THE HAMILTON-JACOBI THEORY FOR SOLVING TWO-POINT BOUNDARY VALUE PROBLEMS: THEORY AND NUMERICS WITH APPLICATION TO SPACECRAFT FORMATION FLIGHT, OPTIMAL CONTROL AND THE STUDY OF PHASE SPACE STRUCTURE

by

Vincent M. Guibout

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Aerospace Engineering)
in The University of Michigan
2004

Doctoral Committee:

Associate Professor Daniel J. Scheeres, Co-Chair
Professor Anthony M. Bloch, Co-Chair
Professor Emeritus Donald T. Greenwood
Professor Pierre T. Kabamba
Professor N. Harris McClamroch

In memory of Olivier Guibout

ACKNOWLEDGEMENTS

I owe a debt of gratitude to many people whose help has been crucial to my success in completing this dissertation. First of all, I have been privileged to have the direction and guidance of two excellent advisors, my co-chairs Daniel J. Scheeres and Anthony M. Bloch. From the inception of this research project, Daniel Scheeres has been generous with his time and has provided eminently helpful advice. His enthusiasm, support, and invaluable insight have sparked my interest in spacecraft formation flight and Hamiltonian dynamics. Anthony Bloch has likewise made invaluable contributions to this research project. His unfailing guidance, wise counsel and encouragements have significantly contributed to my interest in the Hamilton-Jacobi theory and in discrete geometry. I am also deeply grateful to my other committee members, Donald Greenwood, Pierre Kabamba and Harris McClamroch who have provided me with insightful and critical comments and suggestions.

Many thanks are due to my wife for her unconditional support and to my family for their long term encouragement. My heartfelt appreciation also goes out to the Mathieus for their warm hospitality.

I am thankful to my professors who have steered my education in sciences, in particular Ms. Artois, Ms. Delye and M. Martel and to my fellow graduate students for making my graduate experience richer and more pleasant. Special mention must also be made to Margaret Fillion and Michelle Shepherd, who have done so much for me over the years.

The work described in this dissertation was funded in part by a grant from the Jet

Propulsion Laboratory, California Institute of Technology which is under contract with the National Aeronautics and Space Administration and by a grant from the National Science Foundation. Additional travel support was provided by the Rackham School of Graduate Studies.

PREFACE

This thesis was submitted at the University of Michigan in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Aerospace Engineering. The thesis was defended on September 9, 2004, in Ann Arbor, MI. The doctoral committee was composed of:

- Anthony M. Bloch (Co-Chair), Mathematics, University of Michigan,
- N. Harris McClamroch, Aerospace Engineering, University of Michigan,
- Donald T. Greenwood, Professor Emeritus, Aerospace Engineering, University of Michigan,
- Pierre T. Kabamba, Aerospace Engineering, University of Michigan,
- Daniel J. Scheeres (Co-Chair), Aerospace Engineering, University of Michigan.

This thesis draws upon the following papers that have been published, presented at conferences or have been (or will be) submitted for publication. However, the organization of the thesis does not strictly follow the organization of each paper, as this dissertation presents the “full” picture of our research.

- Journal papers

  [34] V. M. Guibout and D. J. Scheeres, Solving Relative Two-Point Boundary Value Problems: Spacecraft Formation Flight Transfers Application, Journal of Guidance,

Control, and Dynamics, 27(4): 693-704, 2004.

[31] V. M. Guibout and A. M. Bloch, Discrete Variational Principles and Hamilton-Jacobi Theory for Mechanical Systems and Optimal Control Problems, Physica D, submitted.

[37] V. M. Guibout and D. J. Scheeres, Solving two-point boundary value problems using generating functions: Theory and Applications to optimal control and the study of Hamiltonian dynamical systems, Journal of Nonlinear Science, submitted.

[35] V. M. Guibout and D. J. Scheeres, Computing the generating functions to solve two-point boundary value problems, Journal of Aerospace Computing, Information, and Communication, submitted.

[38] V.M. Guibout and D.J. Scheeres, Spacecraft formation dynamics and design, Journal of Guidance, Control and Dynamics, submitted.

[74] C. Park, V. M. Guibout and D. J. Scheeres, Solving Optimal Low-Thrust Rendezvous Problems with Generating Functions, in preparation.

- Conference papers

  [32] V.M. Guibout and D.J. Scheeres, Formation Flight with Generating Functions: Solving the Relative Boundary Value Problem, AIAA/AAS Astrodynamics Specialist Conference, Monterey, California, August 2002.

  [33] V. M. Guibout and D.J. Scheeres, Finding periodic orbits with generating functions, AAS/AIAA Astrodynamics Specialist Conference, Big Sky, Montana, August 2003.

  [86] D.J. Scheeres, C. Park, and V.M. Guibout, Solving optimal control problems with generating functions, AAS/AIAA Astrodynamics Specialist Conference, Big Sky, Montana, August 2003.

[38] V. M. Guibout and D.J. Scheeres, Spacecraft Formation Dynamics and Design, AIAA/AAS Astrodynamics Specialist Conference, Rhode Island, Providence, August 2004.

[36] V. M. Guibout and D.J. Scheeres, New techniques for spacecraft formation design and control, 18th International Symposium on Space Flight Dynamics, Munich, October 2004.

[30] V. M. Guibout and A.M. Bloch, A discrete maximum principle for solving optimal control problems, 43rd IEEE Conference on Decision and Control, Bahamas, December 2004.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

CHAPTER I

INTRODUCTION

1.1 Two-point boundary value problems

Thanks to Galileo and his telescope, we have been able to observe a wide range of bodies, from comets to stars. These observations made the exploration of space easier than the exploration of the Earth. For instance, Christopher Columbus' voyage to the Orient would have been successful if he had known the actual location of the Orient. He would have not mistaken it for the Americas. However, Galileo's telescope cannot be used to find the paths (orbits) that lead to the observed bodies. We had to wait two centuries for Kepler's and Newton's works before being able to find these orbits. Kepler's (1571-1630) and Newton's (1643-1727) discoveries allow us to understand gravitational laws and to describe the motion of celestial bodies as solutions of ordinary differential equations. As a result, the orbits that lead to celestial bodies can be found as solutions of the differential equations that meet boundary conditions given by an initial position (where we are) and a final position (where we want to go). Lambert (1728-1777) formalized this problem and transformed it into an algebraic equation whose solutions have inspired many papers in the last centuries [24]. Despite this simplification, we usually leave the problem in its ordinary differential equation formulation and search for the unspecified initial conditions that meet the target.

Problems such as the one Lambert considered are often referred to as two-point boundary value problems. As the terminology indicates, the most common case of two-point boundary value problems is where boundary conditions are supposed to be satisfied at two points, usually the starting and ending values of the integration (as in the Lambert problem). In daily life, everyone from researchers to athletes are faced with such problems, although they may not be formulated in such formal terms. Planning a car trip or a mission to Mars, taking a goal shot in soccer or aiming a missile, all of these are examples of two-point boundary value problems where initial and final positions are specified and the corresponding velocities need to be found.

There are crucial distinctions between initial value problems (problems for which the initial position and velocity are known) and two-point boundary value problems. In the former case we are given an “acceptable” solution at the start (initial value) and just march along by numerical integration to its end (final value). In the latter case, the boundary conditions do not determine a unique solution to start with. A random choice among all solutions that satisfy these (incomplete) starting boundary conditions is almost certain to not satisfy the boundary conditions at the other specified point(s). To illustrate this difference, suppose one is at an intersection between two streets, so that one can choose among four directions. In a typical initial value problem, a starting direction is given and one just drives along the road, without knowledge beforehand of the final destination. In contrast, if one is given a final destination instead of a starting direction, we obtain a two-point boundary value problem. Solving this problem is more difficult since one needs, *a priori*, to try each road to find the one that reaches the required final destination. Thus, for an arbitrary boundary value problem it is not surprising that iteration is required in general to meld boundary conditions into a single global solution of the differential equations. Many iterative techniques have been developed over the years, several in the field of optimal control, since the necessary conditions for optimality can be formulated as two-point boundary value problems. In the following we discuss two classes of numerical methods for solving two point boundary value problems, both being iterative.

The shooting method [79, 12] implements the same strategy as the one used in the above example. It consists of choosing values for all of the dependent variables at one boundary. These values must be consistent with any boundary conditions for that boundary, but otherwise are initially guessed randomly. After integration of the differential equations, we in general find discrepancies between the desired boundary values at the other boundary. Then, we adjust the initial guess to reduce these discrepancies and reiterate this procedure again. The method provides a systematic approach to solving boundary value problems, but suffers several inherent limitations. As summarized by Bryson and Ho ([19] p 214),

> The main difficulty with this method is getting started; i.e., finding a first estimate of the unspecified conditions at one end that produces a solution reasonably close to the specified conditions at the other end. The reason for this peculiar difficulty is that the extremal solutions are often very sensitive to small changes in the unspecified boundary conditions.

To get rid of the sensitivity to small changes in initial guesses, techniques such as the multiple shooting method [58] were developed. They consist of breaking the time domain into segments and solving a boundary value problem on each of these segments. In this manner, nonlinear effects are limited over each segment, but on the other hand the size of the problem is increased. However, the choice of the initial conditions still remains as the main hurdle to successfully apply shooting methods to any kind of problem.

Relaxation methods [80] use a different approach. The differential equations are replaced by finite-difference equations on a mesh of points that covers the range of the integration. A trial solution consists of values for the dependent variables at each mesh point, not satisfying the desired finite-difference equations, nor necessarily even satisfying the required boundary conditions. The iteration, now called relaxation, consists of adjusting all the values on the mesh so as to bring them into successively closer agreement with the finite-difference equations and simultaneously with the boundary conditions. In general, relaxation works better than shooting when the boundary conditions are especially delicate or subtle. However, if the solution is highly oscillatory then many grid points are required for accurate representation. Also, the number and position of the required mesh points are not known *a priori* and must be adjusted manually for each problem. In addition, if solutions to the differential equations develop singularities, attempts to refine the mesh to improve accuracy may fail.

In order to plan future space missions, over the years researchers have developed more and more accurate “maps” of motion in space. In this respect, more sophisticated mathematical models have been developed, such as the three-body problem, the four-body problem, and their many variants whose names begin with “full”, “restricted”, “circular”, “elliptic”, “planar”, etc... As these dynamical models have increased in accuracy, the associated boundary value problems usually become *harder* to solve. This is especially true as many of these problems can contain chaotic trajectories, whose extreme sensitivity makes it difficult to find solutions. With the advent of computers, however, the two methods mentioned earlier are still able to solve most of the two-point boundary value problems. They may require substantial time to find an appropriate initial guess and/or computer memory to refine the mesh, but they often succeed.

However, proposed space missions continue to gain in complexity and, most likely, many of tomorrow’s missions may involve several spacecraft in formation. These missions require one to solve a large number of boundary value problems for which the boundary conditions may in turn depend on parameters. For instance, to reconfigure a formation of N spacecraft, there are $N!$ possibilities in general, that is, $N!$ boundary value problems need to be solved. As N increases, the number of boundary value problems dramatically grows. Similarly, suppose that we plan to reconfigure a spacecraft formation to achieve an interferometry mission. We may require the spacecraft to be equally spaced on a circle perpendicular to the line of sight they should observe. In that case, the final positions are specified in terms of the angle that indicates the position of the spacecraft on the circle. In order to find the value of the angle that minimizes fuel expenditure, infinitely many boundary value problems may need to be solved. As a result, the algorithms mentioned above are no longer appropriate as they require excessive computation and time. The present research has been motivated by the need for new methods to address such complex problems that arise in spacecraft formation design (we actually solve the above two spacecraft formation design problems in Chapter VIII). Specifically, we develop a novel approach to solve Hamiltonian boundary value problems based on the generating functions. Our approach outperforms traditional methods for spacecraft formation design and has a broader impact leading to new results in optimal control theory and in the study of the phase space structure of Hamiltonian systems.

1.2 Scope of the thesis

In this thesis, we present a very general theory to solve two-point boundary value problems for Hamiltonian systems. Our method relies on the Hamiltonian nature of the system to naturally describe the nonlinear phase flow in terms of a boundary value problem. There are very few works in the literature that take a similar point of view. In linear systems theory there exists a matrix, sometimes called the perturbation matrix, that describes the flow as a boundary value problem. This matrix verifies a Riccati equation and can be computed from the state transition matrix. It is widely used in optimal control theory to solve linear quadratic terminal controllers and regulators [19], in guidance and navigation, and in astrodynamics to study the relative motion of two spacecraft [10]. For nonlinear systems, however, we could not find any such work save in the field of geometric integrators. Methods based on generating functions [57, 21, 45] allow one to derive symplectic integrators as solutions to several boundary value problems . However, these boundary value problems are restricted to those for which the initial and final states are almost identical and the transfer time is small, as they are designed to generate single steps in long-time integrations.

The method we have developed is based on Hamilton-Jacobi theory. Using generating functions found by solving the Hamilton-Jacobi equation, we can describe the phase flow as a boundary value problem. *Such an approach is very powerful as it allows one to solve any Hamiltonian two-point boundary value problem using only simple function evaluations; no iterations or initial guesses are required*. In addition, this research has implications in several other fields: 1) In optimal control theory, it allows one to develop an explicit solution procedure that finds an analytical form for the nonlinear optimal feedback control law for a general class of problems. Most importantly, this procedure overcomes some of the barriers to truly reconfigurable control. 2) By posing the search for periodic orbits as a two-point boundary value problem with constraints, we develop techniques to find families of periodic orbits, and to exhibit the geometry of phase space about particular solutions. A similar procedure allows us to find relative periodic orbits that are of particular interest when designing spacecraft formation trajectories. 3) In linear systems theory, we recover and extend the results on perturbation matrices developed by R.H. Battin.

Our research is not confined to theoretical work, however. To implement our insight, we develop a robust algorithm to compute a Taylor series expansion of the generating functions. We pay particular attention to the numerics, as our method is based on the symplectic structure of the phase flow, a property that must be preserved during integration. In particular, we present a general framework that allows one to study discretization of certain dynamical systems. This generalizes earlier work on discretization of Lagrangian and Hamiltonian systems on tangent bundles and cotangent bundles respectively [67, 96, 71, 39, 40, 41, 87]. In addition, as noticed by Arnold [5, 6], generating functions may develop singularities which prevent the integration from going forward in time. Using the Legendre transformation, we are able to avoid these singularities, and therefore continue the integration. Furthermore, our algorithm applies to any Hamiltonian system independent of the complexity of its vector field.

Finally, for optimal control problems, we develop a discrete maximum principle that yields necessary conditions for optimality. These conditions are in agreement with the usual conditions obtained from the Pontryagin maximum principle and define symplectic algorithms that solve the optimal control problem.

1.3 Thesis organization

This thesis is organized into three main parts.

The first part includes Chapters II and III and deals with the theoretical aspects of the present research. In Chapter II, we briefly review some features of Hamiltonian systems and then focus on the Hamilton-Jacobi theory. Three points of view are adopted, as they all provide a different perspective on the theory and offer a convenient framework to work with in the rest of this dissertation. On one hand, the variational point of view relates trajectories of Hamiltonian systems to critical points of a certain function. On the other hand, the two geometric points of view characterize trajectories as paths on the tangent bundle. Although the two geometric approaches require advanced mathematical tools and is therefore less accessible, it remains central to the understanding of the discrete Hamilton-Jacobi theory developed in Chapter IV. We believe that the global picture we give in this chapter is a unique exposition on the Hamilton-Jacobi theory in which different points of view are considered. Chapter III is the backbone of this thesis. We first show that the generating functions solve any two-point boundary value problem in phase space. The properties of the generating functions are studied, with a special emphasis on multiple solutions, symmetries, singularities, and their relation to the state transition matrix.

The second part of the thesis focuses on the numerics of our approach and includes Chapters IV and V, and part of Chapter VI. Since our novel approach to solve boundary value problems relies on the symplectic structure of the phase flow, we must understand how this property is preserved during integration. This motivated the work presented in Chapter IV, although we go far beyond our initial objective. Specifically, we introduce a general framework that allows us to study discretizations of Lagrangian and Hamiltonian systems. In particular, we show how to obtain a large class of discrete algorithms using the geometric approach. We give new insight into the Newmark model, for example, and develop a discrete formulation of the Hamilton-Jacobi theory. Based on some results given in Chapter IV on the numerics of Hamiltonian systems, an algorithm that solves the Hamilton-Jacobi equation for generating functions is developed in Chapter V. This algorithm converges locally in the spatial domain and globally in the time domain. Moreover, using the Legendre transformation we are able to handle generating function singularities, and therefore are able to continue the integration. Then, we introduce an indirect approach to compute generating functions based on the initial value problem. Accuracy, convergence and properties of our algorithm are studied. Finally, the first part of Chapter VI focuses on the numerics of optimal control problems. We extend the framework introduced in Chapter IV to develop a discrete maximum principle that yields necessary conditions for optimality that define symplectic algorithms. We show that we are able to recover most of the classical symplectic algorithms and illustrate its use with an example of a sub-Riemannian optimal control problem.

In the third and last part of this dissertation (second part of Chapter VI and Chapters VII and VIII) we analyze a variety of problems in several fields using the theory developed in the first part, together with the algorithm presented in the second part. Section 6.3 concerns optimal control problems. We show that the generating function theory allows one to develop an explicit solution procedure that finds an analytical form for the nonlinear optimal feedback control law for a general class of problems. We illustrate this procedure with an example of the targeting problem in the Hill three-body problem and show that it overcomes some of the barriers to truly reconfigurable control. In Chapter VII, we use generating functions to derive necessary and sufficient conditions for the existence of periodic orbits of a given period, or going through a given point in space. These conditions reduce the search for periodic orbits to either solving a set of implicit equations, which can often be handled graphically, or to finding the roots of an equation of one variable only. Specific examples of finding periodic orbits in the vicinity of other periodic orbits and around the Libration points in the three-body problem are studied. Finally, in Chapter VIII we study spacecraft formation flight. Specifically, we consider the design of spacecraft formations in Earth orbit. For our analysis the effect of the J_2 and J_3 gravity coefficients are taken into account and the reference trajectory is chosen to be an orbit with high inclination ($i = \pi/3$) and eccentricity ($e = 0.3$). Two missions are considered. First, given several tasks over a one month period, modeled as configurations at given times, we find the optimal sequence of reconfigurations to achieve these tasks with minimum fuel expenditure. Next, we find stable configurations such that the spacecraft stay close to each other for an arbitrary, but finite, period of time. Both of these tasks are extremely difficult using conventional approaches, yet are simple to solve using the theory we developed in this dissertation.

CHAPTER II

HAMILTONIAN SYSTEMS AND THE HAMILTON-JACOBI THEORY

The dynamics of real world systems are often too complex to enable full analytical studies and efficient numerical simulations. Thus such systems need to be efficiently modeled. Models must not only provide an accurate picture of the real system, but they must also be tractable analytically and/or numerically. These requirements make modeling a very challenging task. In many different fields such as chemistry, celestial mechanics and plasma physics, Hamiltonian systems have been identified as a relevant class of models. In particular, they are often a highly accurate approximation because non-dissipative forces are dominant. Most importantly, they have a very rich structure and distinctive properties [1, 5, 63, 66, 14] such as preservation of Poincaré invariants, variational principles, the abundance of periodic and quasi periodic motions, the ubiquity of chaos, symmetry and reduction and canonical transformation theory.

These features make the questions one asks about them, and the methods used to answer these questions, fundamentally different from the case of general dynamical systems. One of these features, the Hamilton-Jacobi theory, describes a class of coordinate transformations, known as canonical transformations, that allows one to transform Hamiltonian systems into dynamically trivial ones while preserving their structure and properties.

Hamilton (1805-1865)

Since Hamilton discovered their existence they have been widely used to solve a variety of challenging problems, from integrating non-trivial dynamical systems to deriving symplectic integrators. In the present work, we propose a novel approach to solving two-point boundary value problems based on these transformations. For the sake of clarity it is important to first derive Hamilton-Jacobi theory and study generating functions. In this chapter, we adopt three different points of view; one is variational and two are geometric. We believe that each of them provide a different perspective on the Hamilton-Jacobi theory and offer a convenient framework to work with in the following chapters.

The variational point of view Even though the Lagrange and Hamilton equations were historically not derived from variational principles (some of the history may be gleaned, for example, from Marsden and Ratiu [66] p231 and Bloch et al. [14]), variational principles play an important role in dynamical systems theory. They essentially state that trajectories as defined by Newton's laws correspond to critical points of a certain function, or in other words, the actual path of a particle is the one that minimizes a certain function. Such a formulation allows one to make analogies with geometric optics (that light always takes the shortest path) and optimal control theory (see e.g. Bloch and Crouch [15], Bloch et al. [14], and Chapter VI) for instance. It also provides a coordinate-free formulation for describing the dynamics. Most importantly, this approach introduces a "main" function (the one that is minimized), knowledge of which is sufficient to recover the full dynamics of the system. In the present research, our interest in the variational approach is three-fold: 1) It naturally yields the Hamilton-Jacobi theory. 2) In Chapter IV, we introduce its *discrete* counterpart to study and derive symplectic integrators. 3) It allows one to study optimal control problems (Chapter VI).

The geometric points of view In the Hamiltonian formalism, the dynamics of a particle is described in the phase space consisting of coordinates and associated momenta. The phase space, together with a symplectic two-form (contact two-form) may be given the structure of a symplectic manifold (contact manifold respectively). Therefore, in the geometric approaches the key idea is no longer the existence of a function that needs to be minimized, but the symplectic and contact structures. For instance, we will show that the flow conserves the symplectic (contact) two-form. These geometric approaches are central in this dissertation because: 1) they allow one to study singularities in the Hamilton-Jacobi theory (Section 3.2.3). 2) they provide a convenient framework for deriving a *discrete* Hamilton-Jacobi theory (Section 4.5).

This chapter is organized as follows: The first section introduces Hamiltonian systems using the variational and the two geometric approaches discussed above. The second section focuses on the Hamilton-Jacobi theory; all three points of view are presented. Sections 2.2.3 and 2.2.4 use advanced concepts of geometry to present the geometric approaches. Hence, they may be less accessible than the section adopting the variational point of view (Section 2.2.1). Although the geometric approaches together with the variational point of view give a global and unique picture on the Hamilton-Jacobi theory, these last two sections 2.2.3 and 2.2.4 may be skipped by those who are not interested in the geometric aspects of this theory.

2.1 Hamiltonian systems

Several approaches may be adopted to study Hamiltonian systems. For the bulk of this dissertation, however, we focus on only three, the variational and two geometric approaches. The purpose of this section is to give the essential ideas on each of them, not to review all of their features. We refer to [1, 5, 27, 28, 60, 63, 66, 14, 82] for more details on these topics.

In the following, x is a vector with components x_i. We choose not to use the usual notation **x** or $\vec{x}$ since we believe that there should not be any confusion. Also, we assume Einstein summation convention, i.e., $x_i y_i = \sum_i x_i y_i$.

Definition II.1 (Hamiltonian system). *A system is called Hamiltonian if there exists a smooth function* $H(q,p,t)$ *from* $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ *to* $\mathbb{R}$ *such that its dynamics can be described by equations of the form:*

$$\left\{ \begin{array}{rcl} \dot{q}_i & = & \frac{\partial H}{\partial p_i}\,, \\ \dot{p}_i & = & -\frac{\partial H}{\partial q_i}\,. \end{array} \right. \tag{2.1}$$

H *is called the Hamiltonian function and Eqns.* (2.1) *are known as Hamilton's equations.*

Remark II.2. Consider a dynamical system with Lagrangian function $L(q,\dot{q},t)$ where $q = (q_1, \cdots, q_n)$ are generalized coordinates. If L is hyperregular, i.e., $\frac{\partial L}{\partial \dot{q}}$ is a global isomorphism, then the system is also Hamiltonian and the Hamiltonian function is $H = \langle p, \dot{q} \rangle - L(q,\dot{q},t)$, where $\langle , \rangle$ is the standard dot product, $\langle p, \dot{q} \rangle = p^T q$. We say that H is the Legendre transform of L.

Proof. The dynamics of a Lagrangian system is given by the Euler-Lagrange equations:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0\,. \tag{2.2}$$

If L is hyperregular we can uniquely define the associated momenta $p = (p_1, \cdots, p_n)$ from the Legendre transformation:

$$p_i = \frac{\partial L}{\partial \dot{q}_i}(q,\dot{q},t)\,. \tag{2.3}$$

Let $H(q,p,t) = \langle p, \dot{q} \rangle - L(q,\dot{q},t)$. Then the system of Euler-Lagrange's equation (Eq. (2.2)) is equivalent to the system of Hamilton's equations (Eqns. (2.1)). □

Remark II.3. The Legendre transformation transforms functions on a vector space into functions on the dual space. It has a geometric interpretation that we will make use of in the following. In fact, the Legendre transformation can be formulated as an optimization problem. For sake of simplicity, we consider a one degree of freedom dynamical system whose Lagrangian function $L(q,\dot{q})$ verifies $\det\left(\frac{\partial^2 L}{\partial \dot{q}^2}\right) > 0$. We can construct the Legendre transformation in the following way (Arnold [5]). We draw the graph of L as a function of $\dot{q}$ assuming q fixed. Let p be a given number and define $H(q,p) = langlep, \dot{q}(p)\rangle - L(q,\dot{q}(p))$, where $\dot{q}(p)$ is to be specified. Then H is the Legendre transform of L if and only if $\dot{q}(p)$ is chosen so that H has a maximum with respect to $\dot{q}$ at $\dot{q}(p)$, i.e., $\frac{\partial H}{\partial \dot{q}} = 0$ or equivalently, $p = \frac{\partial L}{\partial \dot{q}}$.

Remark II.4. Hamiltonian systems may not be mechanical systems. For instance in optimal control theory, necessary conditions for optimality yield a Hamiltonian system under sufficient smoothness conditions (Section 6.1). Such a system does not have any physical significance and may not be Lagrangian. For instance, in Chapter 5 we derive necessary conditions for optimality for a sub-Riemannian optimal control problem and show that they yield a degenerate Hamiltonian function. In such cases, we cannot define the Lagrangian from the Hamiltonian function. We need to use the Lagrange multipliers to perform a well-defined Legendre transform (Bloch et al. [14]).

