectic two-form $\omega = dp \wedge dq$ and further illustrated a practical computational approach for its computation based on summing various subdeterminants of the State Transition Matrix (STM), Φ. Liouville's Theorem can be stated in terms of the STM as $\det \Phi = 1$. There is, however, a great deal more structure contained within the STM itself. We showed that the integral invariants of Poincaré-Cartan manifest themselves into the structure of the STM as well. This manifestation takes the form of the additional constraint on the STM that the sum of determinants of each symplectic 2×2 submatrix of each symplectic column must add to unity. The determinant of the 2×2 matrix found in the intersection of the λ-th symplectic column and τ-th symplectic row is the area expansion projection factor, of an area element initially parallel to the λ-th symplectic plane projected, in the future, to the τ-th symplectic plane. Moreover the Gram determinant of the λ-th column is the square of the area expansion factor of a 2-dimensional area element initially parallel to the λ-th symplectic plane. We then generalized these results to their $2k$-dimensional equivalents.

The integral invariants of Poincaré-Cartan are well known in the field of dynamical systems. This result can be stated as:

$$\frac{1}{k!}\int_{\Omega} \omega^k = \frac{1}{k!}\int_{\phi_t(\Omega)} \omega^k$$

where Ω is a $2k$-volume and ϕ_t is the Hamiltonian phase flow. It states that the sum of the oriented $2k$-volume projections on each symplectic "$2k$ plane" is conserved. We then presented a new class of integral invariants, which differ only subtly from those discussed above. The Wirtinger-type integral invariants can be written:

$$\frac{1}{k!}\int_{\Omega} |\omega^k| = \frac{1}{k!}\int_{\phi_t(\Omega)} |\omega^k|$$

Taking $k = 1$ for simplicity, we note that $|\omega| = |dp_1 \wedge dq_1 + \cdots + dp_n \wedge dq_n|$ is the *absolute value of the sum*, not the sum of the absolute values. Define:

$$\begin{aligned} \Lambda &= \{x \in \Omega \quad : \quad \omega^k(x) \geq 0\} \\ \Pi &= \{x \in \Omega \quad : \quad \omega^k(x) < 0\} \end{aligned}$$

so that Ω is the disjoint union $\Lambda \cup \Pi$. Then, explicitly:

$$\frac{1}{k!}\int_\Omega |\omega^k| = \frac{1}{k!}\left(\int_\Lambda \omega^k\right) - \frac{1}{k!}\left(\int_\Pi \omega^k\right) \tag{8.1}$$

$|\omega^k|$ is obviously an integral invariant, and has been up until now overlooked. The reason we discuss it here, is that it actually has a practical physical interpretation, distinctive from the interpretation of the integral invariants of Poincaré-Cartan. That interpretation is that this integral represents *the minimal obtainable* $2k$*-volume of the distribution.* In other words (symbols):

$$\frac{1}{k!}\int_\Omega |\omega^k| \leq \mathrm{Vol}_{2k}(\phi_t(\Omega)) \qquad \text{for all } t$$

From (8.1) we see that if ω is everywhere positive on the distribution, that the integral invariant of Poincaré-Cartan and Wirtinger are identical. In this case, the sum of the signed projections is equal to the minimal obtainable $2k$-volume of that distribution.

This theorem is related the the differential collapse of phase space in the following way. Consider a $2n$-differential "cube," partitioned into a direct sum $\Omega = \Upsilon + \Upsilon'$. Take Υ to be a $2k$-cube, parallel to k of the symplectic planes, and similarly with Υ' with k replaced by $2n - 2k$. Due to orthogonality:

$$\mathrm{Vol}_{2n}(\Omega) = \mathrm{Vol}_{2k}(\Upsilon) \cdot \mathrm{Vol}_{2n-2k}(\Upsilon')$$

Let the overbar operator be the application of the phase flow, so that $\overline{\Omega} = \phi_t(\Omega)$, etc. Liouville's Theorem states that the volume of Ω is the same as the volume of $\overline{\Omega}$.

Hence:

$$\mathrm{Vol}_{2k}(\Upsilon) \cdot \mathrm{Vol}_{2n-2k}(\Upsilon') = \mathrm{Vol}_{2n}(\Omega) = \mathrm{Vol}_{2n}(\overline{\Omega}) = \mathrm{Vol}_{2k}(\overline{\Upsilon}) \cdot \mathrm{Vol}_{2n-2k}(\overline{\Upsilon'}) \sin\beta$$

By our theorem, both Υ and Υ' can *only increase* in their $2k$- or $(2n-2k)$-volume. Such an increase, therefore, at a local level, must be accompanied by the collapse of the local phase space.