2.1.1 Variational principles and Hamiltonian dynamics

The key element in the variational approach for studying dynamical systems is the existence of some functions whose extrema correspond to actual trajectories of particles. There are many such functions, each of them defining a different variational principle. The most famous ones may be the Hamilton principle, the modified Hamilton's principle and the principal of critical action, but many other variational principles exist and we refer to Arnold [5], Greenwood [28], Bloch et al. [14] and references therein for a more complete presentation. In this section, we focus on only two variational principles, Hamilton's principle and the modified Hamilton's principle. As noticed by Greenwood [28], "the variational principle of most importance in dynamics is Hamilton's principle which was first announced in 1834". Hamilton's principle is a variational principle on the tangent bundle, very powerful for studying Lagrangian systems. However, since Hamiltonian systems lie on the co-tangent bundle, it does not apply directly to these systems. Therefore it needs to be modified. The modified version is called the modified Hamilton's principle, it is the counterpart of the Hamilton principle on the co-tangent bundle.

Consider a configuration manifold $\mathcal{Q}$ and a Lagrangian function L on the extended configuration space $T\mathcal{Q} \times \mathbb{R}$.

Theorem II.5 (Hamilton's principle). *Critical points of $\int_{t_0}^{t_1} Ldt$ in the class of curves $\gamma : \mathbb{R} \to \mathcal{Q}$ whose ends are $q = q_0$ at $t = t_0$ and $q = q_1$ at $t = t_1$, correspond to trajectories of the Lagrangian system whose ends are q_0 at t_0 and q_1 at t_1.*

Proof. We use the calculus of variations to search for the critical points of $\int_{t_0}^{t_1} Ldt$:

$$\begin{aligned}\delta \int_{t_0}^{t_1} L(q,\dot{q},t)dt &= \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q_i}\delta q_i + \frac{\partial L}{\partial \dot{q}_i}\delta \dot{q}_i \right) dt \\ &= \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i dt + \left[\frac{\partial L}{\partial \dot{q}_i}\delta q_i \right]_{t_0}^{t_1} ,\end{aligned}$$

where the Einstein summation convention is used. Since the variations at the end points vanish, we obtain the Euler-Lagrange equations:

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = 0 \, .$$

□

Consider now a Hamiltonian function H on the extended phase space $T^*\mathcal{Q} \times \mathbb{R}$.

Theorem II.6 (Modified Hamilton's principle). *Critical points of $\int_{t_0}^{t_1}(\langle p, \dot{q}\rangle - H)dt$ in the class of paths $\gamma : \mathbb{R} \to T^*\mathcal{Q}$ whose ends lie in the n-dimensional subspaces $q = q_0$ at $t = t_0$ and $q = q_1$ at $t = t_1$ correspond to trajectories of the Hamiltonian system whose ends are q_0 at t_0 and q_1 at t_1.*

Proof. We proceed to the computation of the variation.

$$\begin{aligned}\delta\int_\gamma(\langle p, \dot{q}\rangle - H)dt &= \int_\gamma\left(\dot{q}_i\delta p_i + p_i\delta\dot{q}_i - \frac{\partial H}{\partial q_i}\delta q_i - \frac{\partial H}{\partial p_i}\delta p_i\right)dt\\ &= [p_i\delta q_i]_{t_0}^{t_1} + \int_\gamma\left[\left(\dot{q}_i - \frac{\partial H}{\partial p_i}\right)\delta p_i - \left(\dot{p}_i + \frac{\partial H}{\partial q_i}\right)\delta q_i\right]dt\,.\end{aligned}$$

Therefore, since the variation vanishes at the end points, the integral curves of Hamilton's equations are the only critical points. □

The Hamilton principle allows for variations of the path on a n-dimensional manifold whereas the modified Hamilton principle varies curves on a $2n$-dimensional manifold. Thus, Hamilton's principle is a particular case of the modified Hamilton's principle with the peculiarity that both principles are equivalent for dynamical systems with non-degenerate Lagrangians. Indeed, for these systems H is defined as the Legendre transform of L, that is, $\dot{q}$ and p are such that H is maximized with respect to $\dot{q}$ for every p (see Remark II.3). As a consequence, along extrema we have:

$$\int(\langle p, \dot{q}\rangle - H)dt = \int L dt\,,$$

which proves the equivalence of both principles for non-degenerate Lagrangian systems.

Remark II.7. The conditions for a curve γ to be an extremal of a functional does not depend on the choice of coordinate system. Therefore, these variational principles are coordinate invariant.

2.1.2 Symplectic and contact geometries

The variational approach allows one to characterize the flow of a dynamical system as an extremum of a functional. In contrast, the flow is characterized as a path on the tangent bundle of the configuration manifold in the geometric approaches. In addition, the notion of Hamiltonian systems is generalized to vector fields on symplectic (contact) manifolds. This section is inspired by Abraham and Marsden [1] and is intended to introduce the geometric framework necessary to present the Hamilton-Jacobi theory. We do not intend to provide a complete view on this topic and refer to Abraham and Marsden [1] and references therein for further analysis.

Phase space approach (symplectic geometry)

In this paragraph, we present the Hamiltonian formalism for autonomous systems using symplectic geometry. In particular, we introduce the notion of vector fields and show that the phase space can be given the structure of a symplectic manifold.

Definition II.8. *A symplectic form on a manifold $\mathcal{P}$ is a non-degenerate, closed, two-form ω on $\mathcal{P}$.*

A symplectic manifold $(\mathcal{P}, \omega)$ is a manifold together with a symplectic form ω on $\mathcal{P}$.

A canonical one-form on $\mathcal{P}$, θ, is defined such that $\omega = -d\theta$. By the Poincaré lemma [1], θ is well-defined, at least locally.

The charts guaranteed by Darboux's theorem (see e.g. [1, 66, 14]) are called symplectic charts and the component functions q_i, p_i are called canonical coordinates.

In a symplectic chart,

$$\omega = \sum_{i=1}^{n} dq_i \wedge dp_i \,, \tag{2.4}$$

$$\theta = \sum_{i=1}^{n} p_i dq_i \,. \tag{2.5}$$

Remark II.9. θ as defined by Eq. (2.5) is not unique. In a symplectic chart, the following expressions are also valid: $\theta = -\sum_{i=1}^{n} q_i dp_i$, or more generally

$$\theta = -\sum_{i=1}^{n-k} q_i dp_i + \sum_{i=n-k}^{n} q_i dp_i \, , \; \forall k \leq n \, . \tag{2.6}$$

Definition II.10. *Let $(\mathcal{P}, \omega)$ be a symplectic manifold and $H : \mathcal{P} \to \mathbb{R}$ a smooth function. The vector field X_H defined by*

$$i_{X_H}\omega = dH \, , \tag{2.7}$$

is called a Hamiltonian vector field.

(P, ω, X_H) is called a Hamiltonian system.

Proposition II.11. *Given a configuration space $\mathcal{Q}$, and a hyperregular Lagrangian L on $\mathcal{Q}$, we naturally construct a Hamiltonian system as $(T^*\mathcal{Q}, \omega, X_H)$, where $T^*\mathcal{Q}$ is the cotangent bundle of $\mathcal{Q}$, ω is defined by Eq.* (2.4) *and H is the Legendre transform of L.*

Proposition II.12. *Locally, using the canonical coordinates, a Hamiltonian system on a symplectic manifold reads:*

$$X_H = J \cdot dH \, , \text{ or equivalently, } \dot{q}_i = \frac{\partial H}{\partial p_i} \, , \; \dot{p}_i = -\frac{\partial H}{\partial q_i} \, , \tag{2.8}$$

where $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$.

Proof. The definition of the Hamiltonian vector field (Eq. (2.7)) is expressed in local coordinates as:

$$i_{X_H}(\sum_i dq_i \wedge dp_i) = \sum_i \frac{\partial H}{\partial q_i} dq_i + \sum_i \frac{\partial H}{\partial p_i} dp_i \, . \tag{2.9}$$

Let X_H be :

$$X_H = \sum_i \dot{q}_i \frac{\partial}{\partial q_i} + \sum_i \dot{p}_i \frac{\partial}{\partial p_i} \, .$$

Then,

$$\begin{aligned} i_{X_H}(\sum_i dq_i \wedge dp_i) &= \sum_i (i_{X_H} dq_i) \wedge dp_i - \sum_i dq_i \wedge (i_{X_H} dp_i) \\ &= \sum_i \dot{q}_i dp_i - \dot{p}_i dq_i \,. \end{aligned}$$

Identifying this last equation with Eq. (2.9) leads to Eq. (2.8). □

Extended phase space approach (contact geometry)

Non-autonomous Hamiltonian systems have an extra variable, the time, as compared to autonomous systems. As a result, the phase space becomes $2n+1$-dimensional and the above material does not apply (symplectic manifolds are of even dimension). There are two ways to handle this difference: one may either consider the time as a generalized coordinate and associate with it a generalized momentum, or consider the time as an additional parameter. When the time plays the role of a generalized coordinate, the system is parametrized by an additional independent parameter τ. Obviously, the Hamiltonian function is not a function of τ. Therefore, the $2n$-dimensional non-autonomous Hamiltonian system is transformed into a $2n+2$-dimensional autonomous Hamiltonian system that can be studied using symplectic geometry. An alternative to this approach, the extended phase space approach, consists in giving the extended phase space (q, p, t) the structure of a contact manifold.

Definition II.13. *A contact form ω on a manifold $\mathcal{M}$ is a closed two-form, with maximal rank.*

A contact manifold is a pair $(\mathcal{M}, \omega)$ consisting of an odd-dimensional manifold $\mathcal{M}$ and a contact form ω on $\mathcal{M}$

An exact contact manifold $(\mathcal{M}, \theta)$ consists of a $(2n+1)$-dimensional manifold $\mathcal{M}$ and a one-form θ on $\mathcal{M}$ such that $\theta \wedge (d\theta)^n$ is a volume on $\mathcal{M}$

The next theorem is the equivalent of the Darboux theorem (see e.g. [1, 66, 14]) in symplectic geometry. It gives the canonical form of ω and θ.

Theorem II.14. *Let* $(\mathcal{M}, \omega)$ *be a contact manifold. For each* $x \in \mathcal{M}$ *there is a chart* (U, ϕ) *at* x *with* $\phi(u) = (q_1(u), \cdots, q_n(u), p_1(u), \cdots, p_n(u), w(u))$ *such that*

$$\omega_{|U} = dq_i \wedge dp_i \, . \tag{2.10}$$

Similarly, if $(\mathcal{M}, \theta)$ *is an exact contact manifold, there is a chart* $(\tilde{U}, \tilde{\phi})$ *at* x *such that*

$$\theta_{|U} = dt + p_i dq_i \, . \tag{2.11}$$

Before going into more details of the dynamics on a contact manifold, we introduce the characteristic bundle. It is used to characterize contact forms, and therefore contact manifolds.

Definition II.15. *Let* ω *be a two-form on* $\mathcal{M}$*. Define*

$$R_\omega = \left\{ v = (x, v_1) \in T\mathcal{M} | v^b = 0 \right\} \, ,$$

where v^b *is a one-form such that* $v^b(w) = \omega(x)(v_1, w)$*.*

R_ω *is called the characteristic bundle of* ω*.*

A characteristic vector field is a vector field X *such that* $i_X \omega = 0$*, that is,* $X(x) \in R_\omega \, , \; \forall x \in \mathcal{M}$*.*

Proposition II.16. *The characteristic bundle* R_ω *of a contact form* ω *has one-dimensional fibers, and so is sometimes called the characteristic line bundle.*

Moreover, if ω *is closed and its characteristic line bundle is one-dimensional, then* ω *is a contact form.*

Proof. Let $(\mathcal{M}, \omega)$ be a contact manifold of dimension $(2n+1)$. Then ω has rank $2n$ and for every $x \in \mathcal{M}$, there is a one-dimensional vector space $V \subset T_x\mathcal{M}$ such that, $\forall v \in V$, $\forall w \in T_xM$, $\omega(x)(v, w) = 0$, i.e. R_ω has one-dimensional fibers.

The second part is proved as follows. Since ω is closed, we only need to prove that it is maximal rank. But since R_ω has one-dimensional fibers, ω is of rank $2n$, that is, of maximal rank. □

Proposition II.17. *Let θ be a nowhere zero one-form on a $(2n+1)$-manifold $\mathcal{M}$ and let $R_\theta = \{v = (x, v_1) \in T\mathcal{M} | \theta(x)(v_1) = 0\}$ be its characteristic bundle. Then $(\mathcal{M}, \theta)$ is an exact contact manifold if and only if $d\theta$ is non-degenerate on the fibers of R_θ.*

Proof. R_θ is $2n$-dimensional, so $d\theta$ is non-degenerate on R_θ if and only if $(d\theta)^n \neq 0$. By definition of $\wedge$ and R_θ, this is so if and only if $\theta \wedge (d\theta)^n \neq 0$.

□

In the above we have introduced the contact structure and learned to characterize it using the characteristic bundle. We now prove that this concept generalizes the symplectic structure to non-autonomous Hamiltonian systems. Specifically, we show that the symplectic structure of the phase space of autonomous systems can be extended to a contact structure. Next, we focus on non-autonomous systems. We define the notion of (time-dependent) vector fields and give the extended phase space a contact structure.

Proposition II.18. *Let $(\mathcal{P}, \omega)$ be a symplectic manifold, $\pi : \mathbb{R} \times \mathcal{P} \to \mathcal{P}$ the projection on P and $\tilde{\omega} = \pi^*\omega$. Then $(\mathbb{R} \times \mathcal{P}, \tilde{\omega})$ is a contact manifold.*

The characteristic line bundle of $\tilde{\omega}$ is generated by the vector field $\underline{t}$ on $\mathbb{R} \times \mathcal{P}$ given by:

$$\underline{t}(s, z) = ((s, 1), (z, 0)) \in T_{(s,z)}(\mathbb{R} \times \mathcal{P}) \sim T_s\mathbb{R} \times T_z\mathcal{P}$$

If $\omega = d\theta$ and $\tilde{\theta} = dt + \pi^*\theta$, where $t : \mathbb{R} \times \mathcal{P} \to \mathbb{R}$ is the projection on $\mathbb{R}$, then $\tilde{\omega} = d\tilde{\theta}$ and $(\mathbb{R} \times \mathcal{P}, \tilde{\theta})$ is an exact contact manifold.

Proof. Clearly, $d\tilde{\omega} = \pi^* d\omega = 0$, so $\tilde{\omega}$ is closed. To show that $\tilde{\omega}$ is maximal rank, it suffices to show that the fibers of its characteristic bundle is one-dimensional (Prop. II.16). Let $((s, z), v_1) \in R_{\tilde{\omega}}$, then for all $v_2 \in T_{(s,z)}(\mathbb{R} \times \mathcal{P})$,

$$\tilde{\omega}(s, z)(v_1, v_2) = 0\,,$$

that is:

$$\omega(z)(T\pi^* v_1, T\pi^* v_2) = 0\,,\ \forall v_2\,.$$

Since ω is non-degenerate, we conclude that $T\pi^* v_1 = 0$, i.e.,

$$R_{\tilde{\omega}} = \{((s, z), (v, 0)) | v \in \mathbb{R}\}$$

has a one-dimensional fiber.

To prove that $(\mathbb{R} \times \mathcal{P}, \tilde{\theta})$ is an exact contact manifold, we need to show that $\tilde{\theta} \wedge (d\tilde{\theta})^n \neq 0$, that is, dt is non-zero on the characteristic line bundle of $\tilde{\omega}$. But, we have just proved that the fibers of the characteristic line bundle are of dimension 1 and are generated by $(1, 0) \in T_s\mathbb{R} \times T_z\mathcal{P}$ at any point $(s, z) \in \mathbb{R} \times \mathcal{P}$. Furthermore, $dt(s)(1, 0) = 1$. Thus, $\tilde{\theta}$ is non-zero on the fiber of $R_{\tilde{\omega}}$ and $(\mathbb{R} \times P, \tilde{\theta})$ is an exact contact manifold. □

Similarly to the autonomous case, non-autonomous Hamiltonian systems are characterized by their vector fields. In this case, however, they are defined on contact manifolds.

Definition II.19. *Let $(\mathcal{P}, \omega)$ be a symplectic manifold and $H : \mathbb{R} \times \mathcal{P} \to \mathbb{R}$. For each t define $H_t : \mathcal{P} \to \mathbb{R}; z \mapsto H(t, z)$, $X_H : \mathbb{R} \times \mathcal{P} \to T\mathcal{P}; (t, z) \mapsto X_{H_t}(z)$ and the suspension $\tilde{X}_H : \mathbb{R} \times \mathcal{P} \to T(\mathbb{R} \times \mathcal{P}) \simeq T\mathbb{R} \times T\mathcal{P}; (t, z) \mapsto ((t, 1), X_H(t, z))$.*

With $\underline{t}$ as defined in Prop. II.18, $\tilde{X}_H = \underline{t} + X_H$.

Hence, the time-dependent vector field X_{H_t} is obtained by freezing t and constructing the usual Hamiltonian vector field. Such a definition yields classical Hamilton's equations of motion:

Proposition II.20. *Let (U, ϕ) be a symplectic chart of $\mathcal{P}$ with*

$$\phi(u) = (q_1(u), \cdots, q_n(u), p_1(u), \cdots, p_n(u)) ,$$

and $(\mathbb{R} \times U, t \times \phi)$ a chart of $\mathbb{R} \times P$, where t is the projection onto $\mathbb{R}$ defined previously. Then $c : I \to \mathbb{R} \times U; t \mapsto (t, b(t))$ is an integral curve of $\tilde{X}_H$, or equivalently, $b : I \to U$ is an integral curve of X_H, if and only if:

$$\begin{aligned} \tfrac{d}{dt}[q_i(b(t))] &= \tfrac{\partial H}{\partial p_i}(t, b(t)) , \\ \tfrac{d}{dt}[p_i(b(t))] &= -\tfrac{\partial H}{\partial q_i}(t, b(t)) . \end{aligned}$$

We now bring together the concept of contact manifold and the definition of time-dependent vector fields to give the *extended* phase space a contact structure.

Theorem II.21. *(Cartan) Let $(\mathcal{P}, \omega)$ be a symplectic manifold and $H : \mathbb{R} \times \mathcal{P} \to \mathbb{R}$. Let $\tilde{\omega} = \pi^* \omega$ as defined above and*

$$\omega_H = \tilde{\omega} + dH \wedge dt , \tag{2.12}$$

Then,

1. *$(\mathbb{R} \times \mathcal{P}, \omega_H)$ is a contact manifold,*

2. *$\tilde{X}_H$ generates the line bundle of ω_H; $\tilde{X}_H$ is the unique vector field satisfying*

$$i_{\tilde{X}_H} \omega_H = 0 \quad \textit{and} \quad i_{\tilde{X}_H} dt = 1 .$$

3. *if $\omega = d\theta$ and $\theta_H = \pi^* \theta + H dt$, then $\omega_H = d\theta_H$.*

Proof. 1. $d\omega_H = d\tilde{\omega} + d(dH \wedge dt) = 0$, so ω_H is closed. Further, ω_H coincides with $\tilde{\omega}$ on vectors of the form $((s,0),(z,v)) \in T_s\mathbb{R} \times T_z\mathcal{P}$ on which $\tilde{\omega}$ is non-degenerate as we saw before (Prop. (II.18) states that $((s,0),(z,v))$ is not in $R_{\tilde{\omega}}$). Thus, ω_H is closed and of maximal rank and $(\mathbb{R} \times \mathcal{P}, \omega_H)$ is a contact manifold.

2. For all vector field Y on $\mathbb{R} \times \mathcal{P}$,

$$\begin{aligned} i_{\tilde{X}_H}\tilde{\omega}(Y) &= \tilde{\omega}(\tilde{X}_H, Y) \\ &= \omega(T\pi \cdot \tilde{X}_H, T\pi \cdot Y) \\ &= \omega(X_H, T\pi \cdot Y)\,. \end{aligned} \tag{2.13}$$

Thus,

$$\begin{aligned} i_{\tilde{X}_H}\omega_H &= i_{\tilde{X}_H}\tilde{\omega} + (i_{\tilde{X}_H}dH)dt - dH(i_{\tilde{X}_H}dt) \\ &= \omega(X_H, T\pi \cdot Y) + \frac{\partial H}{\partial t}(T_t \cdot Y) - dH \cdot Y \\ &= 0 \end{aligned} \tag{2.14}$$

since $i_{\tilde{X}_H}dH = dH(\underline{t}) = \frac{\partial H}{\partial t}$ and $i_{\tilde{X}_H}dt = i_{\underline{t}+X_H}dt = dt(\underline{t}) = 1$. The characteristic bundle being one dimensional, $\tilde{X}_H$ is unique.

3. First, we can readily verify that $\omega_H = d\theta_H$. Moreover,

$$\begin{aligned} \theta_H(\tilde{X}_H) &= (\pi^*\theta)(X_H + \underline{t}) + H dt(X_H + \underline{t}) \\ &= \theta(X_H) + H\,. \end{aligned} \tag{2.15}$$

Thus, θ_H does not vanish on the characteristic bundle of ω_H. We conclude that $(\mathbb{R}\times\mathcal{P}, \theta_H)$ is an exact contact manifold. □

2.1.3 Properties of Hamiltonian systems

We briefly introduced Hamiltonian systems using three different points of view. These approaches provide us with different perspectives on the dynamics of Hamiltonian systems and offer a convenient framework to work with in the following. But before going further we need to recall a few properties of Hamiltonian systems. Again, we restrict ourselves to those properties that are of interest for the present research.

Properties of autonomous Hamiltonian systems

Autonomous Hamiltonian systems have distinctive properties. Most of them are not of prime interest for the following and so we do not mention them. However, there are some, such as the conservation of energy and the invariance of the symplectic two-form along the flow, we need to pay particular attention to.

- The conservation of energy constrains the motion of a particle to lie on an energy surface. This property is important for the understanding of the dynamics of a Hamiltonian system. Therefore, the energy may need to be preserved when proceeding to numerical simulations (see Chapter IV for more details on this topic).

- The invariance of the symplectic two-form along the flow is the most important feature of Hamiltonian systems. It implies volume conservation and additional stability properties, for instance. Most importantly, it allows one to embed the phase space in a symplectic manifold.

Let Φ_t be the phase flow of the Hamiltonian system $(\mathcal{P}, \omega, X_H)$

$$\begin{aligned} \Phi_t : \mathcal{P} &\rightarrow \mathcal{P} \\ (q_0, p_0) &\mapsto (\Phi_t^1(q_0, p_0) = q(q_0, p_0, t), \Phi_t^2(q_0, p_0) = p(q_0, p_0, t)) \end{aligned} \tag{2.16}$$

Proposition II.22. *Φ_t preserves the symplectic structure, i.e., $(\Phi_t)^*\omega = \omega$*

Proof. From the Lie derivative theorem (Bloch et al. [14] page 87) and $i_{X_H}\omega = dH$, we obtain:

$$\frac{d}{dt}{\Phi_t}^*\omega = {\Phi_t}^*\mathcal{L}_X\omega = {\Phi_t}^*(i_{X_H}d + di_{X_H})\omega = 0\,.$$

□

Corollary II.23. *Each of the forms* $(\omega)^2$, $(\omega)^3$, $\cdots$ *is an integral invariant of* Φ_t. *If the dimension of* $\mathcal{P}$ *is* $2n$, *then the conservation of* ω^n *is equivalent to volume conservation in the phase space.*

Proposition II.24 (Energy conservation). $H \circ \Phi_t = H$, *i.e., the energy is conserved along trajectories.*

The proof is straightforward using the definition of the vector field X_H.

Properties of non-autonomous Hamiltonian systems

The geometry of the phase space of non-autonomous Hamiltonian systems is different from that of autonomous systems. As a result, non-autonomous systems do not have the same properties. In particular, the energy is not preserved along trajectories ($L_{\tilde{X}_H} H \neq 0$) and the contact two-form is an invariant of the time-dependent flow of $\tilde{X}_H$.

Proposition II.25 (Non energy conservation). *The energy of non-autonomous Hamiltonian systems is not conserved along the flow, i.e.,* $L_{\tilde{X}_H} H = \frac{\partial H}{\partial t} \neq 0$

Proof. Since $\tilde{X}_H = \underline{t} + X_H$, $L_{\tilde{X}_H} H = dH(\underline{t} + X_H) = dH(\underline{t}) = \frac{\partial H}{\partial t}$. □

Proposition II.26. *The contact two-forms* ω_H, ω_H^2, $\cdots$ *are invariant forms of* $\tilde{X}_H$.

Proof. Since ω_H is closed and $\tilde{X}_H$ is a characteristic vector field of ω_H, we have $L_{\tilde{X}_H} \omega_H = i_{\tilde{X}_H} d\omega_H + d i_{\tilde{X}_H} \omega_H = 0$. $L_{\tilde{X}_H}$ being a derivation, we directly obtain that $L_{\tilde{X}_H} \omega_H^k = 0$ as well. □

Proposition II.27. $dt \wedge \omega_H^n = dt \wedge \tilde{\omega}^n$ *is an invariant volume element for* $\tilde{X}_H$.

Proof. Since $L_{\tilde{X}_H} dt = d(dt \cdot \tilde{X}_H) = d(1) = 0$, we have:

$$L_{\tilde{X}_H}(dt \wedge \omega_H^n) = (L_{\tilde{X}_H} dt) \wedge \omega_H^n = 0\,.$$

□

2.2 Local Hamilton-Jacobi theory

We now move on to the derivation of the Hamilton-Jacobi theory. The Hamilton-Jacobi theory describes a class of coordinate transformations, called canonical transformations, that allow one to transform Hamiltonian systems. In the previous section, we introduced the Hamiltonian formalism from three different points of view. Depending on the approach we took, the formalism we used was very different. For instance, in the variational approach the symplectic two-form does not have any significance and in the geometric approach, trajectories are not critical points of any functions. However, all these approaches are equivalent. The same will hold for the Hamilton-Jacobi theory: canonical transformations have different definitions depending on the adopted point of view, yet they are all equivalent.

- In the variational approach, we saw that the key idea is the existence of a function whose critical points are trajectories of the system. As a consequence, canonical transformations are defined with respect to this concept. Specifically, a coordinate transformation is canonical if it preserves this relationship between critical points and trajectories.

- In the symplectic geometry approach, the symplectic two-form is the main object. Therefore, in this context canonical transformations are defined as coordinate transformations that preserve the symplectic two-form.

- Finally, the extended phase space approach relies on the contact structure, that is, on the contact two-form. Thus, canonical transformations are those that preserve the contact two-form.

Although the variational and symplectic geometry approaches can be found in many textbooks, the contact geometry point of view does not seem to be easily accessible, except partially in Abraham and Marsden [1]. We believe that the global picture we give in this chapter is a unique exposition on the Hamilton-Jacobi theory, in which the different points of view are confronted.

2.2.1 The variational approach

Definition II.28. *Let H define a Hamiltonian system. Then $f : \mathcal{P} \times \mathbb{R} \rightarrow \mathcal{P} \times \mathbb{R}$ is a canonical transformation from (q, p, t) to (Q, P, t) if and only if:*

(1)- f is a diffeomorphism,

(2)- f preserves the time, i.e., there exists a function g_t such that $f(x, t) = (g_t(x), t)$,

(3)- Critical points of $\int_{t_0}^{t_1} \Big(\langle P, \dot{Q}\rangle - K(Q, P, t)\Big)\, dt$ correspond to trajectories of the Hamiltonian system, where $K(Q, P, t)$ is the Hamiltonian function expressed in the new set of coordinates.

Consider a canonical transformation between two sets of coordinates in the phase space $f : (q, p, t) \mapsto (Q, P, t)$ and let $H(q, p, t)$ and $K(Q, P, t)$ be the Hamiltonian functions of the same system expressed in different sets of coordinates. From Def. II.28, trajectories correspond to critical points of $\int_{t_0}^{t_1} \Big(\langle P, \dot{Q}\rangle - K(Q, P, t)\Big)\, dt$. Therefore, they are integral of:

$$\left\{ \begin{array}{rcl} \dot{Q}_i & = & \frac{\partial K}{\partial P_i}\,, \\ \dot{P}_i & = & -\frac{\partial K}{\partial Q_i}\,, \end{array} \right. \tag{2.17}$$

i.e., f preserves the canonical form of Hamilton's equations.

Conversely, suppose that f is a coordinate transformation that preserves the canonical form of Hamilton's equations and leaves the time invariant. Let $K(Q, P, t)$ be the Hamiltonian in the new set of coordinates, then from the modified Hamilton's principle (Thm.

II.6), critical points of

$$\int_{t_0}^{t_1} \Big(\langle P, \dot{Q}\rangle - K(Q,P,t)\Big)\, dt$$

correspond to trajectories of the system. Thus, f is a canonical map. These last two remarks are summarized in the following lemma:

Lemma II.29. *The third item in Def. II.28 is equivalent to:*

(4)- f preserves the canonical form of Hamilton's equations and the new Hamiltonian function is $K(Q,P,t)$.

Remark II.30. The definition we give is different from the one given in many textbooks. Often the third item reduces to:

(5)- f preserves the canonical form of Hamilton's equations.

The example given by Arnold in "Mathematical Methods of Classical Mechanics" [5], p 241 sheds light on the difference on these two definitions. For instance, consider the transformation $f : (q,p,t) \mapsto (Q = q, P = 2p, t)$ and the harmonic oscillator with Hamiltonian function $H(q,p) = \frac{1}{2}p^2 + \frac{1}{2}q^2$. The equations of motion for this system are:

$$\dot{q} = p\,, \quad \dot{p} = -q\,. \tag{2.18}$$

In the new set of coordinates, these equations transform into:

$$\dot{Q} = \frac{p}{2}\,, \quad \dot{P} = -2Q\,. \tag{2.19}$$

Define $K(Q,P) = \frac{1}{4}P^2 + Q^2$. Then Eqns. (2.19) may be written as:

$$\dot{Q} = \frac{\partial K}{\partial P}\,, \quad \dot{P} = -\frac{\partial K}{\partial Q}\,, \tag{2.20}$$

that is, Eqns. (2.19) may be written as Hamilton's equations. As a result, f preserves the canonical form of Hamilton's equations. However, according to Def. II.28, f is not a canonical transformation because the Hamiltonian of the new system should be $K(Q,P) = \frac{1}{8}P^2 + \frac{1}{2}Q^2$.

We consider again a canonical transformation f and a Hamiltonian system defined by H. Along trajectories, we have by definition:

$$\delta \int_{t_0}^{t_1} \left(\sum_{i=1}^{n} p_i \dot{q}_i - H(q,p,t) \right) dt = 0 \,, \tag{2.21}$$

$$\delta \int_{t_0}^{t_1} \left(\sum_{i=1}^{n} P_i \dot{Q}_i - K(Q,P,t) \right) dt = 0 \,. \tag{2.22}$$

From Eqns. (2.21) - (2.22), we conclude that the integrands of the two integrals differ at most by a total time derivative of an arbitrary function F:

$$\sum_{i=1}^{n} p_i dq_i - H dt = \sum_{j=1}^{n} P_j dQ_j - K dt + dF \tag{2.23}$$

Such a function is called a generating function for the canonical transformation f. It is, *a priori*, a function of both the old and the new variables and time. The two sets of coordinates being connected by the $2n$ equations, namely, $f(q,p,t) = (Q,P,t)$, F can be reduced to a function of $2n+1$ variables among the $4n+1$. Thus, we can define 4^n generating functions that have n variables in P_1 and n in P_2. Among these are the four kinds defined by Goldstein [27]:

$$F_1(q_1, \cdots, q_n, Q_1, \cdots, Q_n, t) \,, \quad F_2(q_1, \cdots, q_n, P_1, \cdots, P_n, t) \,,$$
$$F_3(p_1, \cdots, p_n, Q_1, \cdots, Q_n, t) \,, \quad F_4(p_1, \cdots, p_n, P_1, \cdots, P_n, t) \,.$$

Let us first consider the generating function $F_1(q,Q,t)$. The total time derivative of F_1 reads:

$$dF_1(q,Q,t) = \sum_{i=1}^{n} \frac{\partial F_1}{\partial q_i} dq_i + \sum_{j=1}^{n} \frac{\partial F_1}{\partial Q_i} dQ_i + \frac{\partial F_1}{\partial t} dt \,.$$

Hence Eq. (2.23) yields:

$$\sum_{i=1}^{n} (p_i - \frac{\partial F_1}{\partial q_i}) dq_i - H dt = \sum_{j=1}^{n} (P_j + \frac{\partial F_1}{\partial Q_j}) dQ_j - K dt + \frac{\partial F_1}{\partial t} dt \,. \tag{2.24}$$

Assume that (q, Q, t) is a set of independent variables. Then Eq. (2.24) is equivalent to:

$$p_i = \frac{\partial F_1}{\partial q_i}(q, Q, t)\,,\ P_i = -\frac{\partial F_1}{\partial Q_i}(q, Q, t)\,,\ K(Q, -\frac{\partial F_1}{\partial Q}, t) = H(q, \frac{\partial F_1}{\partial q}, t) + \frac{\partial F_1}{\partial t}\,. \quad (2.25)$$

Eqns. (2.25) characterize F_1. If (q, Q) is not a set of independent variables, we say that F_1 is singular (see Chapter III for more details on singularities).

Now let us consider more general forms of generating functions. Let $(i_1, \cdots, i_p)$ $(i_{p+1}, \cdots, i_n)$ and $(k_1, \cdots, k_r)(k_{r+1}, \cdots, k_n)$ be two partitions of the set $\{1, \cdots, n\}$ into two non-intersecting parts such that $i_1 < \cdots < i_p$, $i_{p+1} < \cdots < i_n$, $k_1 < \cdots < k_r$ and $k_{r+1} < \cdots < k_n$. In addition, we define $I_p = (i_1, \cdots, i_p)$, $\bar{I}_p = (i_{p+1}, \cdots, i_n)$, $K_r = (k_1, \cdots, k_r)$ and $\bar{K}_r = (k_{r+1}, \cdots, k_n)$. If

$$(q_{I_p}, p_{\bar{I}_p}, Q_{K_r}, P_{\bar{K}_r}) = (q_{i_1}, \cdots, q_{i_p}, p_{i_{p+1}}, \cdots, p_{i_n}, Q_{k_1}, \cdots, Q_{k_r}, P_{k_{r+1}}, \cdots, P_{k_n})$$

are independent variables, then we can define the generating function

$$F_{I_p, K_r}(q_{I_p}, p_{\bar{I}_p}, Q_{K_r}, P_{\bar{K}_r}, t)\,.$$

Expanding dF_{I_p,K_r} yields:

$$\begin{aligned} dF_{I_p,K_r} &= \sum_{a=1}^{p} \frac{\partial F_{I_p,K_r}}{\partial q_{i_a}} dq_{i_a} + \sum_{a=p+1}^{n} \frac{\partial F_{I_p,K_r}}{\partial p_{i_a}} dp_{i_a} + \sum_{a=1}^{r} \frac{\partial F_{I_p,K_r}}{\partial Q_{k_a}} dQ_{k_a} \\ &\quad + \sum_{a=r+1}^{n} \frac{\partial F_{I_p,K_r}}{\partial P_{k_a}} dP_{k_a} + \frac{\partial F_{I_p,K_r}}{\partial t} dt\,, \end{aligned} \quad (2.26)$$

and rewriting Eq. (2.23) as a function of the linearly independent variables leads to:

$$\sum_{a=1}^{p} p_{i_a} dq_{i_a} - \sum_{a=p+1}^{n} q_{i_a} dp_{i_a} - H dt = \sum_{a=1}^{r} P_{k_a} dQ_{k_a} - \sum_{a=r+1}^{n} Q_{k_a} dP_{k_a} - K dt + dF_{I_p,K_r}\,, \quad (2.27)$$

where

$$F_{I_p,K_r} = F_1 + \sum_{a=r+1}^{n} Q_{k_a} P_{k_a} - \sum_{a=p+1}^{n} q_{i_a} p_{i_a}\,. \quad (2.28)$$

Eq. (2.28) is often referred to as the *Legendre transformation*, it allows one to transform one generating function into another.

We then substitute Eq. (2.26) into Eq. (2.27):

$$\begin{aligned}\sum_{a=1}^{r}(P_{k_a}+\frac{\partial F_{I_p,K_r}}{\partial Q_{k_a}})dQ_{k_a}+\sum_{a=r+1}^{n}(\frac{\partial F_{I_p,K_r}}{\partial P_{k_a}}-Q_{k_a})dP_{k_a}-Kdt+\frac{\partial F_{I_p,K_r}}{\partial t}dt\\ =\sum_{a=1}^{p}(p_{i_a}-\frac{\partial F_{I_p,K_r}}{\partial q_{i_a}})dq_{i_a}-\sum_{a=p+1}^{n}(q_{i_a}+\frac{\partial F_{I_p,K_r}}{\partial p_{i_a}})dp_{i_a}-Hdt\,,\end{aligned} \tag{2.29}$$

and obtain the set of equations that characterizes F_{I_p,K_r}:

$$p_{I_p} = \frac{\partial F_{I_p,K_r}}{\partial q_{I_p}}(q_{I_p},p_{\bar{I}_p},Q_{K_r},P_{\bar{K}_r},t)\,, \tag{2.30}$$

$$q_{\bar{I}_p} = -\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}}(q_{I_p},p_{\bar{I}_p},Q_{K_r},P_{\bar{K}_r},t)\,, \tag{2.31}$$

$$P_{K_r} = -\frac{\partial F_{I_p,K_r}}{\partial Q_{K_r}}(q_{I_p},p_{\bar{I}_p},Q_{K_r},P_{\bar{K}_r},t)\,, \tag{2.32}$$

$$Q_{\bar{K}_r} = \frac{\partial F_{I_p,K_r}}{\partial P_{\bar{K}_r}}(q_{I_p},p_{\bar{I}_p},Q_{K_r},P_{\bar{K}_r},t)\,, \tag{2.33}$$

$$\begin{aligned}K(Q_{K_r},\frac{\partial F_{I_p,K_r}}{\partial P_{\bar{K}_r}},-\frac{\partial F_{I_p,K_r}}{\partial Q_{K_r}},P_{\bar{K}_r},t) =\\ H(q_{I_p},-\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}},\frac{\partial F_{I_p,K_r}}{\partial q_{I_p}},p_{\bar{I}_p},t)+\frac{\partial F_{I_p,K_r}}{\partial t}\,.\end{aligned} \tag{2.34}$$

For the case where the partitions are $(1,\cdots,n)()$ and $()(1,\cdots,n)$ (i.e., $p=n$ and $r=0$), we recover the generating function F_2, which verifies the following equations:

$$p_i=\frac{\partial F_2}{\partial q_i}(q,P,t)\,,\ Q_i=\frac{\partial F_2}{\partial P_i}(q,P,t)\,,\ K(\frac{\partial F_2}{\partial P},P,t)=H(q,\frac{\partial F_2}{\partial q},t)+\frac{\partial F_2}{\partial t}\,. \tag{2.35}$$

The case $p=0$ and $r=n$ corresponds to a generating function of the third kind, F_3:

$$q_i=-\frac{\partial F_3}{\partial p_i}(p,Q,t)\,,\ P_i=-\frac{\partial F_3}{\partial Q_i}(p,Q,t)\,,\ K(Q,-\frac{\partial F_3}{\partial Q},t)=H(-\frac{\partial F_3}{\partial p},p,t)+\frac{\partial F_3}{\partial t}\,. \tag{2.36}$$

Finally, if $p=0$ and $r=0$, we obtain F_4:

$$q_i=-\frac{\partial F_4}{\partial p_i}(p,P,t)\,,\ Q_i=\frac{\partial F_4}{\partial P_i}(p,P,t)\,,\ K(-\frac{\partial F_4}{\partial P},P,t)=H(\frac{\partial F_4}{\partial p},p,t)+\frac{\partial F_4}{\partial t}\,. \tag{2.37}$$

For a generating function to be well-defined, we need to make the assumption that its variables are linearly independent. In Chapter III we see that this hypothesis is often not satisfied. The following property grants us that at least one of the generating function is well-defined at every instant.

Proposition II.31. *Let $f : \mathcal{P}_1 \times \mathcal{P}_2$ be a canonical transformation. Using the above notation, there exist at least two partitions I_p and K_r such that $(q_{I_p}, p_{\bar{I}_p}, Q_{K_r}, P_{\bar{K}_r}, t)$ are linearly independent.*

Proof. Suppose we cannot find such I_p and K_r. Then, we could generate the canonical transformation using less than $2n$ variables. Using the local inversion theorem, we conclude that this is in contradiction with f being a diffeomorphism (Def. II.35). □

In the class of canonical transformations, changes of coordinates that transform the system to equilibrium ($K = constant$) are of particular interest: they transform the system into a trivial one. For these particular transformations, Eq. (2.34) simplifies into the Hamilton-Jacobi equation.

Theorem II.32 (Hamilton-Jacobi). *Let f be a canonical transformation and let F_{I_p,K_r} be the associated generating function. Then, f transforms the Hamiltonian to equilibrium if and only if F_{I_p,K_r} verifies the Hamilton-Jacobi equation:*

$$\frac{\partial F_{I_p,K_r}}{\partial t} + H(q_{I_p}, -\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}}, \frac{\partial F_{I_p,K_r}}{\partial q_{I_p}}, p_{\bar{I}_p}, t) = constant\,. \tag{2.38}$$

Proof. On one hand, if f transforms the system to equilibrium then $K = constant$ and Eq. (2.34) simplifies into Eq. (2.38). On the other hand, if f is a canonical transformation and F_{I_p,K_r} verifies the Hamilton-Jacobi equation, then $K = constant$, i.e., the system is transformed to equilibrium. □

Remark II.33. Generating functions are not unique. Both F_{I_p,K_r} and $F_{I_p,K_r} + constant$ verify the Hamilton-Jacobi equation and Eqns. (2.30)-(2.33). Therefore, Eq. (2.38) may be equivalently written as:

$$\frac{\partial F_{I_p,K_r}}{\partial t} + H(q_{I_p}, -\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}}, \frac{\partial F_{I_p,K_r}}{\partial q_{I_p}}, p_{\bar{I}_p}, t) = 0\,. \tag{2.39}$$

In the literature, Eq. (2.39) is often called the Hamilton-Jacobi equation and Eq. (2.38) is not given any name. Starting in Chapter III, we follow this convention and the Hamilton-Jacobi equation will always refer to Eq. (2.39) except as otherwise mentioned.

Thm. II.32 is the backbone of the theory we present in this dissertation. It relates solutions of the Hamilton-Jacobi partial differential equation to canonical transformations that map Hamiltonian dynamical systems into trivial ones. There are many such mappings, all of them satisfy the Hamilton-Jacobi equation but with different boundary conditions.

2.2.2 The generating function for integrating the equations of motion

The Hamilton-Jacobi theory has found many applications over the years. It was first used to integrate the equations of motion of integrable Hamiltonian systems. Branching from this, a variety of applications were developed. For instance, it was used to derive symplectic integrators [57, 21, 45] and prove the existence of the action-angle variables [5]. In this section, through a non-trivial example, we present the use of the Hamilton-Jacobi theory for integrating the equations of motion. This example, taken from Classical Dynamics [28], will help us to highlight fundamental differences between the classical approach and the method we develop in this dissertation.

Example II.34 (The two-body problem). The two-dimensional two-body problem consists of a particle of unit mass attracted by an inverse-square gravitational force to a fixed point. The dynamics of the particle is described by the Hamiltonian function:

$$H(r,\theta,p_r,p_\theta) = \frac{1}{2}\left(p_r^2 + \frac{p_\theta^2}{r^2}\right) - \frac{\mu}{r}\,,$$

where (r,θ) are polar coordinates centered at the fixed point. Since H is time-independent, H is conserved along the trajectory (Prop. II.24). In addition, θ does not appear in H and therefore its conjugate momentum p_θ has a constant value. Now consider the generating function of the second kind associated with the extended phase flow canonical transformation[1]: $F_2(r,\theta,t,p_{r_0},p_{\theta_0},p_t)$. From Eqns. (3.10)-(3.11) and the fact that p_θ and H are constants of motion, we can write F_2 in the following form [28]:

$$F_2(r,\theta,t,p_{r_0},p_{\theta_0},p_t) = p_t t + p_\theta\theta + W(r,-H,p_\theta)\,.$$

Then the Hamilton-Jacobi equation (3.12) reads:

$$\frac{1}{2}\left(\frac{\partial W}{\partial r}\right)^2 + \frac{p_\theta^2}{2r^2} - \frac{\mu}{r} + H = 0\,.$$

Integration of the Hamilton-Jacobi equation yields:

$$F_2(r,\theta,t,p_{r_0},p_{\theta_0},-H) = \int_{r_0}^{r}\sqrt{-2H + 2\frac{\mu}{r} - \frac{p_\theta^2}{r^2}dr}\,,$$

where r_0 is the value of r at the initial time $t_0 = 0$. Now recall Eq. (3.11):

$$\begin{aligned} t_0 &= -\frac{\partial F_2}{\partial H} \\ &= t - \int_{r_0}^{r}\frac{dr}{\sqrt{-2H + 2\frac{\mu}{r} - \frac{p_\theta^2}{r^2}}}\,, \\ \theta_0 &= -\frac{\partial F_2}{\partial p_\theta dr} \\ &= \int_{r_0}^{r}\frac{p_\theta}{r\sqrt{-2Hr^2 + 2\mu r - p_\theta^2}}\,. \end{aligned}$$

Integration of this last equation provides the equation of motion:

$$r = \frac{p_\theta^2/\mu}{1 + e\cos(\theta)}\,,\ \text{where } e = \sqrt{1 + 2Hp_\theta^2/\mu^2}\,.$$

[1] In the extended phase space, the time plays the role of a generalized coordinates with associated momentum $p_t = -H$.

Thus, the Hamilton-Jacobi theory allows us to find the equations of motion for the two-body problem. The methodology used is very general since we just need to find a canonical map that transforms the system into an easily integrable one. The search for such a map remains difficult and this aspect limits the use of the Hamilton-Jacobi theory in practice. Instead, in the present research we focus on a single transformation, the one induced by the phase flow that maps the system to its initial state. Under this transformation, the system is in equilibrium and every point in phase space can be considered to be an equilibrium point. In general, we cannot compute this transformation (if we were able to find this transformation, it would mean that we could integrate the equations of motion) and so we focus on the generating functions that generate this transformation. In particular, we prove that they solve two-point boundary value problems (Chapter III) and we develop an algorithm for approximating them (Chapter V).

We now derive the Hamilton-Jacobi equation from geometric points of view. Both approaches for autonomous and non-autonomous dynamical systems are presented. The outline is inspired from the text “Foundations of Mechanics” [1] but the definition of canonical transformations was modified to allow for comparison with the variational approach. Section 2.2.3 deals with autonomous Hamiltonian systems (symplectic geometry) whereas Section 2.2.4 is focused on non-autonomous systems (contact geometry). Again, these two sections use advanced concepts of geometry and are therefore less accessible. Those who are not interested in the geometry of the Hamilton-Jacobi theory may skip the end of this chapter. Results in the next two sections are not used in the following, however similar reasoning is developed to derive the discrete Hamilton-Jacobi theory (Section 4.5).

2.2.3 From the phase space point of view

We first define the concept of canonical transformations on symplectic manifolds. Then we introduce the generating functions and finally derive the Hamilton-Jacobi equation.

Canonical transformations

Definition II.35. *Let $(\mathcal{P}_1, \omega_1)$ and $(\mathcal{P}_2, \omega_2)$ be symplectic manifolds. A C^∞-mapping $f : \mathcal{P}_1 \longrightarrow \mathcal{P}_2$ is called symplectic or canonical if and only if $f^*\omega_2 = \omega_1$.*

We now prove that this definition is equivalent to Def. II.28. First we show that it implies that f is a diffeomorphism, and then prove the other two items of Def. II.28.

Proposition II.36. *If f is symplectic then f is a diffeomorphism.*

Proof. Suppose f is not a diffeomorphism, i.e., there exists $x \in \mathcal{P}_1$ such that

$$\exists v_1 \in T_x\mathcal{P}_1 \mid Tf \cdot v_1 = 0\,.$$

Since f is symplectic, we have:

$$\forall v_2 \in T_x\mathcal{P}_1 \mid v_2 \neq 0\,,\ \omega_1(x)(v_1, v_2) = \omega_2(f(x))(Tf \cdot v_1, Tf \cdot v_2)\,.$$

The right hand side is zero but the left hand side is not. This is a contradiction and therefore f is a diffeomorphism. □

Lemma II.37. *Let $(\mathcal{P}_1, \omega_1)$ and $(\mathcal{P}_2, \omega_2)$ be symplectic manifolds and f a canonical transformation, $f : \mathcal{P}_1 \longrightarrow \mathcal{P}_2;\ (q,p) \mapsto (Q,P)$. Then, $f^*\omega_2 = \omega_1$ can be written in matrix form as $F^TJF = J$ where F is locally defined by $F_i^k := \frac{\partial f_k}{\partial x_i}$ and J is the local matrix representation of the symplectic two-form using canonical coordinates: $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.*

Proof. Using local canonical coordinates the symplectic forms read (Darboux's theorem):

$$\omega_1 = \sum_i dq_i \wedge dp_i \, , \; \omega_2 = \sum_i dQ_i \wedge dP_i \, .$$

In addition, f being symplectic is equivalent to $f^*\omega_2 = \omega_1$, that is,

$$\forall x = (q_1, \cdots, q_n, p_1, \cdots, p_n) \in P_1 \, , \; \forall v_1, v_2 \in T_x P_1 \, ,$$
$$\omega_2(f(x))(Tf \cdot v_1, Tf \cdot v_2) = \omega_1(x)(v_1, v_2) \, .$$

This last equation reads as $F^T J F = J$ in matrix form. □

Let $(\mathcal{P}_1, \omega_1, X_H)$ define a Hamiltonian system, $(\mathcal{P}_2, \omega_2)$ define a symplectic manifold and f be a canonical transformation, $f : \mathcal{P}_1 \to \mathcal{P}_2; (q, p) \mapsto (Q, P)$.

By definition, X_H is an Hamiltonian vector field if and only if $i_{X_H}\omega_1 = dH$. Define $X_K = f_* X_H$ and let us show that X_K is a Hamiltonian vector field.

$$\begin{aligned} i_{X_H}\omega_1 &= dH \, , \\ i_{f^* X_K} f^* \omega_2 &= dH \, , \\ f^*(i_{X_K}\omega_2) &= dH \, , \\ i_{X_K}\omega_2 &= f_* dH \, . \end{aligned} \tag{2.40}$$

Hence, X_K is a Hamiltonian vector field for the Hamiltonian function f_*H, that is, for the Hamiltonian function H expressed as a function of the new variables.

We summarize this last result in the following proposition.

Proposition II.38. *Let X_H be an Hamiltonian vector field with Hamiltonian function H. Then, f being a canonical transformation is equivalent to $f_* X_H$ being an Hamiltonian vector field with Hamiltonian function $f_* H$.*

Prop. II.36 together with Prop. II.38 show the equivalence between Def. II.35 and Def. II.28.

Generating functions We now introduce the concept of generating functions. Most of the results in this section are taken from Abraham and Marsden, “Foundations of Mechanics” [1].

Proposition II.39. *Let $(\mathcal{P}_1, \omega_1)$ and $(\mathcal{P}_2, \omega_2)$ be symplectic manifolds, $\pi_i : \mathcal{P}_1 \times \mathcal{P}_2 \to \mathcal{P}_i$ the projection onto $\mathcal{P}_i$, $i = 1, 2$ and*

$$\Omega = \pi_1^* \omega_1 - \pi_2^* \omega_2 \, . \tag{2.41}$$

Then:

1. *Ω is a symplectic form on $\mathcal{P}_1 \times \mathcal{P}_2$;*
2. *a map $f : \mathcal{P}_1 \to \mathcal{P}_2$ is symplectic if and only if $i_f^* \Omega = 0$, where $i_f : \Gamma_f \to \mathcal{P}_1 \times \mathcal{P}_2$ is the inclusion map and Γ_f is the graph of f.*

Proof. We need to verify that Ω is closed and non-degenerate. Since ω_i is closed and d commutes with the pull-back, we have:

$$\begin{aligned} d\Omega &= d(\pi_1^* \omega_1 - \pi_2^* \omega_2) \, , \\ &= \pi_1^* d\omega_1 - \pi_2^* d\omega_2 \, , \\ &= 0 \, , \end{aligned}$$

Thus, Ω is closed.