We then introduced the idea of a symplectic eigenskeleton. Despite this local collapse of the phase space, which is especially well known and a hallmark of chaotic systems, there always exists a symplectic basis that resists this collapse. For any fixed t, the symplectomorphism ϕ_t, for a pair of points $x, \phi_t(x)$, has with it associated a preferred basis $\{\xi_i, \eta_i\}$. These basis vectors occur in symplectic pairs, are orthogonal at T_xM, and their image vectors under the state transition matrix are again orthogonal $T_{\phi_t(x)}M$. This basis, which we named the symplectic eigenskeleton, exists for any canonical transformation; no matter how chaotic the system or how long the time.

Space Situational Awareness

Given a single track of data made by an optical sensor, one can determine the topocentric angles and angular rates at an epoch time, centered within the time interval of the track. A large uncertainty distribution exists, however, in the topocentric range and range-rate plane. This uncertainty distribution is therefore a two-dimensional manifold. This manifold can then be mapped into geocentric cartesian or spherical coordinates. In geocentric spherical coordinates, there are nonzero area projections onto each of the symplectic planes, including the area ones. One can then map this distribution into Delaunay space $\mathcal{D}$. Delaunay space has the advantage that it is symplectic. Additionally the time evolution becomes trivial. In fact, 5 of the 6 Delaunay variables are constants of motion.

We proposed a technique called Intersection Theory Analysis, which yields a nominal orbit determination that fits both tracks that utilizes a subvolume approach. The standard approach to this problem is to use a least squares routine. This approach is ill-suited for this problem, as there is no a priori nominal orbit that is known. It is therefore ad hoc and computationally intensive. Our approach reduces the problem of correlating previously uncorrelated tracks and making orbit determinations to that of systematically performing intersections between two separate laminas on two-dimensional planes.

Chapter III was both an introduction to this approach and a feasibility study performed with some toy observation data. The development of efficient computer algorithms to perform intersections of two-dimensional lamina will be a topic of future research. We will also look into metric approaches. It is possible that with the aid of the correct notion of distance in Delaunay space, one might be able to find this intersection directly by comparing distance between pairwise points from either representative uncertainty surface. Finally we will look into the automation of these techniques and efficient methods for incorporating them into the current SSA system.

8.2 The Central Force Field Problem

We continued with our second main application to astrodynamics in Chapter IV, where we introduced a new geometric technique for analyzing particle motion in central force fields. It is classically known that the particle's path is that of a rosette-shaped orbit. It is also known that the system is integrable and solveable by quadrature. We introduced the eccentric frame, a nonuniformly rotating reference frame that is based on the classical eccentricity vector (Runge-Lenz vector) of the two-body problem. With respect to this distinguished frame, the particle's path reduces

to a closed orbit or libration which can be described analytically. A single quadrature is required to determine the orientation of this frame with respect to time. We illustrated the theory with an analysis of the rossette-shaped orbits of particles in the Hernquist potential, a potential used to model spherically symmetric dark matter halos in galaxies. We discussed the motion in terms of the eccentric frame, and showed side-by-side comparisons between the actual motion and the motion relative to the eccentric frame.

The equatorial planar motion of a particle in an ellipsoidal force field (perhaps within an ellipsoidal galaxy) reduces to a central force field due to the symmetry of the problem. The existence of the eccentric frame then justifies the use of Floquet Theory for the stability analysis of planar motion. We then illustrated this technique for a toy mass distribution, based on an elongated version of the Hernquist mass density profile.

8.3 Nonholonomic Systems

In Chapters V, VI, and VII we discussed the geometry and variational principles of nonholonomic constraints in mechanical and control systems. Throughout we utilized a quasi-velocity approach in our understanding and formulating of these constraints. In Chapter VII this discussion was in the context of Lie groups.