Let us choose $x = (x_1, x_2) \in \mathcal{P}_1 \times \mathcal{P}_2$ and $v = (v_1, v_2) \in T_x(\mathcal{P}_1 \times \mathcal{P}_2) \sim T_{x_1}\mathcal{P}_1 \times T_{x_2}\mathcal{P}_2$ such that

$$\forall w = (w_1, w_2) \in T_x(\mathcal{P}_1 \times \mathcal{P}_2) \, , \; \Omega(p)(v, w) = 0 \, .$$

Let us show that v is zero.

$$\begin{aligned} \Omega(x)(v, w) &= \omega_1(\pi_1(x))(T\pi_1 \cdot v, T\pi_1 \cdot w) - \omega_2(\pi_2(x))(T\pi_2 \cdot v, T\pi_2 \cdot w) \\ &= \omega_1(x_1)(v_1, w_1) - \omega_2(x_2)(v_2, w_2) \, . \end{aligned} \tag{2.42}$$

Eq. (2.42) is zero for all (w_1, w_2) if and only if each of the terms are zero, that is,

$$\omega_1(x_1)(v_1, w_1) = 0\,,\ \forall w_1 \ \text{ and } \ \omega_2(x_2)(v_2, w_2) = 0\,,\ \forall w_2\,.$$

Since ω_i is non-degenerate, we conclude that $v_1 = v_2 = 0$. Thus Ω is closed and non-degenerate, i.e., Ω is a symplectic form.

We now prove the second statement. f induces a diffeomorphism of $\mathcal{P}_1$ on Γ_f so we can write

$$T_{(x,f(x))}\Gamma_f = \{(v, Tf\cdot v)|v \in T_xP_1\}\,.$$

Therefore,

$$\begin{aligned}(i_f^*\Omega)(x, f(x))((v_1, Tf\cdot v_1),&(v_2, Tf\cdot v_2))\\ &= \omega_1(x)(v_1, v_2) - \omega_2(f(x))(Tf\cdot v_1, Tf\cdot v_2)\\ &= (\omega_1 - f^*\omega_2)(x)(v_1, v_2)\,.\end{aligned}$$

We conclude that f is symplectic if and only if $i_f^*\Omega = 0$. □

Ω being closed, the Poincaré lemma guaranties the existence of a one-form Θ such that locally $\Omega = -d\Theta$. Now we assume that f is symplectic. Then, we have $di_f^*\Theta = i_f^*d\Theta = 0$, i.e., $i_f^*\Theta$ is closed. Using again the Poincaré lemma, we show that there exists locally a function $F : \Gamma_f \to \mathbb{R}$ such that $i_f^*\Theta = dF$.

Definition II.40. *Such a function F is called a generating function for the symplectic map f. In addition, F is locally defined and is not unique (since Θ is not unique, Eq.* (2.6)*).*

If $(q_1, \cdots, q_n, p_1, \cdots, p_n)$ are coordinates on $\mathcal{P}_1$ and $(Q_1, \cdots, Q_n, P_1, \cdots, P_n)$ are coordinates on $\mathcal{P}_2$, then Γ_f can be given a chart in several ways. For instance, F may appear as a function of (q_i, Q_i) or of (q_i, P_i), and so forth depending of the choice of Θ.

- Let[2] $\theta_1 = p_i dq_i$ and $\theta_2 = P_i dQ_i$,

[2] The Einstein convention for indices is used.

then $i_f^*\Theta = i_f^*\pi_1^*\theta_1 - i_f^*\pi_2^*\theta_2 = (\pi_1 \circ i_f)^* p_i dq_i - (\pi_2 \circ i_f)^* P_i dQ_i$. Suppose F is a function of $(q^1, \cdots, q^n, Q^1, \cdots, Q^n)$. Then from

$$dF = \frac{\partial F}{\partial q_i} dq_i + \frac{\partial F}{\partial Q_i} dQ_i \text{ and } i_f^*\Theta = dF\,,$$

we conclude that:

$$p_i = \frac{\partial F}{\partial q_i} \text{ and } P_i = -\frac{\partial F}{\partial Q_i}\,. \tag{2.43}$$

We recover the generating function of the first kind $F_1(q_1, \cdots, q_n, Q_1, \cdots, Q_n)$.

- Let $\theta_1 = p_i dq_i$ and $\theta_2 = -Q_i dP_i$,

  then $i_f^*\Theta = i_f^*\pi_1^*\theta_1 - i_f^*\pi_2^*\theta_2 = (\pi_1 \circ i_f)^* p_i dq_i + (\pi_2 \circ i_f)^* Q_i dP_i$. Suppose F is a function of (q_i, P_i), then using $i_f^*\Theta = dF$ we conclude that:

$$p_i = \frac{\partial F}{\partial q_i} \text{ and } Q_i = \frac{\partial F}{\partial P_i}\,. \tag{2.44}$$

  We recover the generating of the second kind $F_2(q_1, \cdots, q_n, P_1, \cdots, P_n)$.

- Different choices of Θ yield different generating functions. In the same manner, we can recover the 4^n generating functions introduced previously (Section 2.2.1).

The Hamilton-Jacobi theory The theorems we give in this section are not taken from "Foundations of Mechanics" [1] and, as far as we know, cannot be found in the literature (although we believe that they are well-known). Let $\mathcal{Q}$ be the configuration space of an autonomous Hamiltonian system and consider a canonical transformation $f : T^*\mathcal{Q} \to T^*\mathcal{Q}; (q,p) \mapsto (Q,P)$. Without loss of generality, we focus on the generating functions of the first kind, F_1, associated with f. Since f is time-independent, the energy expressed in either set of coordinates is conserved along trajectories, i.e., $H(q,p) = constant = K(Q,P)$. Using Eq. (2.43) this last equation reads:

$$H(q, \frac{\partial F_1}{\partial q}) = K(Q, -\frac{\partial F_1}{\partial Q})\,. \tag{2.45}$$

In addition, if we assume that f transforms the system to equilibrium then K is a constant and Eq. (2.45) simplifies into

$$H(q, \frac{\partial F_1}{\partial q}) = E\,. \tag{2.46}$$

Eq. (2.46) is the time-independent Hamilton-Jacobi equation. The remainder of this section is devoted to explaining the derivation of this time-independent Hamilton-Jacobi equation in more detail.

In the following f is a canonical transformation from $T^*\mathcal{Q}$ to $T^*\mathcal{Q}$; $f(q,p) = (Q,P)$ and X_H is the Hamiltonian vector field associated with H on $(T^*\mathcal{Q}, \omega_1 = dq_i \wedge dp_i)$. From Prop. II.38, $X_K = f_* X_H$ is the Hamiltonian vector field associated with the function $K = f_* H$ on $(T^*\mathcal{Q}, \omega_2 = dQ_i \wedge dP_i)$.

Theorem II.41. *Let $F_1 : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ be a smooth function. Define $\tilde{p}(q,Q) = \frac{\partial F_1}{\partial q}(q,Q)$ and $\tilde{P}(q,Q) = -\frac{\partial F_1}{\partial Q}(q,Q)$. Then the following two conditions are equivalent:*

1. *F_1 is the generating function associated with f;*

2. - *For every curve $c(t)$ in $\mathcal{Q}$ satisfying:*

     $$c'(t) = T\tau^*_{\mathcal{Q}} \cdot X_H(c(t), \tilde{p}(c(t),Q))\,, \tag{2.47}$$

     the curve $t \mapsto (c(t), \tilde{p}(c(t),Q))$ is an integral curve of X_H, where $\tau^*_{\mathcal{Q}} : T^*\mathcal{Q} \to \mathcal{Q}$ is the cotangent bundle projection.

   - *For every curve $c(t)$ in $\mathcal{Q}$ satisfying:*

     $$c'(t) = T\tau^*_{\mathcal{Q}} \cdot X_K(c(t), \tilde{P}(q, c(t)))\,,$$

     the curve $t \mapsto (c(t), \tilde{P}(q, c(t)))$ is an integral curve of X_K.

The idea of this theorem is rather simple. Let Q be fixed (the following also applies if q is fixed instead), and define the map $\tilde{p} : q \mapsto \frac{\partial F_1}{\partial q}$ which associates a momentum to every point on $\mathcal{Q}$. Then, construct a curve $c : \mathbb{R} \to \mathcal{Q}$ such that $c(t)$ verifies the differential equation (2.47):

$$\dot{q} = \frac{\partial H}{\partial p}(q, \tilde{p}(q))\,. \tag{2.48}$$

Once c is constructed, look at the curve $\tilde{p}(c(t)) = \frac{\partial F_1}{\partial q}(c(t), Q)$. The theorem states that F_1 is a generating function for f if and only if $\tilde{p}(c(t))$ is the momentum associated with c. In other words, F_1 is a generating function for f if and only if $\tilde{p}(c(t))$ verifies the differential equation

$$\dot{\tilde{p}} = -\frac{\partial H}{\partial q}(c(t), \tilde{p})\,, \tag{2.49}$$

or equivalently if and only if $t \mapsto (c(t), \tilde{p}(c(t)))$ verifies Hamilton's equations.

Proof. Suppose F_1 is a generating function of f. Let Q be fixed and consider a curve $c : \mathbb{R} \mapsto \mathcal{Q}$ verifying Eq. (2.47), that is

$$c'(t) = \frac{\partial H}{\partial p}(c(t), \tilde{p}(c(t))\,,$$

where $\tilde{p}(c(t)) = \frac{\partial F_1}{\partial q}(c(t), Q)$. Since F_1 is a generating function, $\tilde{p}(c(t))$ is the generalized momentum associated with $c(t)$. Therefore, we immediately obtain that $t \mapsto (c(t), \tilde{p}(c(t)))$ is an integral of curve of X_H.

We apply the same reasoning for deriving the second condition with q fixed instead. This conclude the proof of $1. \Rightarrow 2.$.

Now suppose item $2.$ is verified and let us show that F_1 is a generating function of f, i.e., $\tilde{p} = \frac{\partial F_1}{\partial q}$ and $\tilde{P} = -\frac{\partial F_1}{\partial Q}$ are the momenta associated with q and Q. But this is exactly the meaning of the statements:

The curve $t \mapsto (c(t), \tilde{p}(c(t), Q))$ is an integral curve of X_H and the curve $t \mapsto (c(t), \tilde{P}(q, c(t)))$ is an integral curve of X_K. □

Theorem II.42. *Let $F_1 : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ be a smooth function. Then the following two conditions are equivalent:*

1. *F_1 is a generating function associated with f;*

2. *For every $H : \mathbb{R} \times T^*\mathcal{Q} \to \mathbb{R}$, there is a function $K : \mathbb{R} \times T^*\mathcal{Q} \to \mathbb{R}$ such that*
$$H(q, \frac{\partial F_1}{\partial q}) = K(Q, -\frac{\partial F_1}{\partial Q})\,. \tag{2.50}$$

Proof. Assume that Q is fixed and consider a curve $c(t)$ in $\mathcal{Q}$ such that:

$$c'(t) = T\tau^*_{\mathcal{Q}} \cdot X_H(c(t), \tilde{p}(c(t)))\,. \tag{2.51}$$

From Thm. II.41, $t \mapsto (c(t), \tilde{p}(c(t)))$ is an integral curve of X_H. Let $X_H = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i}$, then,

$$X_H\left(c(t), \frac{\partial S}{\partial q_i} dq_i\right) = \left(\frac{\partial H}{\partial p}(c(t), \tilde{p}(c(t))), -\frac{\partial H}{\partial q}(c(t), \tilde{p}(c(t))\right)\,.$$

Applying $T\tau^*_{\mathcal{Q}}$ yields:

$$c(t)' = T\tau^*_{\mathcal{Q}} \cdot X_H(c(t), \tilde{p}(c(t))) = \frac{\partial H}{\partial p}(c(t), \tilde{p}(c(t)))\,. \tag{2.52}$$

Further, the statement "$t \mapsto (c(t), \tilde{p}(c(t)))$ is an integral curve of X_H" is equivalent to the following:

$$\begin{aligned}
(c(t), \tilde{p}(c(t)))' &= X_H\left(c(t), \tilde{p}(c(t))\right)\,, \\
\left(c'(t), \frac{\partial S}{\partial q_i}\left(c(t)\right) dq_i\right)' &= X_H\left(c(t), \frac{\partial S}{\partial q}(c(t))\right)\,, \\
\text{Taking only the} \quad i^{th} &\quad \text{component of the second part:} \\
\frac{\partial^2 S}{\partial q_i \partial q^j}(c(t)) \cdot c_j'(t) &= -\frac{\partial H}{\partial q_i}\left(c(t), \frac{\partial S}{\partial q}(c(t))\right)\,, \\
\frac{\partial^2 S}{\partial q_i \partial q^j}(c(t)) \cdot \frac{\partial H}{\partial p_j}\left(c(t), \frac{\partial S}{\partial q}(c(t))\right) &= -\frac{\partial H}{\partial q_i}\left(c(t), \frac{\partial S}{\partial q}(c(t))\right)\,.
\end{aligned} \tag{2.53}$$

On the other hand, deriving the left side of Eq. (2.50) yields:

$$\frac{d}{dt}H(q, \frac{\partial S}{\partial q}) = \frac{\partial H}{\partial q_i}(q, \frac{\partial S}{\partial q})\dot{q}_i + \frac{\partial H}{\partial p_i}(q, \frac{\partial S}{\partial q})\frac{\partial^2 S}{\partial q_i \partial q^j}(q)\dot{q}^j \,.$$

The i^{th} component, α_i, reads

$$\alpha_i = \frac{\partial H}{\partial q_i}(q, \frac{\partial S}{\partial q}) + \frac{\partial H}{\partial p_j}(q, \frac{\partial S}{\partial q}) \cdot \frac{\partial^2 S}{\partial q^j \partial q_i}(q)\,. \tag{2.54}$$

Using Eq. (2.53), we conclude that Eq. (2.54) is identically zero if and only if F_1 is a generating function for f. In the same way we prove that the total time derivative of the right hand side of Eq. (2.50) is zero if and only if F_1 is a generating function of f. Therefore $H(q, \frac{\partial S}{\partial q})$ and $K(Q, -\frac{\partial S}{\partial Q})$ differ at most by a constant that can be added to K. □

The final result of this section is the Hamilton-Jacobi theorem. We have already derived it from the variational point of view and we now present its geometric version.

Theorem II.43 (Hamilton-Jacobi). *Let $F_1 : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ be a generating function associated with the canonical transformation f. Then the following two conditions are equivalent:*

1. *f transforms the Hamiltonian system $(T^*\mathcal{Q}, \omega_1, X_H)$ into a new Hamiltonian system $(T^*\mathcal{Q}, \omega_2, X_K)$ which is in equilibrium, i.e., $K = constant$, $X_K = 0$.*

2. *F_1 satisfies the Hamilton-Jacobi equation $H(q_i, \frac{\partial F_1}{\partial q_i}) = E$ where E is a constant.*

Proof. The proof is obvious using Thm. II.42. Indeed, if F_1 verifies the Hamilton-Jacobi equation, then $K(Q, -\frac{\partial F_1}{\partial Q}) = E$ and $X_K = 0$. On the other hand, if $K = constant$ Eq. (2.50) simplifies to the Hamilton-Jacobi equation. □

Remark II.44. If $H = T + V$, i.e., $p_i = \dot{q}_i$, then,

$$T\tau_Q^* \cdot X_H(q_i, p_i) = T\tau_Q^*((q_i, p_i), (\dot{q}_i, \dot{p}_i)) = (q_i, \dot{q}_i)\,.$$

Thus, $T\tau_Q^* X_H = identity$. If we assume a given Q, the first item in Thm. II.41 reads: "$c(t)$ is a gradient line of F_1", i.e., $c(t)$ is orthogonal to level surfaces of F_1. As a result, solutions to the Hamilton-Jacobi equation can be constructed as follows: for a given Q, F_1 is such that locally the trajectories of the system are orthogonal to its level surfaces (for each q, $\dot{q}$ is orthogonal to a level surface of F_1). This is the beginning of the analogy with geometric optics, we refer to Abraham and Marsden [1], Arnold [5], Lanczos [60], Chetaev [22] and references therein for more details. This remark is also crucial for understanding the geometric construction of the Hamilton-Jacobi theory based on the full picture developed by Caratheodory and then Rund in [81] (see also Bliss [13] and Hestenes [49]).

For non-autonomous Hamiltonian systems, we mentioned previously that there were two ways to handle the time: we could either consider it as a generalized coordinate or as an additional parameter. In the first case, the dimension of the phase space becomes $2n+2$ and the coordinates are no longer independent (the momentum associated with the time coordinate is the opposite of the total energy of the system). If one applies the autonomous Hamilton-Jacobi theory to the $2n + 2$ dimensional system then canonical transformations are no longer generated by the classical generating functions but by generalized canonical transformations (see e.g. Greenwood [28] and Struckmeier and Riedel [89]). On the other hand, if one considers the time as an independent parameter, the above material does not apply and we must derive the Hamilton-Jacobi theory in the framework of contact geometry. In particular, we must re-define canonical transformations and generating functions.

2.2.4 From the extended phase space point of view

This section is also inspired by the excellent book “Foundations of Mechanics” written by Abraham and Marsden [1] but the definition of canonical transformations have been modified to allow comparison with previous sections. The last theorems on the Hamilton-Jacobi theory cannot be found in this book nor in the literature (although we believe that they are well known).

Canonical transformation

Definition II.45. *Let $(\mathcal{P}_1, \omega_1)$ and $(\mathcal{P}_2, \omega_2)$ be symplectic manifolds and $(\mathbb{R} \times \mathcal{P}_i, \tilde{\omega}_i)$ the corresponding contact manifold* [3]*. A smooth mapping $f : \mathbb{R} \times \mathcal{P}_1 \to \mathbb{R} \times \mathcal{P}_2$ is called a canonical transformation if each of the following holds:*

- *(C1) f is a diffeomorphism,*
- *(C2) f preserves time, that is $f^*t = t$,*
- *(C3) there is a function $K_f : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$ such that*

  $$f^*\omega_{K_f} = \tilde{\omega}_1 \,,$$

  where $\omega_{K_f} = \tilde{\omega}_2 + dK_f \wedge dt$.

Moreover, if $\omega_i = -d\theta_i$, then (C3) is equivalent to

- *(C4) there is a K_f such that $f^*\tilde{\theta}_2 - \theta_{K_f}$ is closed, where $\tilde{\theta}_i = dt + \pi^*\theta_i$ and $\theta_{K_f} = \tilde{\theta}_2 - K_f dt$.*

In order to make analogies with the autonomous Hamilton-Jacobi theory, we need to characterize the property of being canonical in a more familiar way.

[3] $\tilde{\omega}_i$ has been defined in Prop. II.18

Proposition II.46. *Let $(\mathcal{P}_1, \omega_1)$ and $(\mathcal{P}_2, \omega_2)$ be symplectic manifolds and $(\mathbb{R} \times \mathcal{P}_i, \tilde{\omega}_i)$ the corresponding contact manifolds. A smooth mapping $f : \mathbb{R} \times \mathcal{P}_1 \to \mathbb{R} \times \mathcal{P}_2$ is called a canonical transformation if each of the following holds:*

- *(C1) f is a diffeomorphism,*
- *(C2) f preserves time, that is $f^*t = t$,*
- *(C5) for all $H : \mathbb{R} \times \mathcal{P}_1 \to \mathbb{R}$, there is a $K : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$ such that*

$$f^*\omega_K = \omega_H \,.$$

(C5) states that canonical transformations must preserve the contact two-form. This condition is similar to the definition of time-independent canonical transformations, namely f preserves the symplectic two-form.

Proof. Suppose (C3) holds and define $K = f_*H + K_f$. Then,

$$\begin{aligned} f^*\omega_K &= f^*(\tilde{\omega}_2 + dK \wedge dt) \\ &= f^*\tilde{\omega}_2 + d(K \circ f) \wedge d(t \circ f) \\ &= f^*\tilde{\omega}_2 + d(K \circ f) \wedge dt \qquad \text{(since } f \text{ preserves time)} \\ &= \tilde{\omega}_1 - f^*dK_f \wedge dt + d(K \circ f) \wedge dt \\ &= \tilde{\omega}_1 + dH \wedge dt \,. \end{aligned}$$

Conversely, choose H to be zero and let $K_f = K$. Then $f^*\omega_K = \omega_H$ reduces to (C3). □

Remark II.47. Suppose $(\mathcal{P}, \omega, X_H)$ is an autonomous Hamiltonian system. From Prop. II.18 we know that the extended phase space maybe be given the contact structure $(\mathbb{R} \times \mathcal{P}, \pi^*\omega)$. In that case, $(C5)$ reduces to "f preserves the symplectic two-form". In addition, if f is time-independent, then $(C2)$ is trivially verified. Thus, the above definition is equivalent to Def. II.35 for autonomous Hamiltonian systems and time-independent canonical transformations.

Proposition II.48. *Let f be a canonical transformation, then f preserves the canonical form of all time-dependent Hamiltonian systems, i.e., for all $H : \mathbb{R} \times \mathcal{P}_1 \to \mathbb{R}$, there is a $K : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$ such that $f_*\tilde{X}_H = \tilde{X}_K$.*

Let $\pi : \mathbb{R} \times \mathcal{P}_i \to \mathcal{P}_i$ be the projection on $\mathcal{P}$ and $j_t : \mathcal{P}_i \to \mathbb{R} \times \mathcal{P}_i; x \mapsto (t, x)$, then for every $t \in \mathbb{R}$, $f_t : \mathcal{P}_1 \to \mathcal{P}_2; \pi \circ f \circ j_t$ is symplectic.

Proof. From Thm. II.21, X_H is uniquely defined by $i_{\tilde{X}_H}\omega_H = 0$ and $i_{\tilde{X}_H}dt = 1$. Thus,

$$i_{f_*\tilde{X}_H}\omega_K = i_{f_*\tilde{X}_H}f_*\omega_H = f_*i_{\tilde{X}_H}\omega_H = 0\,.$$

Moreover, $i_{f_*\tilde{X}_H}dt = 1$ since f preserves time. By uniqueness, we conclude that $\tilde{X}_K = f_*\tilde{X}_H$.

The remainder of the proof follows from the following lemma. □

Lemma II.49. *f_t is symplectic for each t if and only if there is a one-form α on $\mathbb{R} \times \mathcal{P}_2$ such that $f^*(\tilde{\omega}_2 + \alpha \wedge dt) = \tilde{\omega}_1$.*

Proof. If $f^*(\tilde{\omega}_2 + \alpha \wedge dt) = \tilde{\omega}_1$, then

$$\begin{aligned} f_t^*(\omega_2) &= (j_t^* f^* \pi^*)\omega_2 \\ &= j_t^* f^* \tilde{\omega}_2 \\ &= j_t^*\tilde{\omega}_1 - j_t^* f^*(\alpha \wedge dt) \\ &= j_t^*\pi^*\omega_1 - j_t^* f^*\alpha \wedge j_t^* f^* dt \\ &= \omega_1\,, \end{aligned}$$

since $j_t^* f^* dt = d(j_t^* f^* t) = d(j_t^* t) = 0$. Therefore, f_t is symplectic for all t.

Conversely, assume f_t is symplectic and let $\beta = \tilde{\omega}_2 - f_*\tilde{\omega}_1$. Then,

$$j_t^* f^* \beta = j_t^* f^* \tilde{\omega}_2 - j_t^* \tilde{\omega}_1 = f_t^* \omega_2 - \omega_1 = 0\,.$$

Since β is a two-form on $\mathbb{R} \times \mathcal{P}_2$, it can be written as $\beta = \gamma + \alpha \wedge dt$, where γ is a two-form which does not involve dt. From $j_t^* f^* \beta = 0$ and $j_t^* f^* \beta = \gamma$ we conclude that $\gamma = 0$. □

Theorem II.50 (Jacobi). *Let $f : \mathbb{R} \times \mathcal{P}_1 \to \mathbb{R} \times \mathcal{P}_2$ satisfy (C1) and (C2). Then, (C3) is equivalent to*

(C6) There is a function $K_f : \mathbb{R} \times P_2 \to \mathbb{R}$ such that for all $H : \mathbb{R} \times \mathcal{P}_1 \to \mathbb{R}$, $f_* \tilde{X}_H = \tilde{X}_K$, where $K = f_* H + K_f$.

Proof. We have already proven that $(C3)$ implies $(C6)$ in Prop. II.48. For the converse, taking $H = 0$ leads to $f_* \underline{t} = \tilde{X}_{K_f}$. For an arbitrary H, we have,

$$f_* \tilde{X}_H = f_* X_H + \tilde{X}_{K_f} = f_* X_H + X_{K_f} + \underline{t}\,.$$

Further, we also have,

$$\tilde{X}_K = X_K + \underline{t} = X_{f_* H} + X_{K_f} + \underline{t}\,, \text{ and } \tilde{X}_K = f_* \tilde{X}_H\,.$$

Combining these equations yields:

$$\begin{aligned} f_* X_H &= f_*(\tilde{X}_H - \underline{t}) = f_* \tilde{X}_H - \tilde{X}_{K_f} \\ &= \tilde{X}_K - \tilde{X}_{K_f} = X_{f_* H}\,. \end{aligned} \tag{2.55}$$

We define $H_t = j_t^* H$ and recall that $f_{t*} = f_t^{-1*} = j_t^* f_* \pi^*$. Then,

$$\begin{aligned} X_{f_{t*} H_t} &= X_{(j_t \circ \pi \circ f^{-1} \circ j_t)^* H} \\ &= (j_t \circ \pi \circ f^{-1} \circ j_t)^* X_H \quad \text{by Eq. (2.55)} \\ &= j_t^* f_* \pi^* X_{H_t} \\ &= f_{t*} X_{H_t}\,. \end{aligned}$$

Using Prop. II.38, we conclude that f_t is symplectic for all t and that there exists[4] a one-form α such that $f^*(\tilde{\omega}_2 + \alpha \wedge dt) = \tilde{\omega}_1$. Hence,

$$i_{\tilde{X}_{K_f}} f_* \tilde{\omega}_1 = i_{\tilde{X}_{K_f}} \tilde{\omega}_2 + (i_{\tilde{X}_{K_f}} \alpha) \wedge dt - \alpha \wedge i_{\tilde{X}_{K_f}} dt\,. \tag{2.56}$$

In addition, since $\tilde{X}_{K_f} = f_* \underline{t}$, we have:

$$i_{\tilde{X}_{K_f}} f_* \tilde{\omega}_1 = f_* i_{\underline{t}} \tilde{\omega}_1 = 0\,. \tag{2.57}$$

Using Eq. 2.57 together with $i_{\tilde{X}_{K_f}} dt = 1$ and $f^* t = t$, Eq. 2.56 simplifies to:

$$\alpha = i_{f_* \underline{t}} \tilde{\omega}_2 + (i_{f_* \underline{t}} \alpha) dt\,,$$

that is, $f_* \tilde{\omega}_1 = \tilde{\omega}_2 + (i_{f_* \underline{t}} \tilde{\omega}_2) \wedge dt$. Finally, we have

$$i_{f_* \underline{t}} \tilde{\omega}_1 = i_{\tilde{X}_{K_f}} \tilde{\omega}_1 = dK_f - \frac{\partial K_f}{\partial t} dt\,,$$

which allows us to conclude that $f_* \tilde{\omega}_1 = \omega_{K_f}$ or equivalently that f is canonical. □

Generating functions Let $(\mathcal{P}_1, \omega_1)$ and $(\mathcal{P}_2, \omega_2)$ be symplectic manifolds and $(\mathbb{R} \times \mathcal{P}_i, \tilde{\omega}_i)$ the corresponding contact manifolds. Consider a canonical transformation f from $\mathbb{R} \times \mathcal{P}_1$ to $\mathbb{R} \times \mathcal{P}_2$ and denote by Γ_f the graph of f and by $\tilde{\pi}_i : \mathbb{R} \times \mathcal{P}_1 \times \mathcal{P}_2 \to \mathbb{R} \times \mathcal{P}_i$ the projection onto $\mathbb{R} \times \mathcal{P}_i$. We define g_s such that $f(s,x) = (s, g_s(x))$ and identify elements of Γ_f, $((s,x), f(s,x)) \in (\mathbb{R} \times \mathcal{P}_1) \times (\mathbb{R} \times \mathcal{P}_2)$ with elements of $(\mathbb{R} \times \mathcal{P}_1) \times \mathcal{P}_2$ of the form $((s,x), g_s(x))$. The graph of f is thus given by

$$\Gamma_f = \{((s,x), g_s(x)) \in (\mathbb{R} \times \mathcal{P}_1) \times \mathcal{P}_2\}\ .$$

We also define the inclusion map $i_f : \Gamma_f \to \mathbb{R} \times \mathcal{P}_1 \times \mathcal{P}_2$.

[4] Using the previous lemma.

Proposition II.51. *Let*

$$\Omega = \tilde{\pi}_1^* \tilde{\omega}_1 - \tilde{\pi}_2^* \omega_{K_f}$$

be a two form on $\mathbb{R} \times \mathcal{P}_1 \times \mathcal{P}_2$ where K_f is defined as in (C3). Then $i_f^* \Omega = 0$.

Moreover, if $\tilde{\omega}_i$ and ω_{K_f} are defined as in (C4), we have:

$$d(i_f^*(\tilde{\pi}_1^* \theta_1 - \tilde{\pi}_2^* \theta_{K_f})) = 0\,.$$

Proof. The proof of this property is similar to the one in the time-independent case. We take $((s,x), g(x)) \in \Gamma_f$ and $((s_i, v_i), (Tg_s \cdot v_i)) \in T_{((s,x),g(x))}\Gamma_f$ and proceed to the computation of $i_f^* \Omega$:

$$i_f^* \Omega((s,x), g(x))(((s_1, v_1), (Tg_{s_1} \cdot v_1)), ((s_2, v_2), (Tg_{s_2} \cdot v_2))) =$$
$$(\tilde{\omega}_1 - f^* \omega_{K_f})(s,x)((s_1, v_1), (s_2, v_2))\,.$$

Therefore $i_f \Omega = 0$.

The second part of the theorem requires only substitutions. □

From the Poincaré lemma there exists locally a one-form Θ such that $\Omega = -d\Theta$. Thus, $i_f^* \Theta$ is closed and there exists locally a function F such that

$$i_f^* \Theta = dF\,.$$

Definition II.52. *F as defined above is called a generating function for f. F is locally defined and is not unique.*

Depending on the choice of the canonical one-form Θ, F takes different expressions and we can recover all 4^n kinds of generating functions. However, we do not give details of the derivation of each of the generating functions as it proceeds exactly as in the autonomous case. We now move on the Hamilton-Jacobi for time-dependent canonical transformations applied to non-autonomous Hamiltonian systems.

The Hamilton-Jacobi theory Even though the following theorems may not be new, we could not find them in the literature.

Proposition II.53. *Let f be a canonical transformation and F its associated generating function, then*

$$f^*K_f = \frac{\partial F}{\partial t}\,.$$

In addition, for a Hamiltonian H on $\mathbb{R} \times \mathcal{P}_1$,

$$f_*\tilde{X}_H = \tilde{X}_K\,,\ \textit{where}\ f^*K = H + \frac{\partial F}{\partial t}\,.$$

Proof. The definition of F reads:

$$\begin{aligned} dF &= i_f^*\Theta \\ &= i_f^*(\tilde{\pi}_1^*\tilde{\theta}_1 - \tilde{\pi}_2^*\theta_{K_f}) \\ &= i_f^*(\tilde{\pi}_1^*dt + \tilde{\pi}_1^*\pi^*\theta_1 - \tilde{\pi}_2^*dt - \tilde{\pi}_2^*\pi^*\theta_2 - \tilde{\pi}_2^*K_f dt) \\ &= i_f^*\tilde{\pi}_1^*(\pi^*\theta_1 - f^*\pi^*\theta_2 - f^*K_f dt)\,. \end{aligned}$$

Therefore, $f^*K_f = \frac{\partial F}{\partial t}$. The remainder of the proposition is just (C6) (cf. Jacobi's theorem (Thm. II.50)) □

The next two theorems focus on canonical transformations that transform the system to equilibrium. They are the main results of this section: the last one is the Hamilton-Jacobi theorem.

Let $\mathcal{Q}$ be the configuration space of the Hamiltonian system defined by H and consider a canonical transformation f on $\mathbb{R} \times T^*\mathcal{Q}$.

Definition II.54. *We say that f transforms H to equilibrium if $K = f_*H + K_f = constant$.*

Theorem II.55. *Let $F : \mathbb{R} \times \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ be a smooth function. Define $\tilde{p}_t(q, Q) = \frac{\partial F}{\partial q}(t, q, Q)$ and $\tilde{P}_t(q, Q) = -\frac{\partial F}{\partial Q}(t, q, Q)$. Then the following two conditions are equivalent:*

1. *F is the generating function associated with f;*

2. - *For every curve $c(t)$ in $\mathcal{Q}$ satisfying:*

   $$c'(t) = T\tau_{\mathcal{Q}}^* \cdot X_{H_t}(c(t), \tilde{p}_t(c(t), Q)) \, ,$$

   the curve $t \mapsto (c(t), \tilde{p}_t(c(t), Q))$ is an integral curve of X_H, where $\tau_{\mathcal{Q}}^* : T^*\mathcal{Q} \to \mathcal{Q}$ is the cotangent bundle projection.

   - *For every curve $c(t)$ in $\mathcal{Q}$ satisfying:*

   $$c'(t) = T\tau_{\mathcal{Q}}^* \cdot X_K(c(t), \tilde{P}_t(q, c(t))) \, ,$$

   the curve $t \mapsto (c(t), \tilde{P}_t(q, c(t)))$ is an integral curve of X_K.

Proof. The proof is the same as in the autonomous case, so we omit it. □

Theorem II.56 (Hamilton-Jacobi). *Let $F : \mathbb{R} \times \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ be a generating function associated with f. Then f transforms H to equilibrium if and only if $H(q, \frac{\partial F}{\partial q}) + \frac{\partial F}{\partial t} = constant$.*

Proof. If f transforms H to equilibrium, then from Thm. II.50, $H + \frac{\partial F}{\partial t} = constant$. Suppose now that F verifies the Hamilton-Jacobi equation, then again Thm. II.50 allows us to conclude that $K = constant$. □

CHAPTER III

SOLVING TWO-POINT BOUNDARY VALUE PROBLEMS

One of the most famous two-point boundary value problems in astrodynamics is Lambert's problem, which consists of finding a trajectory in the two-body problem which goes through two given points in a given time. Even though the two-body problem is integrable, no closed-form solution has been found to this problem so far. Solving Lambert's problem still requires one to solve Kepler's equation, which has motivated many papers since 1650 (see e.g. Colwell [24]). As mentioned in the introduction, for a general Hamiltonian dynamical system, a two-point boundary value problem is solved using iterative methods that require a "good" initial guess for convergence. Though very systematic, these techniques are not appropriate when several boundary value problems need to be solved as they require excessive computation and time. For example, in order to design a change of configuration of a formation of N spacecraft, $N!$ two-point boundary value problems need to be solved [94]. As N increases, the number of boundary value problems dramatically grows.

The novel approach we propose in this dissertation addresses these limitations. Specifically, it allows us to formally solve any kind of "symmetric" two-point boundary value problems with no need for an initial guess and at the cost of a single function evaluation once the generating functions are known ("symmetric" boundary value problems refer to boundary value problems for which the same number of initial and final states are specified. In the following, we restrict ourselves to those problems and simply refer to them as two-point boundary value problems).

Our method is based on the Hamilton-Jacobi theory (Chapter II). We consider the transformation that maps the state of a Hamiltonian system at time t to its initial state. Such a transformation is canonical and transforms the system to an equilibrium, i.e., to its initial conditions which are constants of motion. As a result, the Hamilton-Jacobi theorem tells us that there exist generating functions associated with this transformation that verify the Hamilton-Jacobi equation (Eq. (2.34)). These generating functions have distinctive properties that we now study.

This chapter is organized as follows. We first establish the canonical nature of the transformation that maps the state of a system to its initial state. Then we focus on the generating functions associated with this transformation. We prove that they solve two-point boundary value problems and analyze their properties. Specifically, for linear systems we show that generating functions and state transition matrices are closely related. The state transition matrix allows one to predict singularities of the generating functions whereas the generating functions provide information on the structure of the state transition matrix. This relationship also allows us to recover and extend some results on the perturbation matrices developed by Battin in [10]. For nonlinear systems, generating functions may also develop singularities (called caustics). Using the geometric framework introduced in Chapter II together with the Legendre transformation, we propose a technique to study the geometry of these caustics. We illustrate our method with the study of the singularities of the F_1 generating function in the Hill three-body problem. Most importantly, we relate the existence of singularities to the presence of multiple solutions to boundary value problems. Finally, we introduce Hamilton's principal function, a function *similar* to the generating functions that also solves two-point boundary value problems. We highlight the differences between Hamilton's function and generating functions and justify our choice of focusing on generating functions.

3.1 The phase flow transformation and its generating functions

In this section, we define the transformation that maps the state of a Hamiltonian system to its initial state and prove that it is canonical. We also summarize the equations verified by the generating functions associated with this transformation.

Consider a Hamiltonian system $(\mathcal{P}, H, \omega)$ and recall Φ_t, the phase flow of the system:

$$\begin{aligned} \Phi_t : P &\rightarrow P \\ (q_0, p_0) &\mapsto \left(\Phi_t^1(q_0, p_0) = q(q_0, p_0, t), \Phi_t^2(q_0, p_0) = p(q_0, p_0, t)\right). \end{aligned} \tag{3.1}$$

Φ_t induces a transformation ϕ on $\mathcal{P} \times \mathbb{R}$ as follows:

$$\phi : (q_0, p_0, t) \mapsto (\Phi_t(q_0, p_0), t)\,.$$

In other words, ϕ^{-1} transforms the state of the system at time t into its state at the initial time while preserving the time. Let us now prove that ϕ, and *a fortiori* ϕ^{-1}, are canonical transformations.

Proposition III.1. *The transformation ϕ induced by the phase flow is canonical.*

Proof. From the theory of differential equations[1], ϕ is an isomorphism. Moreover, Prop. II.22 states that Φ_t is symplectic. Thus, ϕ is canonical. □

ϕ^{-1} maps the Hamiltonian system to equilibrium. Therefore, the associated generating functions, F_{I_p,K_r}, verify the Hamilton-Jacobi equation (Eq. (2.39)). In addition, they must

[1] Uniqueness of solutions of ordinary differential equations

also verify Eqns. (2.30)-(2.34), where (Q, P) now denotes the initial state (q_0, p_0):

$$p_{I_p} = \frac{\partial F_{I_p,K_r}}{\partial q_{I_p}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t)\,, \tag{3.2}$$

$$q_{\bar{I}_p} = -\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t)\,, \tag{3.3}$$

$$p_{0_{K_r}} = -\frac{\partial F_{I_p,K_r}}{\partial q_{0_{K_r}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t)\,, \tag{3.4}$$

$$q_{0_{\bar{K}_r}} = \frac{\partial F_{I_p,K_r}}{\partial p_{0_{\bar{K}_r}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t)\,, \tag{3.5}$$

$$0 = H(q_{I_p}, -\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}}, \frac{\partial F_{I_p,K_r}}{\partial q_{I_p}}, p_{\bar{I}_p}, t) + \frac{\partial F_{I_p,K_r}}{\partial t}\,. \tag{3.6}$$

Similarly, Eqns. (2.25), (2.35), (2.36) and (2.37) simplifies to:

$$p = \frac{\partial F_1}{\partial q}(q, q_0, t)\,, \tag{3.7}$$

$$p_0 = -\frac{\partial F_1}{\partial q_0}(q, q_0, t)\,, \tag{3.8}$$

$$H(q, \frac{\partial F_1}{\partial q}, t) + \frac{\partial F_1}{\partial t} = 0\,. \tag{3.9}$$

$$p = \frac{\partial F_2}{\partial q}(q, p_0, t)\,, \tag{3.10}$$

$$q_0 = \frac{\partial F_2}{\partial p_0}(q, p_0, t)\,, \tag{3.11}$$

$$H(q, \frac{\partial F_2}{\partial q}, t) + \frac{\partial F_2}{\partial t} = 0\,. \tag{3.12}$$

$$q = -\frac{\partial F_3}{\partial p}(p, q_0, t)\,, \tag{3.13}$$

$$p_0 = -\frac{\partial F_3}{\partial q_0}(p, q_0, t)\,, \tag{3.14}$$

$$H(-\frac{\partial F_3}{\partial p}, p, t) + \frac{\partial F_3}{\partial t} = 0\,. \tag{3.15}$$

$$q = \frac{\partial F_4}{\partial p}(p, p_0, t)\,, \tag{3.16}$$

$$q_0 = -\frac{\partial F_4}{\partial p_0}(p, p_0, t)\,, \tag{3.17}$$

$$H(\frac{\partial F_4}{\partial p}, p, t) + \frac{\partial F_4}{\partial t} = 0\,. \tag{3.18}$$

3.2 Properties of the generating functions

We now study the properties of the generating functions associated with ϕ^{-1}. First we show that they solve any two-point boundary value problem. Then we focus on linear and non-linear systems. Specifically, we relate the state transition matrix to the generating functions and extend some results on perturbation matrices presented by Battin in [10]. Most importantly, we study the singularities of the generating functions and prove that they correspond to multiple solutions to boundary value problems.

3.2.1 Solving a two-point boundary value problem

Consider two points in phase space, $X_0 = (q_0, p_0)$ and $X_1 = (q, p)$, and two partitions of $(1, \cdots, n)$ into two non-intersecting parts, $(i_1, \cdots, i_p)$ $(i_{p+1}, \cdots, i_n)$ and $(k_1, \cdots, k_r)$ $(k_{r+1}, \cdots, k_n)$. A two-point boundary value problem is formulated as follows:
Given $2n$ coordinates $(q_{i_1}, \cdots, q_{i_p}, p_{i_{p+1}}, \cdots, p_{i_n})$ and $(q_{0_{k_1}}, \cdots, q_{0_{k_r}}, p_{0_{k_{r+1}}}, \cdots, p_{0_{k_n}})$, find the remaining $2n$ variables such that a particle starting at X_0 reaches X_1 in T units of time.

From the relationship defined by Eqns. (3.2), (3.3), (3.4) *and* (3.5), *we see that the generating function F_{I_p,K_r} solves this problem. This remark is of prime importance as it provides us with a very general technique for solving any Hamiltonian boundary value problems.*

Example III.2. Lambert's problem is a particular case of a boundary value problem where $p = r = n$. Though, given two positions q and q_0 and a transfer time T, the corresponding momentum vectors are found from Eqns. (3.7) and (3.8):

$$p_i = \frac{\partial F_1}{\partial q_i}(q, q_0, T)\,,\ p_{0_i} = -\frac{\partial F_1}{\partial q_{0_i}}(q, q_0, T)\,.$$

3.2.2 Linear systems theory

In this section we study the generating functions associated with the flow of linear Hamiltonian systems. Specifically, we reduce the Hamilton-Jacobi equation to a set of four matrix ordinary differential equations. Then, we relate the state transition matrix and generating functions. We show that properties of one may be deduced from properties of the other. The theory we present has implications in the study of relative motion and in optimal control theory (Chapter VI).

Hamilton-Jacobi equation

To study the relative motion of two particles, one often linearizes the dynamics about the trajectory (called the reference trajectory) of one of the particles. Then one uses this linear approximation to study the motion of the other particle relative to the reference trajectory (perturbed trajectory). Thus, the dynamics of relative motion reduces at first order to a time-dependent linear Hamiltonian system, i.e., a system with a quadratic Hamiltonian function without any linear terms (Appendix A, Eq. A.10):

$$H^h = \frac{1}{2} X^{hT} \begin{pmatrix} H_{qq}(t) & H_{qp}(t) \\ H_{pq}(t) & H_{pp}(t) \end{pmatrix} X^h \,, \tag{3.19}$$

where $X^h = \left(\begin{smallmatrix} \Delta q \\ \Delta p \end{smallmatrix}\right)$ is the relative state vector.

Lemma III.3. *The generating functions associated with the phase flow transformation of the system defined by Eq.* (3.19) *are quadratic without linear terms.*

The proof of this lemma is trivial once we understand the link between the generating functions and the state transition matrix (see later in the section).

From the above lemma, a general form for F_2 is:

$$F_2 = \frac{1}{2} Y^T \begin{pmatrix} F^2_{11}(t) & F^2_{12}(t) \\ F^2_{21}(t) & F^2_{22}(t) \end{pmatrix} Y \,, \tag{3.20}$$

where $Y = \left(\begin{smallmatrix} \Delta q \\ \Delta p_0 \end{smallmatrix} \right)$ and $\left(\begin{smallmatrix} \Delta q_0 \\ \Delta p_0 \end{smallmatrix} \right)$ is the relative state vector at the initial time. We point out that both matrices defining H^h and F_2 are symmetric by definition. Then Eq. (3.10) reads:

$$\begin{aligned} \Delta p &= \frac{\partial F_2}{\partial \Delta q} \\ &= \begin{pmatrix} F^2_{11}(t) & F^2_{12}(t) \end{pmatrix} Y \,, \end{aligned}$$

Substituting into Eq. (3.12) yields:

$$\begin{aligned} Y^T \Bigg\{ & \begin{pmatrix} \dot{F}^2_{11}(t) & \dot{F}^2_{12}(t) \\ \dot{F}^2_{12}(t)^T & \dot{F}^2_{22}(t) \end{pmatrix} \\ & + \begin{pmatrix} I & F^2_{11}(t)^T \\ 0 & F^2_{12}(t)^T \end{pmatrix} \begin{pmatrix} H_{qq}(t) & H_{qp}(t) \\ H_{pq}(t) & H_{pp}(t) \end{pmatrix} \begin{pmatrix} I & 0 \\ F^2_{11}(t) & F^2_{12}(t) \end{pmatrix} \Bigg\} Y = 0 \,. \end{aligned} \tag{3.21}$$

Though the above equation has been derived using F_2, it is also valid for F_1 (replacing $Y = \left(\begin{smallmatrix} \Delta q \\ \Delta p_0 \end{smallmatrix} \right)$ by $Y = \left(\begin{smallmatrix} \Delta q \\ \Delta q_0 \end{smallmatrix} \right)$) since F_1 and F_2 solve the same Hamilton-Jacobi equation (Eqns. (3.9) and (3.12)). Eq. (3.21) is equivalent to the following four matrix equations:

$$\begin{gathered} \dot{F}^{1,2}_{11}(t) + H_{qq}(t) + H_{qp}(t)F^{1,2}_{11}(t) + F^{1,2}_{11}(t)H_{pq}(t) + F^{1,2}_{11}(t)H_{pp}(t)F^{1,2}_{11}(t) = 0 \,, \\ \dot{F}^{1,2}_{12}(t) + H_{qp}(t)F^{1,2}_{12}(t) + F^{1,2}_{11}(t)H_{pp}(t)F^{1,2}_{12}(t) = 0 \,, \\ \dot{F}^{1,2}_{21}(t) + F^{1,2}_{21}(t)H_{pq}(t) + F^{1,2}_{21}(t)H_{pp}(t)F^{1,2}_{11}(t) = 0 \,, \\ \dot{F}^{1,2}_{22}(t) + F^{1,2}_{21}(t)H_{pp}(t)F^{1,2}_{12}(t) = 0 \,, \end{gathered} \tag{3.22}$$

where we replaced F^2_{ij} by $F^{1,2}_{ij}$ to signify that these equations are valid for both F_1 and F_2. We also recall that $F^{1,2}_{21} = F^{1,2^T}_{12}$. A similar set of equations can be derived for any generating function F_{I_p,K_r}. However, in this section we only give the equations verified by F_3 and F_4:

$$\begin{gathered} \dot{F}^{3,4}_{11}(t) + H_{pp}(t) - H_{pq}(t)F^{3,4}_{11}(t) - F^{3,4}_{11}(t)H_{qp}(t) + F^{3,4}_{11}(t)H_{qq}(t)F^{3,4}_{11}(t) = 0 \,, \\ \dot{F}^{3,4}_{12}(t) - H_{pq}(t)F^{3,4}_{12}(t) + F^{3,4}_{11}(t)H_{qq}(t)F^{3,4}_{12}(t) = 0 \,, \\ \dot{F}^{3,4}_{21}(t) - F^{3,4}_{21}(t)H_{qp}(t) + F^{3,4}_{21}(t)H_{qq}(t)F^{3,4}_{11}(t) = 0 \,, \\ \dot{F}^{3,4}_{22}(t) + F^{3,4}_{21}(t)H_{qq}(t)F^{3,4}_{12}(t) = 0 \,. \end{gathered} \tag{3.23}$$

The first equations of Eqns. (3.22) and (3.23) are Riccati equations. The second and third are non-homogeneous, time varying, linear equations once the Riccati equations are solved and are equivalent to each other (i.e., transform into each other under transpose). The last are just a quadrature once the previous equations are solved.

Initial conditions

Although F_1 and F_2 (or more generally F_{I_p,K_r} and F_{I_p,K_s} for all r and s) verify the same Hamilton-Jacobi partial differential equation, these generating functions are different. This difference is characterized by the boundary conditions. At the initial time, the flow induces the identity transformation, thus the generating functions should also do so. In other words, at the initial time,

$$\Delta q(t_0) = \Delta q_0\,, \quad \Delta p(t_0) = \Delta p_0\,.$$

In terms of generating functions this translates for F_2 to:

$$\frac{\partial F_2}{\partial \Delta q}(\Delta q_0, \Delta p_0, t_0) = \Delta p_0\,, \quad \frac{\partial F_2}{\partial \Delta p_0}(\Delta q_0, \Delta p_0, t_0) = \Delta q_0\,,$$

that is,

$$F^2_{11}\Delta q + F^2_{12}\Delta p_0 = \Delta p_0\,, \quad F^2_{21}\Delta q + F^2_{22}\Delta p_0 = \Delta q_0\,,$$

or equivalently:

$$F^2_{11} = F^2_{22} = 0\,, \quad F^2_{12} = F^2_{21} = Identity\,.$$

On the other hand, F_1 is ill-defined at the initial time. Indeed, at the initial time Eqns. (3.7) and (3.8) read:

$$F^1_{11}\Delta q + F^1_{12}\Delta q_0 = \Delta p_0\,, \quad F^1_{21}\Delta q + F^1_{22}\Delta q_0 = \Delta p_0\,.$$

These equations do not have any solutions. This was expected since at the initial time $(\Delta q, \Delta q_0)$ are not independent variables ($\Delta q = \Delta q_0$).

The same reasoning applies to all 4^n generating functions and in the same manner we can prove that only F_2 and F_3 have well-defined boundary conditions at the initial time. They are the only two kinds of generating functions that can generate the identity transformation. We come back to this important issue in Chapter V.

Legendre transformation

We saw in Chapter II that the Legendre transformation (Eq. (2.28)) allows one to transform one generating function into another. It plays a central role in the present research. It is used to avoid singularities in the algorithm presented in Chapter V. In addition, it allows us to overcome some of the barriers to truly reconfigurable control in optimal control theory (Section 6.3). As an introduction to this technique, we detail in this section the Legendre transformation for transforming F_2 into F_1 for linear systems.