Given a quasi-basis (co-moving frame), we introduced the notion of an associated connection. This connection is not deriveable from a metric and has non-vanishing torsion. It is defined by the property that

$$\tilde{\nabla}_{E_i} E_j = 0$$

where $\{E_i\}$ is the quasi-basis. In other words, given a quasi-basis, $\tilde{\nabla}$ is the unique connection with respect to which the quasi-basis is covariantly constant. As it turns

out, the Hamel coefficients (or structure coefficients of the co-moving frame) are simply the negative of the components of the torsion of $\tilde{\nabla}$ when viewed from the co-moving frame. We then show that the transpositional relations of nonholonomic mechanics are produced simply by application of the identity

$$\mathcal{L}_X Y = \tilde{\nabla}_X Y - \tilde{\nabla}_Y X - \tilde{T}(X, Y)$$

to the vector fields $X = \dot{q}$ and $Y = \delta q$. The transpositional relations,

$$\left(\frac{\partial \delta q^i}{\partial t} - \frac{\partial \dot{q}^i}{\partial s}\right) \Psi_i^j = \left(\frac{\partial \zeta^j}{\partial t} - \frac{\partial u^j}{\partial s}\right) + \gamma_{pq}^j u^p \zeta^q,$$

show that one cannot have both

$$\left(\frac{\partial \delta q^i}{\partial t} - \frac{\partial \dot{q}^i}{\partial s}\right) = 0 \qquad \text{and} \qquad \left(\frac{\partial \zeta^j}{\partial t} - \frac{\partial u^j}{\partial s}\right) = 0$$

One must choose one or the other. This gives rise to the transitivity choice, which is related to how one *defines* variations consistent with the constraints. If one defines the variations to be continuous (T1), then one must choose between applying the constraints to the infinitesimal variations (Principle of Virtual Work) or to the varied paths (Vakonomic Principle). Applying the constraints to both violates the transpositional relations. On the other hand, if one defines a set of m *dependent* velocity variables, which are functions of the remaining $n - m$ *independent* velocity variables and the n generalized coordinates, one only has closure in the independent or base directions, but nonclosure in the fiber or dependent directions (T2). The constraints can then be applied to the full differential non-closed quadrilaterals, both to the infinitesimal variations and to the varied paths. This approach is also known as Suslov's Principle.

After our discussion on the transpositional relations and the transitivity choice, we discussed the difference between the mechanical and vakonomic equations of motion in

relation to (T1). We extended Maggi's equation and the Boltzmann-Hamel equations to the vakonomic case, by applying the Vakonomic Principle instead of the Principle of Virtual Work. We further show how the vakonomic motion can be achieved physically by application of a set of external gyroscopic forces. Physical advantages of actually applying such gyroscopic forces will be the topic of future research.

In Chapter VI we generalized the Boltzmann-Hamel equations and the Euler-Poincaré equations to the optimal control setting, considering both the case of kinematic and dynamic optimal control. In particular, the Boltzmann-Hamel equations for dynamical optimal control make use of the *second transpositional relations*, which state that the operators d and δ commute when applied to the quasi-velocities themselves. The dynamical optimal control Boltzmann-Hamel equations are a fourth order version of the classical Boltzmann-Hamel equations for nonholonomic mechanics. By using quasi-velocities, one need no longer solve for the m quasi-velocities, the m *quasi-accelerations* and the m *quasi-jerks*. By taking a quasi-velocity approach to optimal dynamical control problems, we showed one requires a minimal set of $4n - 2m$ first order differential equations of motion, $3m$ fewer equations than required by standard techniques. This saving could offer a great advantage for high-dimensional systems with a large number of nonholonomic constraints.

In Chapter VII we discussed some geometric aspects of mechanics and control on Lie groups. We showed that the Euler-Poincaré equations are simply a special case of the Boltzmann-Hamel equations suited for matrix Lie groups. The Hamel coefficients are related to the Lie algebra structure coefficients, and the quasi-velocities are the components of the pullback of the material velocity to the Lie algebra under the left-translation map. We then generalize the Euler-Poincaré equations for the case when the Lagrangian function is not left-invariant. This was accomplished by defining an

appropriate set of coordinates on the group, a generalization of the Euler angles in rigid body mechanics. We related our discussion to the corresponding rigid body equations throughout the chapter, including an analysis of the heavy top. Finally we generalized the Euler-Poincaré equations to a higher order version, in a similar manner to the generalization of the Boltzmann-Hamel equations discussed in Chapter VI, to a form suitable in determining optimal control trajectories on Lie groups.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] R. Abraham and J.E. Marsden. *Foundations of Mechanics*. Addison-Wesley, 2nd edition, 1978.

[2] F.C. Adams and A.M. Bloch. Orbits in extended mass distributions: general results and the spirographic approximation. *The Astrophysical Journal*, 629:204–218, 2005.

[3] A. Agrachev and Y. Sachkov. *Control Theory from the Geometric Viewpoint*. Springer, 2004.

[4] V.I. Arnold. *Dynamical Systems III: Mathematical Aspects of Classical and Celestial Mechanics*. Springer, 1988.

[5] V.I. Arnold. *Ergodic Problems of Classical Mechanics*. Addison-Wesley, 1989.

[6] V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer, 1989.

[7] M. Audin, A.C. da Silva, and E. Lerman. *Symplectic Geometry of Integrable Hamiltonian Systems*. Birkauser, 2003.

[8] N. Bedrossian and S. Bhatt. Space station zero propellant maneuver flight results compared to eigenaxis. AAS-08-108, 2008.

[9] N. Bedrossian, S. Bhatt, M. Lammers, L. Nguyen, and Y. Zhang. First ever flight demonstration of zero propellant maneuver attitude control concept. In

Proceedings of the 2007 Guidance, Navigation and Control Conference, 2007. AIAA 2007-6734.

[10] G. Benettin, L. Galgani, A. Giorgilli, and J.M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, 15(1):9–20, 1980.

[11] G. Benettin, L. Galgani, A. Giorgilli, and J.M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 2: Numerical application. *Meccanica*, 15(1):21–30, 1980.

[12] G. Bennetin, L. Galgani, and J.M. Strelcyn. Kolmogorov entropy and numerical experiments. *Physical Review A*, 14(6):2338–2345, 1976.

[13] J. Betts. *Practical Methods for Optimal Control Using Nonlinear Programming*. SIAM, 2001.

[14] S.A. Bhatt. Optimal reorientation of spacecraft using only control moment gyroscopes. Master's thesis, Rice University, 2007.

[15] J. Binney and S. Tremaine. *Galactic Dynamics*. Princeton University Press, 1988.

[16] A.M. Bloch. *Nonholonomic Mechanics and Control*. Springer, 2003.

[17] A.M. Bloch and P.E. Crouch. Nonholonomic and vakonomic control systems on riemannian manifolds. In M.J. Enos, editor, *Dynamics and Control of Mechanical Systems*, pages 25–52. Fields Inst. Comm 1, AMS, 1993.

[18] A.M. Bloch and P.E. Crouch. Reduction of euler lagrange problems for constrained variational problems and relation with optimal control problems. In *The Proceedings of the 33rd IEEE Conference on Decision and Control*, pages 2584–2590, 1994.

[19] A.M. Bloch and P.E. Crouch. Nonholonomic control systems on riemannian manifolds. *SIAM Journal on Control*, 33(1):126–148, 1995.

[20] A.M. Bloch, P.S. Kirshnaprasad, J.E. Marsden, and T.S. Ratiu. Nonholonomic mechanical systems with symmetry. *Arch. Rat. Mech. Anal*, 136:21–99, 1996.

[21] A.M. Bloch, P.S. Krishnaprasad, J.E. Marsden, and T.S. Ratiu. The euler-poincare equations and double bracket dissipation. *Communications in Mathematical Physics*, 175(2):1–42, 1996.

[22] M. Born. *The Mechanics of the Atom.* G. Bells and Sons, Ltd, 1927.

[23] M. Brin and G. Stuck. *Introduction to Dynamical Systems.* Cambridge University Press, 2002.

[24] R.W. Brockett. *Control Theory and Singular Riemannian Geometry*, pages 11–27. Springer-Verlag, 1981.

[25] D. Brouwer and G.M. Clemence. *Methods of Celestial Mechanics.* Academic, 1961.

[26] J.L. Brun and A.F. Pacheco. On closed but non-geometrically similar orbits. *The Journal of Celestial Mechanics and Dynamical Astronomy*, 96:311–316, 2006.

[27] F. Bullo and A.D. Lewis. *Geometric Control of Mechanical Systems.* Springer, 2005.

[28] S.L. Campbell and C.D. Meyer Jr. *Generalized Inverses of Linear Transformations*. Dover, 1991.

[29] F. Cardin and M. Favretti. On nonholonomic and vakonomic dynamics of mechanical systems with nonintegrable constraints. *Journal of Geometry and Physics*, 18(4):295–325, 1996.

[30] M.P. Do Carmo. *Riemannian Geometry*. Birkhauser, 1992.