Recall the Legendre transformation:

$$F_1(\Delta q, \Delta q_0, t) = F_2(\Delta q, \Delta p_0, t) - \langle \Delta p_0, \Delta q_0 \rangle \,, \tag{3.24}$$

where Δp_0 is to be viewed as a function of $(\Delta q, \Delta q_0)$. Let us first find $\Delta p_0(\Delta q, \Delta q_0)$. From Eq. (3.11) we have:

$$\begin{aligned} \Delta q_0 &= \frac{\partial F_2}{\partial \Delta p_0} \\ &= F^2_{21} \Delta q + F^2_{22} \Delta p_0 \,. \end{aligned}$$

Solving the above equation for Δp_0 yields:

$$\Delta p_0 = {F^2_{22}}^{-1} \left(\Delta q_0 - F^2_{21} \Delta q \right) \,.$$

We now substitute Δp_0 into Eq. (3.24) and obtain F_1:

$$\begin{aligned} F_1(\Delta q, \Delta q_0, t) &= \frac{1}{2}\Delta q^T F_{11}^2 \Delta q + \Delta q F_{12}^2 {F_{22}^2}^{-1} \left(\Delta q_0 - F_{21}^2 \Delta q\right) \\ &+\frac{1}{2} {F_{22}^2}^{-1} \left(\Delta q_0 - F_{21}^2 \Delta q\right)^T {F_{22}^2}^{-T} F_{22}^2 {F_{22}^2}^{-1} \left(\Delta q_0 - F_{21}^2 \Delta q\right) \\ &-{\Delta q_0}^T {F_{22}^2}^{-1} \left(\Delta q_0 - F_{21}^2 \Delta q\right) \\ &= \frac{1}{2} Y_1^T \begin{pmatrix} F_{11}^1(t) & F_{12}^1(t) \\ F_{21}^1(t) & F_{22}^1(t) \end{pmatrix} Y_1 \,, \end{aligned}$$

where $Y_1 = \begin{pmatrix} \Delta q \\ \Delta q_0 \end{pmatrix}$ and

$$\begin{cases} F_{11}^1 &= F_{11}^2 - F_{12}^2 {F_{22}^2}^{-1} F_{21}^2 \,, \\ F_{12}^1 &= F_{12}^2 {F_{22}^2}^{-1} \,, \\ F_{21}^1 &= {F_{22}^2}^{-1} F_{21}^2 \,, \\ F_{22}^1 &= -{F_{22}^2}^{-1} \,. \end{cases}$$

Therefore, using the Legendre transformation we are able to find a closed-form expression of F_1 from knowledge of F_2, at the cost of one matrix inversion only. This result generalizes to some extent to nonlinear systems as we show in Section 3.2.3.

Perturbation matrices

Another approach for studying relative motion at linear order relies on the state transition matrix. This method is developed by Battin in the textbook “An Introduction to the Mathematics and Methods of Astrodynamics” [10] for the case of a spacecraft moving in a point mass gravity field. Let Φ be the state transition matrix which describes the relative motion:

$$\begin{pmatrix} \Delta q \\ \Delta p \end{pmatrix} = \Phi \begin{pmatrix} \Delta q_0 \\ \Delta p_0 \end{pmatrix} \,,$$

where $\Phi = \begin{pmatrix} \Phi_{qq} & \Phi_{qp} \\ \Phi_{pq} & \Phi_{pp} \end{pmatrix}$. Battin[10] defines the fundamental perturbation matrices C and $\tilde{C}$ as:

$$\begin{aligned} \tilde{C} &= \Phi_{pq}{\Phi_{qq}}^{-1}\,, \\ C &= \Phi_{pp}{\Phi_{qp}}^{-1}\,. \end{aligned}$$

That is, given $\Delta p_0 = 0$, $\tilde{C}\Delta q = \Delta p$ and given $\Delta q_0 = 0$, $C\Delta q = \Delta p$. He shows that for the relative motion of a spacecraft about a circular trajectory in a point mass gravity field the perturbation matrices verify a Riccati equation and are therefore symmetric. Using the generating functions for the canonical transformation induced by the phase flow, we immediately recover these properties. We also generalize these results to any linear Hamiltonian system.

Using the notations of Eq. (3.20), Eqns. (3.10) and (3.11) read:

$$\begin{aligned} \Delta p &= \frac{\partial F_2}{\partial \Delta q} \\ &= F^2_{11}\Delta q + F^2_{12}\Delta p_0\,, \\ \Delta q_0 &= \frac{\partial F_2}{\partial \Delta p_0} \\ &= F^2_{21}\Delta q + F^2_{22}\Delta p_0\,. \end{aligned}$$

We solve for $(\Delta q, \Delta p)$:

$$\Delta q = {F^2_{21}}^{-1}\Delta q_0 - {F^2_{21}}^{-1}F^2_{22}\Delta p_0\,,$$

$$\Delta p = F^2_{11}{F^2_{21}}^{-1}\Delta q_0 + (F^2_{12} - F^2_{11}{F^2_{21}}^{-1}F^2_{22})\Delta p_0\,,$$

and identify the right hand side with the state transition matrix:

$$\left\{ \begin{aligned} \Phi_{qp} &= -{F^2_{21}}^{-1}F^2_{22}\,, \\ \Phi_{qq} &= {F^2_{21}}^{-1}\,, \\ \Phi_{pp} &= F^2_{12} - F^2_{11}{F^2_{21}}^{-1}F^2_{22}\,, \\ \Phi_{pq} &= F^2_{11}{F^2_{21}}^{-1}\,. \end{aligned} \right.$$

We conclude that

$$\tilde{C} = \Phi_{pq}\Phi_{qq}^{-1} = F_{11}^{2}\,. \tag{3.25}$$

In the same manner, but using F_1, we can show that:

$$C = \Phi_{pp}\Phi_{qp}^{-1} = F_{11}^{1}\,. \tag{3.26}$$

Thus, C and $\tilde{C}$ are symmetric by nature (as $F_{11}^{1,2}$ is symmetric by definition) and they verify the Riccati equation given in Eq. (3.22).

Singularities of generating functions and their relation to the state transition matrix

In Chapter II we presented the Hamilton-Jacobi theory. The results we gave there were local and did not concern the global behavior of the generating functions. We proved that at least one of the generating functions is well-defined at every instant (Prop. II.31). In general, we can notice that each of them can become singular at some point, even for simple systems. As an example let us look at the harmonic oscillator.

Example III.4. The Hamiltonian for the harmonic oscillator is given by:

$$H(q,p) = \frac{1}{2m}p^2 + \frac{k}{2}q^2\,,$$

The F_1 generating function for the phase flow canonical transformation can be found to be:

$$F_1(q,q_0,t) = \frac{1}{2}\sqrt{km}\,\csc(\omega t)\left(-2qq_0 + (q^2+q_0^2)\cos(\omega t)\right)\,,$$

where $\omega = \sqrt{\frac{k}{m}}$. One can readily verify that F_1 is a solution of the Hamilton-Jacobi equation (Eq. (3.6)). Although it is well-defined most of the time, at $T = m\pi/\omega$, $m \in \mathbb{Z}$, F_1 becomes singular in that the values of the coefficients of the q's and q_0's increase without bound. To understand these singularities, recall the general solution to the equations of motion:

$$q(t) = q_0 \cos(\omega t) + p_0/\omega \sin(\omega t)\,,$$

$$p(t) = -q_0\omega \sin(\omega t) + p_0 \cos(\omega t)\,.$$

At $t = T$, $q(T) = q_0$, that is q and q_0 are not independent variables. Therefore the generating function F_1 is undefined at this instant. We say that it is singular at $t = T$. However, F_1 may be defined in the limit: at $t = T$, $q = q_0$, and thus F_1 behaves as $m\frac{(q-q_0)^2}{2(t-T)}$ as $t \mapsto T$. Finally, at $t = T$, q is equal to q_0 whatever values p and p_0 take, i.e., singularities correspond to multiple solutions to the boundary value problems.

The harmonic oscillator is a useful example. Since the flow is known analytically, we are able to explicitly illustrate the relationship between the generating functions and the phase flow ϕ. We can go a step further by noticing that both the state transition matrix and the generating functions generate the flow. Therefore, singularities of the generating functions should be related to properties of the state transition matrix:

$$\begin{aligned}
\Delta p &= \frac{\partial F_2}{\partial \Delta p} \\
&= F^2_{11}\Delta q + F^2_{12}\Delta p_0\,, \\
&\ \text{but we also have} \\
\Delta p &= \Phi_{pq}\Phi_{qq}^{-1}\Delta q + (\Phi_{pp} - \Phi_{pq}\Phi_{qq}^{-1}\Phi_{qp})\Delta p_0\,. \\
&\ \text{Similarly,} \qquad\qquad (3.27) \\
\Delta q_0 &= \frac{\partial F_2}{\partial \Delta p_0} \\
&= F^2_{21}\Delta q + F^2_{22}\Delta p_0\,, \\
&\ \text{but we also have} \\
\Delta q_0 &= \Phi_{qq}^{-1}\Delta q - \Phi_{qq}^{-1}\Phi_{qp}\Delta p_0\,. \qquad (3.28)
\end{aligned}$$

A direct identification yields:

$$F^2_{11} = \Phi_{pq}\Phi^{-1}_{qq}\,, \tag{3.29}$$

$$F^2_{12} = \Phi_{pp} - \Phi_{pq}\Phi^{-1}_{qq}\Phi_{qp}\,, \tag{3.30}$$

$$F^2_{21} = \Phi^{-1}_{qq}\,, \tag{3.31}$$

$$F^2_{22} = \Phi^{-1}_{qq}\Phi_{qp}\,. \tag{3.32}$$

Thus, F_2 is singular when and only when ϕ_{qq} is not invertible. This relation between singularities of F_2 and invertibility of a sub-matrix of the state transition matrix readily generalizes to other kinds of generating functions. In particular, we can show that

- F_1 is singular when Φ_{qp} is singular,
- F_2 is singular when Φ_{qq} is singular,
- F_3 is singular when Φ_{pp} is singular,
- F_4 is singular when Φ_{pq} is singular.

To extend these results to other generating functions, we must consider other block decompositions of the state transition matrix. Every $n \times n$ block of the state transition matrix is associated with a different generating function. Since the determinant of the state transition matrix is 1, there exists at least one $n \times n$ sub-matrix that must have a non-zero determinant. The generating function associated with this block is non-singular, and we recover Prop. II.31 for linear systems.

3.2.3 Nonlinear systems theory

We have proved the local existence of generating functions and mentioned that they may not be globally defined. Using linear systems theory we are also able to predict where the singularities are and to interpret their meaning as multiple solutions to the two-point boundary value problem. In this section we generalize these results to singularities of nonlinear systems.

The following proposition relates singularities of the generating functions to the invertibility of sub-matrices of the Jacobi matrix of the canonical transformation.

Proposition III.5. *The generating function F_{I_p,K_r} for the canonical transformation ϕ is singular at time t if and only if*

$$\det\left(\frac{\partial\phi_i}{\partial z_j}\right)_{i\in I,j\in J} = 0\,, \tag{3.33}$$

where $I = \{i \in I_p\}\bigcup\{n+i, i \in \bar{I}_p\}$, $J = \{j \in \bar{K}_r\}\bigcup\{n+j, j \in K_r\}$ and $z = (q_0, p_0)$ is the state vector at the initial time.

Proof. For the sake of clarity, let us prove this property for F_1. In that case, $I = [1, n]$ and $J = [n + 1, 2n]$. First we remark that

$$\left(\frac{\partial\phi_i}{\partial z_j}\right)_{i\in I,j\in J} = \left(\frac{\partial q_i}{\partial p_{0_j}}\right)\,.$$

Thus, from the inversion theorem, if $\det\left(\frac{\partial\phi_i}{\partial z_j}\right)_{i\in I,j\in J} = 0$, there is no open set in which we can solve p_0 as a function of q and q_0.

On the other hand, suppose that F_1 is non singular. Then, from Eq. (3.8), we have:

$$p_0 = -\frac{\partial F_1}{\partial q_0}(q, q_0, t)\,, \tag{3.34}$$

that is, we can express p_0 as a function of (q, q_0). This is in contradiction with the result obtained form the local inversion theorem. Therefore, F_1 is singular. □

Example III.6. From the above proposition, we conclude that the F_1 generating function associated with the phase flow of the harmonic oscillator is singular if and only if:

$$\det\left(\frac{\partial\phi_i}{\partial z_j}\right)_{i\in I,j\in J} = 0\,. \tag{3.35}$$

In this example, $I = 1$, $J = 2$ and $\phi = (q_0 \cos(\omega t) + p_0/\omega \sin(\omega t), -q_0\omega \sin(\omega t) + p_0 \cos(\omega t))$. Therefore F_1 is singular if and only if $\sin(\omega t) = 0$, i.e., $t = 2\pi/\omega + 2k\pi$. We recover previous results obtained by direct computation of F_1.

Prop. III.5 generalizes to nonlinear systems the relation between singularities and non-uniqueness of the solutions to boundary value problems. Indeed, F_{I_p,K_r} is singular if and only if $z_J \mapsto \phi_I(t,z)$ is not an isomorphism. By definition of the flow, $z_J \mapsto \phi_I(t,z)$ is surjective. Thus it is not injective, that is, singularities arise when there exist multiple solutions to the boundary value problem.

To study the singularities of nonlinear systems, we need to introduce the concept of Lagrangian submanifolds. The theory of Lagrangian submanifolds goes far beyond the results we present in this section: “Some believe that the Lagrangian submanifold approach will give deeper insight into quantum theories than does the Poisson algebra approach. In any case, it gives deeper insight into classical mechanics and classical field theories” (Abraham and Marsden [1]). We refer to Abraham and Marsden [1], Marsden [66] and Weinstein [95] and references given therein for further information on these subjects.

Lagrangian submanifolds

Consider an arbitrary generating function F_{I_p,K_r}. Then the graph of dF_{I_p,K_r} defines a $2n$-dimensional submanifold called a canonical relation [95] of the $4n$-dimensional symplectic space $(\mathcal{P}_1 \times \mathcal{P}_2, \Omega = \pi_1^*\omega_1 - \pi_2^*\omega_2)$. On the other hand, since the variables (q_0, p_0) do not appear in the Hamilton-Jacobi equation (Eq. (3.6)), we may consider them as parameters. In that case the graph of $(q_{I_p}, p_{\bar{I}_p}) \mapsto dF_{I_p,K_r}$ defines an n-dimensional submanifold of the symplectic space $(\mathcal{P}_1, \omega_1)$ called a Lagrangian submanifold [95]. The study of singularities can be achieved using either canonical relations [1] or Lagrangian submanifolds [5, 66].

Theorem III.7. *The generating function F_{I_p,K_r} is singular if and only if the local projection of the canonical relation $\mathcal{L}$ defined by the graph of dF_{I_p,K_r} onto $(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$ is not a local diffeomorphism.*

Definition III.8. *The projection of a singular point F_{I_p,K_r} onto $(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$ is called a caustic.*

If one works with Lagrangian submanifolds then the previous theorem becomes:

Theorem III.9. *The generating function*[2] *F_{I_p,K_r} is singular if the local projection of the Lagrangian submanifold defined by the graph of $(q_{I_p}, p_{\bar{I}_p}) \mapsto dF_{I_p,K_r}$ onto $(q_{I_p}, p_{\bar{I}_p})$ is not a local diffeomorphism.*

These theorems are the geometric formulation of Prop. III.5. If the projection of the canonical relation defined by the graph of dF_{I_p,K_r} onto $(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$ is not a local diffeomorphism, then there exists multiple solutions to the problem of finding (q_0, p_0, q, p) knowing $(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$. From the local inversion theorem, this is equivalent to Prop. III.5.

In the light of these theorems, we can give a geometrical interpretation to Thm. II.31 on the existence of generating functions. Given a canonical relation $\mathcal{L}$ (or a Lagrangian submanifold) defined by a canonical transformation, there exists a $2n$-dimensional (or n-dimensional) submanifold $\mathcal{M}$ of $\mathcal{P}_1 \times \mathcal{P}_2$ (or $\mathcal{P}_1$) such that the local projection of $\mathcal{L}$ onto $\mathcal{M}$ is a local diffeomorphism.

Study of caustics

To study caustics two approaches, at least, are possible depending on the problem. A good understanding of the physics may provide information very easily. For instance,

[2] We consider here that the generating function is a function of n variables only, and has n parameters.

consider the two-body problem in dimension 2, and the problem of going from a point A to a point B, symmetric with respect to the central body, in a certain lapse of time, T. For certain values of T, the trajectory that links A to B is an ellipse whose perigee and apogee are A and B. Therefore, there are two solutions to this problem depending upon which way the particle is going. In terms of generating functions, we deduce that F_3 is non-singular (there is a unique solution once the final momentum is given) but F_1 is singular (existence of two solutions).

Another method for studying caustics consists of using a known non-singular generating function to define the Lagrangian submanifold $\mathcal{L}$ and then study its projection. A very illustrative example is given by Ehlers and Newman [25]. Using the Hamilton-Jacobi equation they treat the evolution of an ensemble of free particles whose initial momentum distribution is $p = \frac{1}{1+q^2}$. They identify a time t_1 at which F_1 is singular. Then, using a closed-form expression of F_3, they find the equations defining the Lagrangian submanifold at t_1. Its projection can be studied and they eventually find that the caustic is two folds. Nevertheless, such an analysis is not always possible as solutions to the Hamilton-Jacobi equation are usually found numerically, not analytically. In the remainder of this section, we focus on systems with polynomial generating functions. Specifically, we show that, in this case, the generating functions can be computed numerically and we develop a method for studying their caustics.

Suppose we are interested in the relative motion of a particle whose coordinates are (q, p) with respect to another one on a known reference trajectory whose coordinates are (q^0, p^0), both moving in an Hamiltonian field. If both particles stay “close” to each other, we can expand (q, p) as a Taylor series about the reference trajectory. The dynamics of the relative motion is described by the Hamiltonian function H^h (Appendix A, Eq. (A.13)):

$$H^h(X^h, t) = \sum_{p=2}^{N} \sum_{\substack{i_1,\cdots,i_{2n}=0\,, \\ i_1+\cdots+i_{2n}=p}}^{p} \frac{1}{i_1!\cdots i_{2n}!} \frac{\partial^p H}{\partial q_1^{i_1}\cdots\partial q_n^{i_n}\partial p_1^{i_{n+1}}\cdots\partial p_n^{i_{2n}}}(q^0, p^0, t) X_1^{h^{i_1}} \dots X_{2n}^{h\ i_{2n}} \,. \tag{3.36}$$

For sake of clarity, $(\Delta q, \Delta p)$ and $(\Delta q_0, \Delta p_0)$ are replaced by (q, p) and (q_0, p_0) in the following, so that $X^h = \begin{pmatrix} q \\ p \end{pmatrix}$ is the relative state vector. Using the algorithm we present in Chapter V we are able to find an approximation of the generating function F_{I_p,K_r} as a polynomial of order N in its spatial variables with time-dependent coefficients. Once F_{I_p,K_r} is known, we find the other generating functions from the Legendre transformation (Eq. (2.28)), at the cost of a series inversion. If a generating function is singular, the inversion does not have a unique solution and the number of solutions characterizes the caustic. To illustrate this method, let us consider the following example.

Example III.10 (Motion about the Libration point L_2 in the Hill three-body problem). Consider a spacecraft moving about and staying close to the Libration point L_2 in the normalized Hill three-body problem (See Appendix C for a description of the Hill three-body problem). Its relative motion with respect to L_2 is described by the Hamiltonian function H^h (Eq. (3.36), or equivalently Eq. (C.11)) and approximated at order N by truncation of terms of order greater than N in the Taylor series defining H^h. The flow associated with the truncation of H^h defines a canonical transformation. Using the algorithm presented in Chapter V, the associated generating function F_2 can be approximated by a Taylor series expansion of order N:

$$
\begin{aligned}
F_2(q_x, q_y, p_{0_x}, p_{0_y}, t) = f_{11}^2(t)q_x^2 + f_{12}^2(t)q_xq_y + f_{13}^2(t)q_xp_{0_x} + f_{14}^2(t)q_xp_{0_y} \\
+ f_{22}^2(t)q_y^2 + f_{23}^2(t)q_yp_{0_x}(t) + f_{24}^2(t)q_yp_{0_y} \\
+ f_{33}^2(t)p_{0_x}^2 + f_{34}^2(t)p_{0_x}p_{0_y} + f_{44}^2(t)p_{0_y}^2 + r(q_x, q_y, p_{0_x}, p_{0_y}, t)\,,
\end{aligned}
$$

where (q, p, q_0, p_0) are relative position and momenta of the spacecraft with respect to L_2 at t and t_0, the initial time, and r is a polynomial of degree N in its spatial variables with time dependent coefficients and without any quadratic terms. At $T = 1.6822$, F_1 is singular but F_2 is not. Eqns. (3.10) and (3.11) reads:

$$p_x = 2f_{11}^2(T)q_x + f_{12}^2(T)q_y + f_{13}^2(T)p_{0_x} + f_{14}^2(T)p_{0_y} + D_1r(q_x, q_y, p_{0_x}, p_{0_y}, T)\,, \tag{3.37}$$

$$p_y = f_{12}^2(T)q_x + 2f_{22}^2(T)q_y + f_{23}^2(T)p_{0_x} + f_{24}^2(T)p_{0_y} + D_2r(q_x, q_y, p_{0_x}, p_{0_y}, T)\,, \tag{3.38}$$

$$q_{0_x} = f_{13}^2(T)q_x + f_{23}^2(T)q_y + 2f_{33}^2(T)p_{0_x} + f_{34}^2(T)p_{0_y} + D_3r(q_x, q_y, p_{0_x}, p_{0_y}, T)\,, \tag{3.39}$$

$$q_{0_y} = f_{14}^2(T)q_x + f_{24}^2(T)q_y + f_{34}^2(T)p_{0_x} + 2f_{44}^2(T)p_{0_y} + D_4r(q_x, q_y, p_{0_x}, p_{0_y}, T)\,, \tag{3.40}$$

where D_ir represents the derivative of r with respect to its i^{th} variable. Eqns. (3.37)-(3.40) define a canonical relation $\mathcal{L}$. By assumption F_1 is singular, therefore the projection of $\mathcal{L}$ onto (q, q_0) is not a local diffeomorphism and there exists a caustic.

Let us now study this caustic. Eqns. (3.37)-(3.40) provide p and q_0 as a function of (q, p_0), but to characterize the caustic we need to study the projection of the Lagrangian manifold on[3] (q, q_0). Hence, we must express p and p_0 as a function of (q, q_0). F_1 being singular, there are multiple solutions to the problem of finding p and p_0 as a function of (q, q_0), and one valuable piece of information is the number k of such solutions. To find p and p_0 as a function of (q, q_0) we first invert Eqns. (3.39) and (3.40) to express p_0 as a function of (q, q_0). Then we substitute this relation into Eqns. (3.37) and (3.38). The first

[3] Since F_1 is a function of (q, q_0).

step requires a series inversion that can be carried out using the technique developed by Moulton in “Differential equations” [72]. Let us rewrite Eqns. (3.39) and (3.40):

$$2f^2_{33}(T)p_{0_x}+f^2_{34}(T)p_{0_y} = q_{0_x}-f^2_{13}(T)q_x-f^2_{23}(T)q_y-D_3r(q_x,q_y,p_{0_x},p_{0_y},T)\,, \tag{3.41}$$

$$f^2_{34}(T)p_{0_x}+2f^2_{44}(T)p_{0_y} = q_{0_y}-f^2_{14}(T)q_x-f^2_{24}(T)q_y-D_4r(q_x,q_y,p_{0_x},p_{0_y},T)\,. \tag{3.42}$$

The determinant of the coefficients of the linear terms on the left hand side is zero (otherwise there is a unique solution to the series inversion) but each of the coefficients is non-zero, that is, we can solve for p_{0_x} as a function of $(p_{0_y}, q_{0_x}, q_{0_y})$ using Eq. (3.41). Then we substitute this solution into Eq. (3.42) and we obtain an equation of the form

$$R(p_{0_y}, q_{0_x}, q_{0_y}) = 0\,, \tag{3.43}$$

that contains no terms in p_{0_y} alone of the first degree. In addition, R contains a non-zero term of the form $\alpha p^2_{0_y}$, where α is a real number. In this case, Weierstrass proved that there exist two solutions, $p^1_{0_y}$ and $p^2_{0_y}$, to Eq. (3.43).

In the same way, we can study the singularity of F_1 at the initial time. At $t = 0$, F_2 generates the identity transformation, hence $f^2_{33}(0) = f^2_{34}(0) = f^2_{43}(0) = f^2_{44}(0) = 0$. This time there is no non-zero first minor, and we find that there exists infinitely many solutions to the series inversion. Another way to see this is to use the Legendre transformation:

$$F_1(q, q_0, t) = F_2(q, p_0, t) - q_0p_0\,,$$

As t tends toward 0, (q, p) goes to (q_0, p_0) and F_2 converges toward the identity transformation $\lim\limits_{t\to 0} F_2(q, p_0, t) = qp_0 \xrightarrow[t\to 0]{} q_0p_0$. Therefore, as t goes to 0, F_1 also goes to 0, i.e., the projection of $\mathcal{L}$ onto (q, q_0) reduces to a point.

The use of series inversion to quantify the number of solutions to the boundary value problem is a very efficient technique for systems with polynomial generating functions.

From the series inversion theory we know that the uniqueness of the inversion is determined by the linear terms whereas the number of solutions (if many) depends on properties of nonlinear terms (we illustrated this property in the above example). In addition, this technique allows us to study the projection of the canonical relation at the cost of a single matrix inversion only.

In the case where generating functions are (or can be approximated by a) polynomial, we can recover the phase flow (or its approximation) as a polynomial too. For instance, from

$$p_0 = \frac{\partial F_1}{\partial q_0}(q, q_0, t)\,,$$

we can find $q(q_0, p_0)$ at the cost of a series inversion. Then, $q(q_0, p_0)$ together with $p = \frac{\partial F_1}{\partial q}(q, q_0, t)$ define the flow (or its polynomial approximation). This procedure is described in greater detail in Section 5.3.2. We want to point out that only a series inversion (i.e., a matrix inversion and a few substitutions) is necessary for transforming the flow into the generating functions and vice versa. On the other hand, generating functions are well-defined if and only if the transformation from the flow to the generating function has a unique solution (Prop. III.5). From series inversion theory, we conclude that generating functions are well-defined if and only if the inversion of the linear approximation of the flow has a unique solution. Therefore, we have the following property:

Proposition III.11. *Singularities of* ***polynomial*** *generating functions correspond to degeneracy of sub-matrices of the state transition matrix as in the linear case. In other words, using our previous notation,*

- F_1 *is singular when* $\det(\Phi_{qp}) = 0$,
- F_2 *is singular when* $\det(\Phi_{qq}) = 0$,

- *F_3 is singular when* $\det(\Phi_{pp}) = 0$*,*

- *F_4 is singular when* $\det(\Phi_{pq}) = 0$*.*

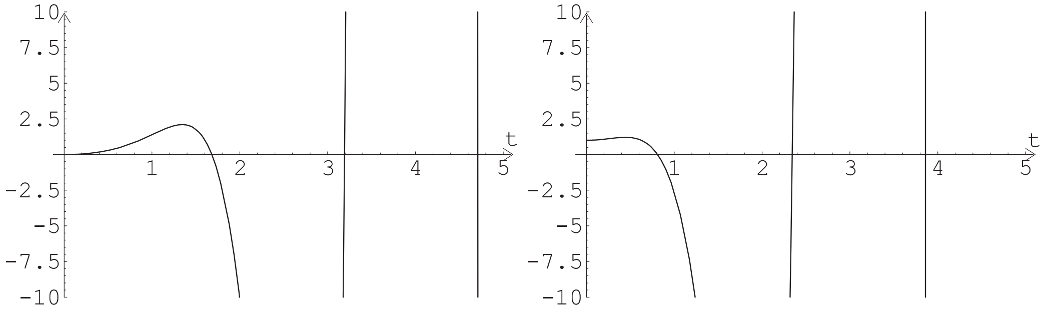
Using other block decompositions of the state transition matrix, these results can be extended to the generating function F_{I_p,K_r}.

Example III.12 (Singularities of the generating functions in the Hill three-body problem). To illustrate Prop. III.11, let us determine the singularities of F_1 and F_2 in the normalized Hill three-body problem linearized about L_2.

The state transition matrix for this problem satisfies (see Appendix C):

$$\dot{\phi}(t) = \begin{pmatrix} -8 & 0 & 0 & -1 \\ 0 & 4 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \phi(t)\,, \quad \phi(0) = Identity\,.$$

We use the $Mathematica^{©}$ built in function $DSolve$ to compute a symbolic expression of the state transition matrix. We plot in Fig. 3.1 the determinant of Φ_{qq} and Φ_{qp} as a function of time. As noticed before F_1 is singular at the initial time and at



(a) Determinant of Φ_{qp} (b) Determinant of Φ_{qq}

Figure 3.1: Determinant of Φ_{qq} and Φ_{qp}

$t = \{1.6821, 3.1938, 4.710\}$ and F_2 is singular at $t = \{0.809, 2.3443, 3.86\}$. The singularity at $t = 1.6821$ was studied above. In addition, one can show that $\det(\Phi_{pp}) = \det(\Phi_{qq})$, i.e., $F_2(q, p_0, t)$ and $F_3(p, q_0, -t)$ have the same singularities.

In the above example, we predicted the singularities of the *nonlinear* generating functions F_1, F_2 and F_3. In particular, we noticed that F_2 and F_3 have the same singularities. This property is specific to this problem. It is the consequence of two results. First, the determinant of the sub-matrices of the state transition matrix corresponding to F_2 and F_3 are invariant under the transformation $t \mapsto -t$. The second result can be formulated as follows:

Proposition III.13. *Consider an autonomous Hamiltonian system. Then the generating functions $F_{I_p,K_r}(t)$ and $F_{K_r,I_p}(-t)$ associated with the phase flow transformation develop singularities at the same instant. For instance, if $p = n$ and $r = 0$, we obtain the fact that F_2 and F_3 have the same singularities.*

Proof. Autonomous Hamiltonian systems are reversible, therefore, the two following boundary value problems are equivalent:

- Going from $(q_{0_{K_r}}, p_{0_{\bar{K}_r}})$ to $(q_{I_p}, p_{\bar{I}_p})$ in T units of time.
- Going from $(q_{0_{I_p}}, p_{0_{\bar{I}_p}})$ to $(q_{K_r}, p_{\bar{K}_r})$ in $-T$ units of time.

As a result, if one of these problems has multiple solutions the other one also has. In other words, if $F_{I_p,K_r}(t)$ is singular, $F_{K_r,I_p}(-t)$ also is. In addition, the caustic for these two generating functions are the same. □

3.3 Hamilton's principal function

Though generating functions are used in the present research to solve boundary value problems, they were introduced by Jacobi, and mostly used thereafter, as fundamental functions which can solve the equations of motion by simple differentiations and eliminations, without integration (Section 2.2.2). Nevertheless, it was Hamilton who first hit upon the idea of finding such a fundamental function. He first proved its existence in geometrical optics (i.e., for time-independent Hamiltonian systems) in 1834 and called it the characteristic function [46]. One year later he published a second essay [47] on systems of attracting and repelling points in which he showed that the evolution of dynamical systems is characterized by a single function called Hamilton's principal function:

> The former Essay contained a general method for reducing all the most important problems of dynamics to the study of one characteristic function, one central or radical relation. It was remarked at the close of that Essay, that many eliminations required by this method in its first conception, might be avoided by a general transformation, introducing the time explicitly into a part S of the whole characteristic function V ; and it is now proposed to fix the attention chiefly on this part S, and to call it the Principal Function. (William R. Hamilton, in the introductory remarks of "Second essay on a General Method in Dynamics" [47]).

Although Hamilton's principal function has been introduced to derive solutions to the equations of motion, it may also be used to solve boundary value problems as well. As far as we know, no one has ever noticed this fact before. Therefore, in the next section we introduce Hamilton's principal function and prove that it solves two-point boundary value problems. Then we discuss how it compares to the generating functions.

3.3.1 Existence of the Hamilton principal function

Similarly to the generating functions, Hamilton's principal function may be derived using the calculus of variations. Consider the extended action integral:

$$A = \int_{\tau_0}^{\tau_1} (pq' + p_t t') d\tau \, , \tag{3.44}$$

under the auxiliary condition $K(q, t, p, p_t) = 0$, where $q' = dq/d\tau$, p_t is the momentum associated with the generalized coordinates t and $K = p_t + H$.

Define a line element[4] $d\sigma$ for the extended configuration space (q, t) by

$$d\sigma = L dt = L t' d\tau \, .$$

Then, we can connect two points (q_0, t_0) and (q_1, t_1) of the extended configuration space by a shortest line γ and measure its length from:

$$A = \int_\gamma d\sigma = \int_\gamma L t' d\tau \, .$$

The distance we obtain is a function of the coordinates of the end-points and, by definition, is given by the Hamilton principal function: $W(q_0, t_0, q_1, t_1)$.

From the calculus of variations (see e.g. Lanczos [60]) we know that the variation of the action A can be expressed as a function of the boundary terms if we vary the limits of the integral:

$$\delta A = p_1 \delta q_1 + p_{t_1} \delta t_1 - p_0 \delta q_0 - p_{t_0} \delta t_0 \, .$$

On the other hand, we have:

$$\delta A = \delta W(q_0, t_0, q_1, t_1) = \frac{\partial W}{\partial q_0} \delta q_0 + \frac{\partial W}{\partial t_0} \delta t_0 + \frac{\partial W}{\partial q_1} \delta q_1 + \frac{\partial W}{\partial t_1} \delta t_1 \, ,$$

[4] The geometry established by this line element is not Riemannian [60]

that is:

$$
\begin{aligned}
p_0 &= -\frac{\partial W}{\partial q_0}(q_0, t_0, q_1, t_1)\,, && (3.45)\\
p_1 &= \frac{\partial W}{\partial q_1}(q_0, t_0, q_1, t_1)\,, && (3.46)
\end{aligned}
$$

and

$$
\begin{aligned}
-\frac{\partial W}{\partial t_0}(q_0, t_0, q_1, t_1) + H(q_0, -\frac{\partial W}{\partial q_0}, t_0) &= 0\,, && (3.47)\\
\frac{\partial W}{\partial t_1}(q_0, t_0, q_1, t_1) + H(q_1, \frac{\partial W}{\partial q_1}, t_1) &= 0\,, && (3.48)
\end{aligned}
$$

where K has been replaced by $p_t + H$. As with generating functions of the first kind, Hamilton's principal function solves boundary value problems of Lambert's type through Eqns. (3.45) and (3.46). To find W, however, we need to solve a system of two partial differential equations (Eqns. (3.47) and (3.48)).

3.3.2 Hamilton's principal function and generating functions

In this section we highlight the main differences between generating functions associated with the phase flow and Hamilton's principal function. For sake of simplicity we compare $F_1(q, q_0, t)$ and $W(q, t, q_0, t_0)$.

Calculus of variations Even if both functions are derived from the calculus of variations, there are fundamental differences between them. To derive generating functions the time t is considered as an independent variable in the variational principle. In contrast, we increase the dimensionality of the system by adding the time t to the generalized coordinates to derive Hamilton's principal function. As a consequence, generating functions generate a transformation between two points in the phase space, i.e., they act without passage of time. On the other hand, Hamilton's principal function generates a transformation between two points in the extended phase space, i.e., between two points in the phase space with different times. This difference may be viewed as follows: Generating functions allow us to characterize the phase flow given an initial time, t_0 (i.e., to characterize all trajectories whose initial conditions are specified at t_0), whereas Hamilton's principal function does not impose any constraint on the initial time. The counterpart being that Hamilton's principal function must satisfy two partial differential equations (Eq. (3.47) defines W as a function of t_0 and Eq. (3.48) defines W as a function of t_1) whereas generating functions satisfy only one.

Moreover, to derive the generating functions fixed endpoints are imposed, that is, we impose the trajectory in both sets of variables to verify the principle of least action. On the other hand, the variation used to derive Hamilton's principal function involves moving endpoints and an energy constraint. This difference may be interpreted as follows: Hamilton's principal function generates a transformation which maps a point of a given energy surface to another point on the same energy surface and is not defined for points that do not lie on this surface. As a consequence of the energy constraint, we have [60]:

$$|\frac{\partial^2 W}{\partial q_0 \partial q_1}| = 0\,. \tag{3.49}$$

As noticed by Lanczos [60], "this is a characteristic property of the W-function which has no equivalent in Jacobi's theory". On the other hand, generating functions map any point of the phase space into another one, the only constraint is imposed through the variational principle (or equivalently by the definition of canonical transformation): we impose the trajectory in both sets of coordinates to be Hamiltonian with Hamiltonian functions H and K respectively.

Fixed initial time In the derivation of Hamilton's principal function dt_0 may be chosen to be zero, that is, the initial time is imposed. Then Hamilton's principal function loses its dependence with respect to t_0. Eq. (3.47) is trivially verified and Eq. (3.49) does not hold anymore, meaning that W and F_1 become equivalent.

Finally, in [47] Hamilton also derives another principal function $Q(p_0, t_0, p_1, t_1)$ which compares to W as F_4 compares to F_1. The derivation being the same we will not go through it.

To conclude, Hamilton's principal function appears to be more general than the generating functions for the canonical transformation induced by the phase flow. On the other hand, the initial and final times are usually specified when solving two-point boundary problems and therefore, any of these functions will identically solve the problem. However, to find Hamilton's principal function we need to solve two partial differential equations with a constraint whereas only one needs to be solved to find the generating functions. For these reasons, generating functions are more appropriate for addressing the problem of solving two-point boundary value problems.

CHAPTER IV

DISCRETE VARIATIONAL PRINCIPLES AND HAMILTON-JACOBI THEORY

In the last two chapters the Hamilton-Jacobi theory as well as a new approach for solving two-point boundary value problems were introduced. This approach relies on knowledge of the generating functions. In general these functions are not known analytically, and so they need to be computed numerically. The purpose of the next two chapters is to understand the numerics of Hamiltonian systems and develop a robust algorithm to compute the generating functions. Such an algorithm needs to address several challenges: 1) the initial conditions for integration are specified in terms of functions with parameters, 2) generating functions may develop singularities that prevent the integration from going further.

In this chapter, we explore the numerics of Hamiltonian systems. In Chapter V, we develop an algorithm for computing the generating functions for a certain class of problems.

The numerics of Hamiltonian systems Hamiltonian systems have a very rich structure and distinctive properties that we can take advantage of. For instance, we have seen (Chapter II) that the energy and the symplectic two-form were invariant along the flow. Most importantly, the invariance of the symplectic two-form gives rise to the Hamilton-Jacobi theory which is the backbone of the approach we developed in Chapter III. Therefore, when simulating Hamiltonian systems we must make sure that at least the symplectic two-form is conserved. The aim of the work presented in this chapter is first to address this issue, but as we will see our results go far beyond our objective.

Standard methods (called numerical integrators) for simulating motion usually take an initial condition and move objects in the direction specified by the differential equations. These methods do not directly satisfy the physical conservation laws associated with the system. An alternative approach to integration, the theory of geometric integrators [68, 20], has been developed over the last two decades. These integrators strictly obey some of these physical laws, and take their name from the law they preserve. For instance, the class of energy-momentum integrators conserves energy and momenta associated with ignorable coordinates. Another class of geometric integrators is the class of symplectic integrators which preserves the symplectic structure. This last class is of particular interest when studying Hamiltonian and Lagrangian systems since the symplectic structure plays a crucial role in these systems (see e.g. Chapter II, Bloch et al. [14], Arnold [5] and Abraham and Marsden [1]). The work done by Wisdom [97, 98] on the n-body problem perfectly illustrates the benefits of such integrators.

At first, symplectic integrators were derived mostly as a subclass of Runge-Kutta algorithms for which the Runge-Kutta coefficients satisfy specific relationships [84]. Such a methodology, though very systematic, does not provide much physical insight and may be limited when we require several laws to be conserved. Other methods were developed in the 90's, among which we may cite the use of generating functions for the canonical transformation induced by the phase flow [21, 26, 45] and the use of discrete variational principles. This last method "gives a comprehensive and unified view on much of the literature on both discrete mechanics as well as integration methods" (Marsden and West [67]). Several discrete variational principles can be found in the literature: Discrete modified Hamilton's principles were introduced by Shibberu [87] and Wu [99] whereas Moser and Veselov [71] and then Marsden, West and Wendlandt [67, 96] developed a fruitful approach based on a discrete Hamilton's principle. Also, Jalnapurkar, Pekarsky and West [54] developed a variational principle on the cotangent bundle based on generating function theory.

In the present research, we focus on the discrete variational principles introduced by Guo, Li and Wu [39, 40, 41] because the theory they have developed provides both a discrete modified Hamilton's principle (DMHP) and a discrete Hamilton's principle (DHP) that are equivalent. We modify and generalize both variational principles they introduce by changing the time discretization so that a suitable analogue of the continuous boundary conditions may be enforced. These boundary conditions are crucial for the analysis of optimal control problems (Section 6.2) and play a fundamental role in dynamics. Our approach not only allows us to obtain a large class of discrete algorithms but it also gives new geometric insight into the Newmark model [73]. Most importantly, using our improved version of the discrete variational principles introduced by Guo et al., we develop a discrete Hamilton-Jacobi theory that yields new results on symplectic integrators. Finally, we derive some properties of symplectic integrators that are of prime importance for building a robust algorithm to compute the generating functions (Chapter V).

In the first part of this chapter (Sections 4.1, 4.2 and 4.3), we present a discrete Hamilton's principle on the tangent bundle and a discrete modified Hamilton's principle on the cotangent bundle (Section 4.1), we discuss the differences with other works on variational integrators (Section 4.2) and show that we are able to recover classical symplectic schemes (Section 4.3). The second part (Sections 4.4 and 4.5) is devoted to issues related to energy conservation and energy error. We first show that by considering time as a generalized coordinate we can ensure energy conservation (Section 4.4). Then we introduce the framework for discrete symplectic geometry and the notion of discrete canonical transformations. We obtain a discrete Hamilton-Jacobi theory that allows us to show that the energy error in the symplectic integration of a dynamical system is invariant under discrete canonical transformations (Section 4.5).

In each part, we illustrate some of the ideas with simulations. In particular we show in the first part that symplectic methods allow one to recover the generating function from the phase flow while standard numerical integrators fail because they do not enforce the necessary exactness condition. In the second part we look at the energy error in the integration of the equations of motion of a particle in a double well potential using a set of coordinates and their transform under a discrete symplectic map.

4.1 Discrete principles of critical action: DMHP and DHP

In this section, we develop a modified version of both variational principles introduced by Guo, Li and Wu [39, 40, 41] and present the geometry associated with them.

4.1.1 Discrete geometry

Consider a discretization of the time t into n instants $\mathcal{T} = \{(t_k)_{k\in[1,n]}\}$. Here $t_{k+1} - t_k$ may not be equal to $t_k - t_{k-1}$ but for sake of simplicity we assume in the following that $t_{k+1} - t_k = \tau$, $\forall k \in [1, n]$. The configuration space at t_k, is the n-dimensional manifold M_k and $\mathcal{M} = \bigcup M_k$ is the configuration space on $\mathcal{T}$. Define a discrete time derivative operator Δ^d_τ on $\mathcal{T}$. Note that Δ^d_τ may not verify the usual Leibnitz law but a modified one. For instance, if we choose Δ^d_τ to be the forward difference operator on $T\mathcal{T}$:

$$\Delta^d_\tau q(t_k) := \frac{1}{\tau}(q(t_k + \tau) - q(t_k)) = \frac{q_{k+1} - q_k}{\tau} := \Delta_\tau q_k \, ,$$

then Δ^d_τ verifies:

$$\Delta^d_\tau(f(t)g(t)) = \Delta^d_\tau f(t) \cdot g(t) + f(t+\tau) \cdot \Delta^d_\tau g(t) \, . \tag{4.1}$$

4.1.2 Discrete Hamilton's principle

Our modified version of the discrete Hamilton's principle derived by Guo, Li and Wu [39] is the discrete time counterpart of Hamilton's principle for Lagrangian systems (Thm. II.5). Consider a discrete curve of points $(q_k)_{k\in[0,n]}$ and a discrete Lagrangian $L_d(q^d_k, \Delta^d_\tau q^d_k)$ where Δ^d_τ is a discrete time derivative operator and q^d_k is a function of (q_k, q_{k+1}).

Definition IV.1 (Discrete Hamilton's principle). *Trajectories of the discrete Lagrangian system L_d going from (t_0, q_0) to (t_n, q_n) correspond to critical points of the discrete action*

$$S^L_d = \sum_{k=0}^{n-1} L_d(q^d_k, \Delta^d_\tau q_k)\tau\,, \tag{4.2}$$

in the class of discrete curves $(q^d_k)_k$ whose ends are (t_0, q_0) and (t_n, q_n). In other words, if we require that the variations of the discrete action S^L_d be zero for any choice of δq^d_k, and $\delta q_0 = \delta q_n = 0$, then we obtain the discrete Euler-Lagrange equations.

Remark IV.2. If we do not impose $t_{k+1} - t_k = t_k - t_{k-1}$, then the discrete action would be defined as:

$$S^L_d = \sum_{k=0}^{n-1} L_d(q^d_k, \Delta^d_\tau q_k)(t_{k+1} - t_k)\,, \tag{4.3}$$

but the discrete Hamilton's principle would be stated in the same manner[1].

To proceed to the derivation of the equations of motion, we need to specify the derivative operator, Δ^d_τ. As we will explain below, its definition depends on the scheme we consider. We should also mention that our variational principle differs from Guo, Li and Wu's since we consider that the action has only finitely many terms and we impose fixed end points. Such a formulation is more in agreement with continuous time variational principles and preserves the fundamental role played by boundary conditions. For a discussion on this topic, we refer to Lanczos [60] Section 15.

[1] In this formulation, the t_k's are known, so there are no additional variables.

4.1.3 Discrete modified Hamilton's principle

As in the continuous case, there exists a discrete variational principle on the cotangent bundle that is equivalent to the above discrete Hamilton's principle (Thm. II.6).

Definition IV.3. *Let L_d be a discrete Lagrangian on $T\mathcal{M}$ and define the discrete Legendre transform (or discrete fiber derivative) $\mathbb{F}L : T\mathcal{M} \to T^*\mathcal{M}$ which maps the discrete state space $T\mathcal{M}$ to $T^*\mathcal{M}$ by*

$$(q_k^d, \Delta_\tau^d q_k^d) \mapsto (q_k^d, p_k^d)\,, \tag{4.4}$$

where

$$p_k^d = \frac{\partial L_d(q_k^d, \Delta_\tau^d q_k^d)}{\partial \Delta_\tau^d q_k^d}\,. \tag{4.5}$$

If the discrete fiber derivative is a local isomorphism, L_d is called regular and if it is a global isomorphism we say that L_d is hyperregular.

If L_d is hyperregular, we define the corresponding discrete Hamiltonian function on $T^*\mathcal{M}$ by

$$H_d(q_k^d, p_k^d) = \langle p_k^d, \Delta_\tau^d q_k^d \rangle - L_d(q_k^d, \Delta_\tau^d q_k^d)\,, \tag{4.6}$$

where $\Delta_\tau^d q_k^d$ is defined implicitly as a function of (q_k^d, p_k^d) through Eq. (4.5). Let S_d^H be the discrete action summation:

$$S_d^H = \sum_{k=0}^{n-1} \left(\langle p_k^d, \Delta_\tau^d q_k^d \rangle - H_d(q_k^d, p_k^d) \right) \tau\,, \tag{4.7}$$

where τ is to be replaced by $t_{k+1} - t_k$ if $t_{k+1} - t_k \neq t_k - t_{k-1}$. Then the discrete principle of least action may be stated as follows:

Definition IV.4 (Discrete modified Hamilton's principle). *Trajectories of the discrete Hamiltonian system H_d going from (t_0, q_0) to (t_n, q_n) correspond to critical points of the* *discrete action S_d^H in the class of discrete curves (q_k^d, p_k^d) whose ends are (t_0, q_0) and (t_n, q_n).*

Again, for deriving the equations of motion we need to specify the discrete derivative operator, Δ_τ^d and its associated Leibnitz law. It will generally depend upon the scheme we consider as we will see through examples later.

4.2 Comparison with other classical variational principle

At this point it is of interest to compare discrete variational principles introduced in this chapter and other classical discrete variational principles. As we mentioned above, the discrete variational principles we develop are inspired by the work of Guo, Li and Wu [39] and we explained above the key difference between our work and this earlier work. We now point out the main differences of the work discussed here with that of Marsden and West, based on the variational principle introduced by Moser and Veselov. In the following, DVPI refers to the discrete variational principle developed by Moser, Veselov, Marsden, Wendlandt et al. whereas DVPII denotes the discrete variational principles developed by Guo and this work.

The first main difference lies in the geometry of both variational principles. Whereas the discrete Lagrangian is a functional on $\mathcal{Q} \times \mathcal{Q}$ where $\mathcal{Q}$ is the configuration space in DVPI, it is a functional on TQ in DVPII. As a consequence, DVPII has a form more like that of the continuous case but has a major drawback: we have to specify the derivative operator and the Leibnitz law it verifies in order to derive the discrete Euler-Lagrange equation. Such a law allows us to perform the discrete counterpart of the integration by parts and depends on the scheme we consider. On the other hand, the Euler-Lagrange equation obtained by DVPI is scheme independent and one benefit is that these equations ensure satisfaction of physical laws such as Noether's theorem for any numerical scheme which can be derived from them.

The next important difference between the two discrete variational principles lies in the role of the Legendre transformation in defining a discrete Hamiltonian function from the discrete Lagrangian. In DVPI, one defines a discrete Legendre transform for computing the momenta from the discrete Lagrangian function, so one may study the discrete dynamics on both $\mathcal{Q} \times \mathcal{Q}$ and $T^*\mathcal{Q}$. However, it does not seem possible to define a discrete Hamiltonian function from the discrete Lagrangian and develop a DMHP. Given a Hamiltonian system, to derive discrete equations of motion using DVPI one needs to first find a continuous Lagrangian function by performing a Legendre transform on the continuous Hamiltonian function, then apply DVPI and finally use the discrete Legendre transform to study the dynamics on $T^*\mathcal{Q}$ (see e.g. Marsden and West [67] page 408). While this point may not be of importance when dealing with dynamical systems, it is crucial if one wants to discretize an optimal control problem, where the continuous Hamiltonian function does not have any physical meaning and the Legendre transformation may not be well-defined (See Section 6.2). In contrast, DVPII naturally defines a discrete Legendre transform and a DMHP.

As mentioned in the introduction, previous researchers have already introduced DMHPs on the cotangent bundle, but, as far as we know, no one has developed an approach that allows one to equivalently consider both the Hamiltonian and Lagrangian approaches in discrete settings (i.e., a DMHP and a DHP that are equivalent for non-degenerate Lagrangian systems). In addition, the DMHPs that can be found in the literature do not allow one to recover most of the classical schemes. For instance, Shibberu's DMHP [87] focuses on the midpoint scheme and Wu [99] developed a different DMHP for each scheme.

Let us now look at some classical schemes and see how they can be derived from DVPII.

4.3 Examples

4.3.1 Störmer's rule and Newmark methods

Störmer's scheme is a symplectic algorithm that was first derived for molecular dynamics problems. It can be viewed as a Runge-Kutta-Nyström method induced by the leap-frog partitioned Runge-Kutta method [84]. The derivation of the Störmer rule as a variational integrator came later and can be found in [99, 96]. Guo, Li and Wu [41] recovered this algorithm using their discrete variational principles. In the next subsection, we briefly go through the derivation and add to their work the velocity Verlet [90] and Newmark methods [67]. In particular, we will show how the conservation of the Lagrangian and symplectic two-forms is built into DVPII.

From the Lagrangian point of view

We first let $q_k^d = q_k$ and define the discrete Lagrangian by $L_d(q_k^d, \Delta_\tau^d q_k) = L(q_k, \Delta_\tau^d q_k)$ and the discrete derivative operator as the forward difference $\Delta_\tau^d = \Delta_\tau$. Δ_τ satisfies the modified Leibnitz law (Eq. (4.1)). Discrete equations of motion are obtained from discrete Hamilton's principle (Def. (IV.1)):

$$\begin{aligned}
\delta S_d^L &= \tau \sum_{k=0}^{n-1} \delta L_d(q_k, \Delta_\tau q_k) \\
&= \tau \sum_{k=0}^{n-1} \langle D_1 L_d(q_k, \Delta_\tau q_k), \delta q_k \rangle + \langle D_2 L_d(q_k, \Delta_\tau q_k), \delta \Delta_\tau q_k \rangle \\
&= \tau \sum_{k=1}^{n-1} \langle D_1 L_d(q_k, \Delta_\tau q_k) - \Delta_\tau D_2 L_d(q_{k-1}, \Delta_\tau q_{k-1}), \delta q_k \rangle \\
&\quad + \Delta_\tau \langle D_2 L_d(q_{k-1}, \Delta_\tau q_{k-1}), \delta q_k \rangle \\
&\quad + \tau \langle D_1 L_d(q_0, \Delta_\tau q_0) \delta q_0 \rangle + \tau \langle D_2 L_d(q_0, \Delta_\tau q_0), \delta \Delta_\tau q_0 \rangle \\
&= \tau \sum_{k=1}^{n-1} \langle D_1 L_d(q_k, \Delta_\tau q_k) - \Delta_\tau D_2 L_d(q_{k-1}, \Delta_\tau q_{k-1}), \delta q_k \rangle - \\
&\quad - \langle D_2 L_d(q_0, \Delta_\tau q_0), \delta q_0 \rangle + \tau \langle D_1 L_d(q_0, \Delta_\tau q_0), \delta q_0 \rangle \\
&\quad + \langle D_2 L_d(q_{n-1}, \Delta_\tau q_{n-1}), \delta q_n \rangle \,, \qquad (4.8)
\end{aligned}$$

where the commutativity of δ and Δ_τ and the modified Leibnitz law defined by Eq. (4.1) have been used.