[31] A. Celletti and L. Chierchia. Kam tori for n-body problems: a brief history. *The Journal of Celestial Mechanics and Dynamical Astronomy*, 95:117–139, 2006.

[32] L. Cesari. *Asymptotic Behavior and Stability Problems in Ordinary Differential Equations*. Springer, 1963.

[33] D.E. Chang and J.E. Marsden. Geometric derivation of the delaunay variables and geometric phases. *Celestial Mechanics and Dynamical Astronomy*, 86(2):185–208, 2003.

[34] J. Cortes, M. de Leon, and D.M. de Diego. Geometric description of vakonomic and nonholonomic dynamics. comparison of solutions. *SIAM Journal of Control and Optimization*, 41(5):1389–1412, 2003.

[35] J. Cortes, S. Martinez, J.P. Ostrowski, and K.A. McIsaac. Optimal gaits for dynamic robotic locomotion. *The International Journal of Robotics Research*, 20(9):707–728, 2001.

[36] S. Craig, F. Diacu, E.A. Lacomba, and E. Perez. On the anisotropic manev problem. *Journal of Mathematical Physics*, 40:1359–1375, 1999.

[37] J. L. Crassidis and J.L. Junkins. *Optimal Estimation of Dynamic Systems*. Chapman & Hall/CRC, 2004.

[38] P. Crouch and F. Silva Leite. The dynamic interpolation problem: On riemannian manifolds, lie groups, and symmetric spaces. *Journal of Dynamical and Control Systems*, 1(2):177–202, 1995.

[39] P.E. Crouch and R. Grossman. Numerical integration of ordinary differential equations on manifolds. *J. Nonlinear Sci.*, 3:1–33, 1993.

[40] J.M.A. Danby. *Fundamentals of Celestial Mechanics*. Willmann-Bell, 1988.

[41] M. de Leon, J.C. Marrero, and D.M. de Diego. Vakonomic mechanics versus nonholonomic mechanics: a unified geometrical approach. *Journal of Geometry and Physics*, 35(2):126–144, 2000.

[42] C. Delaunay. Theorie due mouvement de la lune. In *Memoires de l'Academie des Sciences*, volume 28. 1860.

[43] C. Delaunay. Theorie due mouvement de la lune. In *Memoires de l'Academie des Sciences*, volume 29. 1867.

[44] F. Diacu and M. Santoprete. Nonintegrability and chaos in the anisotropic manev problem. *Physica D*, 156:39–52, 2001.

[45] L. Dieci, R.D. Russell, and E.S. Van Vleck. On the compuation of lyapunov exponents for continuous dynamical systems. *SIAM Journal on Numerical Analysis*, 34(1):402–423, 1997.

[46] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov. *Modern Geometry–Methods and Applications, Part I.* Springer, 1992.

[47] K. Ehlers, J. Koiller, R. Montgomery, and P.M. Rios. Nonholonomic systems via moving frames: Cartan equivalence and chaplygin hamiltonization. In J.E.

Marsden and T.S. Ratiu, editors, *The Breadth of Symplectic and Poisson Geometry*, pages 75–120. Birkhauser, 2005.

[48] M. Favretti. Equivalence of dynamics for nonholonomic systems with transverse constraints. *Journal of Dynamical and Differential Equations*, 10(4):511–536, 1998.

[49] O. Fernandez and A.M. Bloch. Equivalence of the dynamics of nonholonomic and variational nonholonomic systems for certain initial data. preprint, 2008.

[50] S. Ferraz-Mello. *Canonical Perturbation Theories: Degenerate Systems and Resonance*. Springer, 2007.

[51] C. Froeschle and Ch. Froeschle. Generalized lyapunov characteristic indicators and corresponding kolmogorov like entropy of the standard mapping. *Celestial Mechanics and Dynamical Astronomy*, 56(2):307–314, 1993.

[52] C. Froeschle, E. Lega, and R. Gonczi. Fast lyapunov indicators. application to asteroidal motion. *Celestial Mechanics and Dynamical Astronomy*, 67(1):41–62, 1997.

[53] I.M. Gelfand and S.V. Fomin. *Calculus of Variations*. Dover, 2000.

[54] H. Goldstein and C. Poole. *Classical Mechanics*. Addison Wesley, 2001.

[55] X. Gracia, J.Marin-Solano, and M.C. Munoz-Lecanda. Some geometric aspects of variational calculus in constrained systems. *Reports on Mathematical Physics*, 51(1):127–148, 2003.

[56] D.T. Greenwood. *Classical Dynamics*. Dover, 1977.

[57] D.T. Greenwood. *Principles of Dynamics*. Prentice Hall, 1987.

[58] D.T. Greenwood. *Advanced Dynamics*. Cambridge University Press, 2003.

[59] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*. Springer, 2002.

[60] G. Hamel. Die lagrange-eulersche gleichungen der mechanik. *Zeitschrift fuer Mathematik und Physik*, 50:1–57, 1904.

[61] G. Hamel. Ueber die virtuellen verschiebungen in der mechanik. *Mathematische Annalen*, 59:416–434, 1904.

[62] G. Hamel. *Theoretische Mechankik. Eine Einheitliche Einfuehrung in die gesamt Mechanik*, volume 57 of *Die Grundlehren der math. Wissenschaften*. Springer, 1949.

[63] T.W. Hamilton and W.G. Melbourne. Information content of a single pass of doppler data from a distant spacecraft. *JPL Space Programs Summary*, 3(37):18–23, 1966.

[64] L. Hernquist. An analytical model for spherical galaxies and bulges. *The Astrophysical Journal*, 356:359–364, 1990.

[65] O. Hoelder. Ueber die principien von hamilton und maupertuis. *Nachrichten von der Koenigl. Gesellschaft der Wissenschaften zu Goettingen, Math-Phys. Kl*, 2:122–157, 1896.

[66] H. Hofer and E. Zehnder. *Symplectic Invariants and Hamiltonian Dynamics*. Birkhauser, 1994.

[67] F.Y. Hsiao and D.J. Scheeres. Fundamental constraints on uncertainty evolution in hamiltonian systems. Accepted for publication in IEEE Transactions on Automatic Control.

[68] T.R. Kane. Dynamics of nonholonomic systems. *J. Appl. Mech*, 28:574–578, 1961.

[69] W. Kang and N. Bedrossian. Pseudospectral optimal control theory makes debut flight, saves nasa $1m in under three hours. *SIAM News*, 40(7), 2007.

[70] D.E. Kirk. *Optimal Control Theory*. Dover, 1998.

[71] W.S. Koon and J.E. Marsden. Optimal control for holonomic and nonholonomic mechanical systems with symmetry and lagrangian reduction. *SIAM Journal on Control and Optimization*, 35(3):901–929, 1997.

[72] C. Lanczos. *The Variational Principles of Mechanics*. Dover, 1970.

[73] L.D. Landau and E.M. Lifshitz. *The Classical Theory of Fields*. Elsevier Butterworth Heinemann, 1975.

[74] T. Lee, M. Leok, and N.H. McClamroch. Lie group variational integrators for the full body problem. *Comput. Methods Appl. Mech. Engrg.*

[75] T. Lee, M. Leok, and N.H. McClamroch. Lie group variational integrators for the full body problem in orbital mechanics. *The Journal of Celestial Mechanics and Dynamical Astronomy*, 98:121–144, 2007.

[76] J. Lei and M. Santorprete. Rosette central configurations, degenerate central configurations and bifurcations. *The Journal of Celestial Mechanics and Dynamical Astronomy*, 94:271–287, 2006.

[77] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge University Text, 2004.

[78] T. Levi-Civita and U. Amaldi. *Kurs teoreticheskoi mekhaniki (Course of Theoretical Mechanics)*, volume 2. IIL, 1951.

[79] A.D. Lewis and R.M. Murray. Variational principles for constrained systems: Theory and experiment. *Int. J. Non-Linear Mechanics*, 30(6):793–815, 1995.

[80] A.J. Lichtenberg and M.A. Lieberman. *Regular and Chaotic Dynamics*. Springer, 1992.

[81] G.A. Maggi. Di alcune nuove forme delle equazioni della dinamica applicabili ai sistemi anolonomi. *Atti Accad. Naz. Lincei Rend Cl. Fis. Mat. Nat.*, 10:287–291, 1901.

[82] W. Magnus. On the exponential solution of differential equations for a linear operator. *Comm. Pure Apply. Math.*, 7:649–673, 1954.

[83] J.E. Marsden and T.S. Ratiu. *Introduction to Mechanics and Symmetry*. Springer, 1994.

[84] J.E. Marsden, T.S. Ratiu, and A. Weinstein. Reduction and hamiltonian structures on duals of semidirect product lie algebras. *Cont. Math.*, 28:55–100, 1984.

[85] J.E. Marsden, T.S. Ratiu, and A. Weinstein. Semi-direct products and reduction in mechanics. *Trans. Amer. Math. Soc.*, 281:147–177, 1984.

[86] J.E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numerica*, 10:357–514, 2001.

[87] S. Martinez, J. Cortes, and M. de Leon. Symmetries in vakonomic dynamics: applications to optimal control. *Journal of Geometry and Physics*, 38.

[88] J.M. Maruskin and A.M. Bloch. The boltzmann-hamel equations for the optimal control of mechanical systems with nonholonomic constraints. Accepted for publication in Systems and Control Letters, 2007.

[89] J.M. Maruskin and A.M. Bloch. The boltzmann-hamel equations optimal control. In *The Proceedings of the 46th IEEE Conference on Decision and Control*, pages 554–559, 2007.

[90] J.M. Maruskin, A.M. Bloch, J.E. Marsden, and D. Zenkov. A geometric foundation of quasi-velocities in nonholonomic and vakonomic dynamics. preprint, 2008.

[91] J.M. Maruskin, D.J. Scheeres, F.C. Adams, and A.M. Bloch. The eccentric frame decomposition for central force fields. *Journal of Celestial Mechanics and Dynamical Astronomy*, 100(1):43–62, 2008.

[92] J.M. Maruskin, D.J. Scheeres, and K.T. Alfriend. Correlation of optical observations of objects in earth orbit. In *The Proceedings of the Seventh US/Russian Space Surveillance Workshop*, 2007.

[93] J.M. Maruskin, D.J. Scheeres, and K.T. Alfriend. Correlation of optical observations of objects in earth orbit. Submitted to the Journal of Guidance, Control, and Dynamics, 2008.

[94] J.M. Maruskin, D.J. Scheeres, and A.M. Bloch. Dynamics of symplectic subvolumes. In *The Conference Proceedings of the 46th IEEE Conference on Decision and Control*, pages 5600–5605, 2007.

[95] J.M. Maruskin, D.J. Scheeres, and A.M. Bloch. On the tracking of space debris. In *The Proceedings of the 2007 AAIA/AAS Astrodynamics Specialist Conference*, 2007.

[96] J.M. Maruskin, D.J. Scheeres, and A.M. Bloch. Dynamics of symplectic sub-

volumes. Accepted for publication in the SIAM Journal of Applied Dynamical Systems, 2008.

[97] D.Q. Mayne, S.V. Rakobic, R. Findelsen, and F. Allgower. Robust output feedback model predictive control of constrained linear systems. *Automatica*, 42:1217–1222, 2006.

[98] D. McDuff and D. Salamon. *Introduction to Symplectic Topology*. Oxford University Press, 1998.

[99] A. Milani, G.F. Gronchi, M. Vitturi, and Z. Knezevic. Orbit determination with very short arcs. i admissable regions. *Celestial Mechanics and Dynamical Astronomy*, 90:59–87, 2004.

[100] O. Montenbruck and E. Gill. *Satellite Orbits: Models, Methods, Applications*. Springer, 2000.

[101] R. Montgomery. *A Tour of SubRiemannian Geometry, Their Geodesics and Applications*. American Mathematical Society, 2002.

[102] H. Munthe-Kaas. Lie butcher theory for runge-kutta methods. *BIT*, 35:572–587, 1995.

[103] H. Munthe-Kaas. Runge-kutta methods on lie groups. *BIT*, 38:92–111, 1998.

[104] H. Munthe-Kaas. High order runge-kutta methods on manifolds. *J. Apply. Num. Maths.*, 29:115–127, 1999.

[105] J.I. Neimark and N.A. Fufaev. Transpositional relations in the analytical mechanics of nonholonomic systems. *J. Appl. Math. Mech.*, 24:1541–1548, 1960.

[106] J.I. Neimark and N.A. Fufaev. *Dynamics of Nonholonomic Systems*. Number 33 in Translations of Mathematical Monographs. American Mathematical Society, 1972.

[107] H. Nijmeijer and A.J. van der Schaft. *Nonlinear Dynamical Control Systems*. Springer, 1990.

[108] L. Noakes. Null cubics and lie quadratics. *Journal of Mathematical Physics*, 44:1436–1448, 2003.

[109] L. Noakes. Non-null lie quadratics in e^3. *Journal of Mathematical Physics*, 45:4334–4351, 2004.

[110] L. Noakes, G. Heinzinger, and B. Paden. Cubic splines on curved spaces. *IMA Journal of Mathematical Control and Information*, 6:465–473, 1989.

[111] V.I. Oseledec. A multiplicative ergodic theorem. ljapunov characteristic numbers for dynamical systems. *Transactions of Moscow Mathematical Society*, 19(2):197–231, 1968.

[112] R. Osserman. *A Survery of Minimal Surfaces*. Dover, 1986.

[113] J.G. Papastavridis. *Analytical Mechanics*. Oxford University Press, 2002.

[114] W. Pauli. *The Theory of Relativity*. Dover, 1958.

[115] M.Y. Poon and D. Merritt. A self-consistent study of triaxial black hole nuclei. *The Astrophysical Journal*, 606:774–787, 2004.

[116] A. Rossi. The earth orbiting space debris. *Serb. Astron.*, 170:1–12, 2005.

[117] A. Rossi. Population models of space debris. In A. Milani and Z. Knezevic,

editors, *Dynamics of Population of Planetary Systems*, volume 197, pages 427–438, 2005. Proceedings of IAU Coll.

[118] A.E. Roy. *Orbital Motion*. Institute of Physics, 1988.

[119] V.V. Rumiantsev. On hamilton's principle for nonholonomic systems. *PMM*, 42:407–419, 1978.

[120] H. Sagan. *Introduction to the Calculus of Variations*. Dover, 1969.

[121] P.A. Samuelson. Law of conservation of the capital-output ratio. *Appl. Math. Sci.*, 67:1477–1479, 1970.

[122] R. Sato and R.V. Ramachandran. *Conservation laws and symmetry: applications to economics and finance*. Kluwer Academic Publishers, 1990.

[123] D.J. Scheeres, F.Y. Hsiao, R.S. Park, B.F. Villac, and J.M. Maruskin. Fundamental limits on spacecraft orbit uncertainty and distribution propogation. *The Journal of the Astronautical Sciences*, 54(4):505–523, 2006.

[124] P. Seitzer. Modest observations of space debris at geosynchronous orbit. *Advances in Space Research*, 34:1139–1142, 2004.

[125] A. Shapere and F. Wilczek. Geometry of self-propulsion at low reynolds number. *J. Fluid Mech.*, 198:557–585, 1989.

[126] A. Silva. *Lectures on Symplectic Topology*. Springer, 2001.

[127] D.R. Smith. *Variational Methods in Optimization*. Dover, 1974.

[128] G.K. Suslov. On a particular variant of d'alembert's principle. *Matematisheskii Svornki*, 22:687–691, 1901.

[129] G.K. Suslov. *Theoretical Mechanics*. Moscow and Leningrad: Technical/Theoretical Literature (Gostehizdat), 1946.

[130] G. Tommei, A. Milani, and A. Rossi. Orbit determination of space debris: Admissible regions. *Celestial Mechanics and Dynamical Astronomy*, 97:289–304, 2007.

[131] P. Voronets. On the equations of motion for nonholonomic systems. *Matematicheskii Svornki*, 22:659–686, 1901.

[132] F.W. Warner. *Foundations of Differential Manifolds and Lie Groups*. Number 94 in Graduate Texts in Mathematics. Springer-Verlag, 2nd edition, 1989.

[133] E.T. Whittaker. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*. Cambridge University Press, 1988.

ABSTRACT

On the Dynamical Propagation of Subvolumes and on the Geometry and Variational Principles of Nonholonomic Systems

by

Jared Michael Maruskin

Co-Chairs: Anthony M. Bloch and Daniel J. Scheeres

Their are two main themes of this thesis. The first is the theory and application of the propagation of subvolumes in dynamical systems. We discuss the integral invariants of Poincaré-Cartan and introduce a new and closely related set of integral invariants, those of Wirtinger type, and relate these new invariants to a minimum obtainable symplectic volume. We will then consider the application of this approach to the orbit determination and correlation problem for tracking particles of space debris. The second theme is on the geometry of nonholonomic systems. In particular we will focus on the precise geometric understanding of quasi-velocity techniques and its relation to the formulation of variational principles for these systems. We will relate the Euler-Poincaré equations for Lie groups to the Boltzmann-Hamel equations and further extend both these equations to a higher order form that is applicable to optimal dynamical control problems on manifolds.