Discrete Euler-Lagrange equations follow by requiring the variations of the action to be zero for any choice of δq_k, $k \in [1, n-1]$ and $\delta q_0 = \delta q_n = 0$:

$$D_1 L_d(q_k, \Delta_\tau q_k) - \Delta_\tau D_2 L_d(q_{k-1}, \Delta_\tau q_{k-1}) = 0 \,. \qquad (4.9)$$

Suppose $L(q, \dot{q}) = \frac{1}{2} \dot{q} M \dot{q} - V(q)$, then Eq. (4.9) yields Störmer's rule:

$$q_{k+1} = 2q_k - q_{k-1} + h^2 M^{-1}(-\nabla V(q_k)) \,. \qquad (4.10)$$

Consider the one-form[2]

$$\theta_k^L = \frac{\partial L_d(q_{k-1}, \Delta_\tau q_{k-1})}{\partial \Delta_\tau q_{k-1}^i} dq_k^i \,,$$

[2] Einstein's summation convention is assumed

and define the Lagrangian two-form ω_k^L on $T_{q_k}\mathcal{M}$:

$$\begin{aligned}
\omega_k^L &= d\theta_k^L \\
&= \frac{\partial^2 L_d(q_{k-1}, \Delta_\tau q_{k-1})}{\partial q_{k-1}^i \partial \Delta_\tau q_{k-1}^j} dq_k^i \wedge dq_k^j + \frac{\partial^2 L_d(q_{k-1}, \Delta_\tau q_{k-1})}{\partial \Delta_\tau q_{k-1}^i \partial \Delta_\tau q_{k-1}^j} d\Delta_\tau q_k^i \wedge dq_k^j \,.
\end{aligned}$$

Lemma IV.5. *The algorithm defined by Störmer's rule preserves the Lagrangian two-form, ω_k^L.*

Proof. Consider a discrete trajectory $(q_k)_k$ that verifies Eq. (4.10). Then we have:

$$\begin{aligned}
dS_d^L = \tau \sum_{k=1}^{n-1} \left(\frac{\partial L_d(q_k, \Delta_\tau q_k)}{\partial q_k^i} - \Delta_\tau \frac{\partial L_d(q_{k-1}, \Delta_\tau q_{k-1})}{\partial \Delta_\tau^d q_{k-1}^i} \right) dq_k^i \\
+ \Delta_\tau \left(\frac{\partial L_d(q_{k-1}, \Delta_\tau q_{k-1})}{\partial \Delta_\tau q_k^i} dq_k^i \right) \,. \quad (4.11)
\end{aligned}$$

Since the q_k's verify Eq. (4.10), and $d^2 = 0$, Eq. (4.11) yields:

$$d(\Delta_\tau \theta_k^L) = 0 \,, \text{ that is, } \omega_{k+1}^L = \omega_k^L \,. \quad (4.12)$$

We conclude that ω_k^L is preserved along the discrete trajectory □

As we mentioned earlier, because DVPII acts on the tangent bundle it provides results very similar to the continuous case as attested by the form of the Lagrangian two-form. This is to be compared with the Lagrangian two-form arising in the continuous case:

$$\omega^L = \frac{\partial^2 L}{\partial q^i \partial \dot{q}^j} dq^i \wedge dq^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} d\dot{q}^i \wedge dq^j \,.$$

Note that conservation of the Lagrangian two-form is a consequence of using the Leibnitz law, and therefore does not depend on the definition of the discrete Lagrangian. In the remainder of this section we use different discrete Lagrangian functions, but the same Leibnitz law. Thus Lemma IV.5 still applies.

More generally, we can derive Störmer's rule using

$$L_d(q_k, \Delta_\tau q_k) = \lambda L(q_k, \Delta_\tau q_k) + (1 - \lambda) L(q_k + \tau \Delta_\tau q_k, \Delta_\tau q_k) \,,$$

for any λ in $\mathbb{R}$. A particular case of interest is $\lambda = \frac{1}{2}$ which yields a symmetric version of Störmer's rule also called the velocity Verlet method [90]. For this value of λ, we define the associated discrete momenta using the Legendre transform (Eq. (4.5)):

$$\begin{aligned} p_{k+1} &= p_k^d \\ &= D_2 L_d(q_k, \Delta_\tau q_k) \\ &= M\Delta_\tau q_k - \frac{1}{2}\tau\nabla V(q_k + \Delta_\tau q_k)\,, \end{aligned} \tag{4.13}$$

that is:

$$q_{k+1} = q_k + \tau M^{-1}(p_{k+1} + \frac{1}{2}\tau\nabla V(q_{k+1}))\,. \tag{4.14}$$

Moreover, from Eq. (4.9) we obtain:

$$p_{k+1} = p_k + \tau\frac{-\nabla V(q_k) - \nabla V(q_{k+1})}{2}\,. \tag{4.15}$$

Eqns. (4.14) and (4.15) define the velocity Verlet algorithm.

We now focus on the Newmark algorithm which is usually written for the system $L = \frac{1}{2}\dot{q}^T M\dot{q} - V(q)$ as a map given by $(q_k, \dot{q}_k) \mapsto (q_{k+1}, \dot{q}_{k+1})$ satisfying the implicit relations:

$$q_{k+1} = q_k + \tau\dot{q}_k + \frac{\tau^2}{2}[(1-2\beta)a_k + 2\beta a_{k+1}]\,, \tag{4.16}$$

$$\dot{q}_{k+1} = \dot{q}_k + \tau[(1-\gamma)a_k + \gamma a_{k+1}]\,, \tag{4.17}$$

$$a_k = M^{-1}(-\nabla V(q_k))\,, \tag{4.18}$$

where the parameters $\gamma \in [0,1]$ and $\beta \in [0, \frac{1}{2}]$. For $\gamma = \frac{1}{2}$ and any β the Newmark algorithm can be generated from DVPII as a particular case of the Störmer rule where q_k^d and L_d are chosen as follows:

$$q_k^d = q_k - \beta\tau^2 a_k\,,$$

and

$$L_d(q_k^d, \Delta_\tau^d q_k^d) = \frac{1}{2}\dot{q}_k^{d\,T} M \dot{q}_k^d - \tilde{V}(q_k^d)\,,$$

with $\tilde{V}$, the modified potential, satisfying $\nabla\tilde{V}(q_k^d) = \nabla V(q_k)$. Since the derivative operator is the same as above, the discrete Hamilton's principle yields Störmer's equation where q_k is replaced by q_k^d, that is:

$$q_{k+1}^d = 2q_k^d - q_{k-1}^d + \tau^2 M^{-1}(-\nabla\tilde{V}(q_k^d))\,. \tag{4.19}$$

Eq. (4.19) simplifies to

$$q_{k+1} - 2q_k + q_{k-1} = \tau^2(\beta a_{k+2} + (1-2\beta)a_{k+1} + \beta a_{k-1})\,.$$

This last equation corresponds to the Newmark algorithm for the case $\gamma = \frac{1}{2}$. Lemma IV.5 guarantees that the Lagrangian two-form

$$\omega_k^L = d(D_2 L_d(q_k^d, \Delta_\tau^d q_k^d) dq_{k+1}^d)$$

is preserved along the discrete trajectory.

From the Hamiltonian point of view

The Störmer, velocity Verlet, and Newmark algorithms can also be derived using a phase space approach, i.e., the DMHP (Def. IV.4). For Störmer's rule, the Legendre transform yields:

$$p_{k+1} = M\Delta_\tau q_k\,. \tag{4.20}$$

The discrete Hamiltonian function is defined from Eq. (4.6):

$$H_d(q_k, p_{k+1}) = \frac{1}{2}p_{k+1}^T M^{-1} p_{k+1} + V(q_k)\,,$$

and discrete equations of motion are obtained from the DMHP[3] (Def. IV.4). We skip a few steps in the evaluation of the variations of S_d^H to finally find:

$$\begin{aligned}\delta S_d^H &= \delta\left(\tau\sum_{k=0}^{n-1}\langle p_{k+1},\Delta_\tau q_k\rangle - H_d(q_k,p_{k+1})\right)\\ &= \tau\sum_{k=0}^{n-1}\langle \Delta_\tau q_k - D_2H_d(q_k,p_{k+1}),\delta p_{k+1}\rangle - \langle \Delta_\tau p_k + D_1H_d(q_k,p_{k+1}),\delta q_k\rangle\\ &\qquad +\langle p_n,\delta q_n\rangle - \langle p_0,\delta q_0\rangle\,.\end{aligned}$$

If we impose the variations of the action S_d^H to be zero for any $(\delta q_k,\delta p_{k+1})$ and $\delta q_0 = \delta q_n = 0$, we obtain:

$$\left\{\begin{aligned}\Delta_\tau q_k &= p_{k+1}\,,\\ \Delta_\tau p_k &= -\nabla V(q_k)\,.\end{aligned}\right. \tag{4.21}$$

Elimination of the p_k's yields Störmer's rule.

To recover the velocity Verlet scheme from the Hamiltonian point of view, one needs to solve for $\Delta_\tau q_k$ as a function of (q_k,p_{k+1}) in Eq. (4.13). Suppose this has been done and that $\Delta_\tau q_k = f(q_k,p_{k+1})$, then

$$H_d(q_k,p_{k+1}) = \langle p_{k+1}, f(q_k,p_{k+1})\rangle - L_d(q_k,f(q_k,p_{k+1}))\,. \tag{4.22}$$

Taking the variation of the action S_d^H yields the following discrete Hamilton's equations:

$$\Delta_\tau q_k = D_2H_d(q_k,p_{k+1})\,, \tag{4.23}$$

$$\Delta_\tau p_k = -D_1H_d(q_k,p_{k+1})\,. \tag{4.24}$$

On the other hand, Eq. (4.22) provides the following relationships:

$$\begin{aligned}D_1H_d(q_k,p_{k+1}) = \langle D_1 f(q_k,p_{k+1}), p_{k+1} - D_2L_d(q_k,f(q_k,p_{k+1}))\rangle\\ - D_1L_d(q_k,f(q_k,p_{k+1}))\,,\end{aligned} \tag{4.25}$$

[3] $q_k^d = q_k$ and $p_k^d = p_{k+1}$

$$D_2 H_d(q_k, p_{k+1}) = \Delta_\tau^d q_k + \langle D_2 f(q_k, p_{k+1}), p_{k+1} - D_2 L_d(q_k, f(q_k, p_{k+1}))\rangle \,. \quad (4.26)$$

Combining Eqns. (4.23) and (4.24) together with Eqns. (4.25) and (4.26) yields the Velocity Verlet algorithm (Eqns. (4.14) and (4.15)).

We now prove that the scheme we obtained is symplectic. As in the Lagrangian case, the proof differs from the usual one that consists in computing $dp_{k+1} \wedge dq_{k+1}$, in that it relies on fundamental properties of DVPII and on the use of the Leibnitz law.

Lemma IV.6. *The algorithm defined by Eqns. (4.23)-(4.24) is symplectic.*

Proof. We have:

$$\begin{aligned} dS_d^H &= d\left(\tau \sum_{k=0}^{n-1} \langle p_{k+1}, \Delta_\tau q_k\rangle - H_d(q_k, p_{k+1})\right) \,, \\ &= \tau \sum_{k=0}^{n-1} \langle \Delta_\tau q_k - D_2 H_d(q_k, p_{k+1}), dp_{k+1}\rangle - \langle \Delta_\tau p_k + D_1 H_d(q_k, p_{k+1}), dq_k\rangle \\ &\quad + \Delta_\tau \langle p_k, dq_k\rangle \,. \end{aligned}$$

Hence, since (q_k, p_k) verifies Eqns. (4.23)-(4.24) and $d^2 = 0$, we obtain:

$$\Delta_\tau(dp_k \wedge dq_k) = 0 \,.$$

The symplectic two-form $dp_k \wedge dq_k$ is preserved along the trajectory. □

Finally, we can also derive the Newmark methods from the Hamiltonian point of view. The Legendre transform yields:

$$p_k^d = \frac{\partial L_d(q_k^d, \Delta_\tau^d q_k^d)}{\partial \Delta_\tau^d q_k^d} = M \Delta_\tau^d q_k^d \,.$$

The Newmark algorithm is again a particular case of the Störmer rule where (q_k, p_{k+1}) is replaced by (q_k^d, p_k^d):

$$\Delta_\tau^d q_k^d = p_k^d \,, \quad (4.27)$$

$$\Delta_\tau^d p_k^d = -\nabla \tilde{V}(q_k^d) \,. \quad (4.28)$$

Defining $\dot{q}_k$ from p_k as

$$\dot{q}_k = M^{-1}p_k^d + \frac{\tau}{2}a_k$$

allows us to recover the Newmark scheme for $\gamma = \frac{1}{2}$ (Eqns. (4.16) and (4.17)). From the above lemma, we obtain that the symplectic two-form $dp_k^d \wedge dq_k^d$ is preserved along the trajectory.

4.3.2 Midpoint rule

The midpoint rule has been extensively studied and a complete study of its properties can be found in the literature. It is a particular case of the Runge-Kutta algorithm, but can also be derived as a variational integrator (see for instance [99, 87, 67]). The derivation of this scheme has been done by Guo, Li and Wu [41] for the Hamiltonian point of view. In the next section we present the Lagrangian point of view and then recall the Guo, Li and Wu main results, the goal of this section being to illustrate the use of DVPII with other discretization and discrete derivative operator.

From the Lagrangian point of view

Given a Lagrangian $L(q, \dot{q})$, define the discrete Lagrangian by:

$$L_d(q_k^d, \Delta_\tau^d q_k^d) = L(q_k^d, \Delta_\tau^d q_k^d)\,,$$

where $q_k^d = \frac{q_{k+1}+q_k}{2}$, and $\Delta_\tau^d = R_{\tau/2} - R_{-\tau/2}$ where the operator R_τ is the translation by τ. One can readily verify that $\Delta_\tau^d q_k^d = \Delta_\tau q_k$ and that Δ_τ^d verifies the usual Leibnitz law:

$$\Delta_\tau^d(f_k^d g_k^d) = \Delta_\tau^d f_k^d \cdot g_k^d + f_k^d \cdot \Delta_\tau^d g_k^d\,, \tag{4.29}$$

where $f_k = f(t_k)$ and $g_k = g(t_k)$ are functions of time and $f_k^d = \frac{f_{k+1}+f_k}{2}$.

Applying the discrete Hamilton's principle yields:

$$\begin{aligned}\delta S_d^L &= \tau\sum_{k=0}^{n-1}\delta L_d(q_k^d,\Delta_\tau^d q_k^d)\\ &= \tau\sum_{k=0}^{n-1}\langle D_1L_d(q_k^d,\Delta_\tau^d q_k^d),\delta q_k^d\rangle+\langle D_2L_d(q_k^d,\Delta_\tau^d q_k^d),\delta\Delta_\tau^d q_k^d\rangle\,. \end{aligned}\tag{4.30}$$

From the Legendre transform (Eq. (4.5)), we define the associated momentum:

$$\frac{p_{k+1}+p_k}{2}=p_k^d=D_2L_d(q_k^d,\Delta_\tau^d q_k^d)\,.\tag{4.31}$$

Then, Eq. (4.30) becomes:

$$\begin{aligned}\delta S_d^L &= \tau\sum_{k=0}^{n-1}\langle D_1L_d(q_k^d,\Delta_\tau^d q_k^d),\delta q_k^d\rangle+\langle p_k^d,\delta\Delta_\tau^d q_k^d\rangle\\ &= \tau\sum_{k=0}^{n-1}\langle D_1L_d(q_k^d,\Delta_\tau^d q_k^d)-\Delta_\tau^d p_k^d,\delta q_k^d\rangle+\langle p_n,\delta q_n\rangle-\langle p_0,\delta q_0\rangle\,.\end{aligned}$$

If we require the variations of the action to be zero for any choice of $\delta q_k^d,\ k\in[1,n-1]$, and $\delta q_0=\delta q_n=0$, we obtain the discrete Euler-Lagrange equations for the midpoint scheme:

$$\begin{aligned}\frac{p_{k+1}-p_k}{h} &= \Delta_\tau^d p_k^d\\ &= D_1L_d(q_k^d,\Delta_\tau^d q_k^d)\\ &= D_1L_d(\frac{q_{k+1}+q_k}{2},\frac{q_{k+1}-q_k}{h})\,,\end{aligned}\tag{4.32}$$

$$\begin{aligned}\frac{p_{k+1}+p_k}{2} &= p_k^d\\ &= D_2L_d(q_k^d,\Delta_\tau^d q_k^d)\\ &= D_2L_d(\frac{q_{k+1}+q_k}{2},\frac{q_{k+1}-q_k}{h})\,.\end{aligned}\tag{4.33}$$

Lemma IV.7. *The midpoint scheme (Eqns.* (4.32) *and* (4.33)*) defines a symplectic algorithm.*

Proof. The proof proceeds as for the Störmer rule:

$$\begin{aligned} dS_d^L &= \tau \sum_{k=0}^{n-1} \langle D_1 L_d(q_k^d, \Delta_\tau^d q_k^d), dq_k^d \rangle + \langle p_k^d, d\Delta_\tau^d q_k^d \rangle \\ &= \tau \sum_{k=0}^{n-1} \langle D_1 L_d(q_k^d, \Delta_\tau^d q_k^d) - \Delta_\tau^d p_k^d, dq_k^d \rangle + \Delta_\tau^d \langle p_k^d, dq_k^d \rangle \,. \end{aligned}$$

Since $d^2 = 0$ and (q_k, p_k) verifies Eqns. (4.32)-(4.33), we obtain:

$$\Delta_\tau^d (dp_k^d \wedge dq_k^d) = 0 \,.$$

A straightforward computation shows that $\Delta_\tau^d (dp_k^d \wedge dq_k^d) = \Delta_\tau (dp_k \wedge dq_k)$, i.e., the symplectic two-form $\omega_k = dp_k \wedge dq_k$ is preserved along the trajectory. □

From the Hamiltonian point of view

Let $H_d(q_k^d, p_k^d) = H(q_k^d, p_k^d)$ or equivalently define H_d from L_d via Eq. (4.6) and let $(q_k^d, p_k^d) = (\frac{q_{k+1}+q_k}{2}, \frac{p_{k+1}+p_k}{2})$. Then the DMHP (Def. IV.4) yields:

$$\begin{aligned} \frac{q_{k+1} - q_k}{h} &= \Delta_\tau^d p_k^d \\ &= D_2 H_d(q_k^d, p_k^d) \\ &= \frac{\partial H}{\partial p}(\frac{q_{k+1} + q_k}{2}, \frac{p_{k+1} + p_k}{2}) \,, \qquad (4.34) \\ \frac{p_{k+1} - p_k}{h} &= \Delta_\tau^d p_k^d \\ &= -D_1 H_d(q_k^d, p_k^d) \\ &= -\frac{\partial H}{\partial q}(\frac{q_{k+1} + q_k}{2}, \frac{p_{k+1} + p_k}{2}) \,. \qquad (4.35) \end{aligned}$$

Lemma IV.8. *The midpoint scheme defines a symplectic algorithm.*

Proof. The proof is straightforward. We compute $d^2 S_d^H$ assuming (q_k, p_k) verifies the above equations of motion. □

To conclude, we have illustrated the use of the discrete variational principles (Def. IV.1) and (Def. IV.4) and derived discrete equations of motion. One can readily verify that both variational principles yield the same discrete equations, as in the continuous case. Other schemes can be recovered in the same way, and we do not know yet if all classical symplectic algorithms can be derived from DVPII. For instance, we have been able to recover the conditions for the partitioned Runge-Kutta algorithm to be symplectic from the Lagrangian point of view but so far it is not clear to us how it can be done using the Hamiltonian approach (Def. IV.4).

4.3.3 Numerical example

Symplectic integrators are usually used as numerical integrators that preserve the qualitative behavior of dynamical systems and are especially valuable for long time simulations. However, these are not the only uses of symplectic integrators. In this section, we present an aspect of symplectic integrators that we have not seen pointed out in the literature: we show that they allow one to recover the generating functions for the phase flow canonical transformation, whereas numerical integrators do not, even over a short period of time. This remark is of prime importance for deriving a robust algorithm to solve the Hamilton-Jacobi equation (see Chapter V).

Let us first recall Eqns. (3.7) and (3.8) for the F_1 generating function:

$$p = \frac{\partial F_1}{\partial q}\,,\; p_0 = -\frac{\partial F_1}{\partial q_0}\,. \tag{4.36}$$

Eq. (4.36) defines a relationship between the phase flow and the gradient of the generating function. Specifically, if the flow is defined by:

$$\phi : (q_0, p_0, t) \mapsto ((q(t), p(t), t) = (\Phi_t^1(q_0, p_0), \Phi_t^2(q_0, p_0)), t)\,,$$

then, from the local inverse function theorem[4], there exist two functions S_1 and S_2 such

[4] $|\frac{\partial \phi}{\partial p_0}| \neq 0$ since we assume that F_1 exists

that:

$$\begin{aligned} p_0 &= S_1(q, q_0, t)\,, && (4.37)\\ p &= \Phi_t^2(q_0, S_1(q, q_0, t)) \equiv S_2(q, q_0, t)\,. && (4.38)\end{aligned}$$

From Eq. (4.36), we conclude that S_1 and S_2 are the gradient of S and therefore should verify[5]:

$$\frac{\partial^2 F_1}{\partial q_0 \partial q} \equiv \frac{\partial S_1}{\partial q}(q, q_0, t) = \frac{\partial S_2}{\partial q_0}(q, q_0, t) \equiv \frac{\partial^2 F_1}{\partial q \partial q_0}\,. \qquad (4.39)$$

These exactness conditions arise from the symplectic nature of the flow. Therefore, only numerical algorithms that preserve the symplectic two-form (that is symplectic integrators) yield numerical results that agree with Eq. (4.39). Classical numerical integrators fail to provide numerical simulations in agreement with Eq: (4.39).

Example IV.9 (Harmonic Oscillator). The Hamiltonian function for the harmonic oscillator is quadratic:

$$H(q, p) = \frac{1}{2m}p^2 + \frac{k}{2}q^2\,.$$

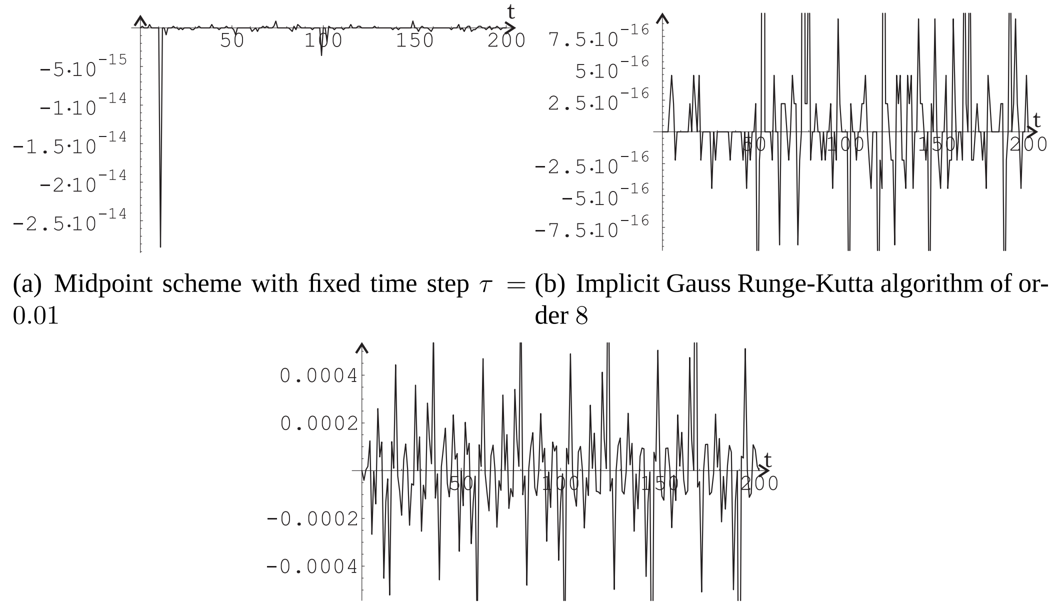
It is a linear system so the phase flow is also linear:

$$\begin{aligned} \Phi_1(q_0, p_0) &= a_{11}(t)q_0 + a_{12}(t)p_0\\ \Phi_2(q_0, p_0) &= a_{21}(t)q_0 + a_{22}(t)p_0\,.\end{aligned}$$

Substituting these expressions into Hamilton's equations (2.1) and balancing terms of the same order yield:

$$\begin{cases} \dot{a}_{11}(t) = a_{21}(t)/m\,,\\ \dot{a}_{12}(t) = a_{22}(t)/m\,,\\ \dot{a}_{21}(t) = ka_{11}(t)\,,\\ \dot{a}_{22}(t) = ka_{12}(t)\,. \end{cases} \qquad (4.40)$$

[5] Since their exists an open set on which the generating functions are smooth, Schwartz's theorem yields $\frac{\partial^2 S}{\partial q_0 \partial q} = \frac{\partial^2 S}{\partial q \partial q_0}$.



(a) Midpoint scheme with fixed time step $\tau = 0.01$

(b) Implicit Gauss Runge-Kutta algorithm of order 8

(c) Explicit Runge-Kutta algorithm of order 8

Figure 4.1: Exactness condition using 3 different integrators

In Fig. 4.1, we plot $\Delta = \frac{\partial S_1}{\partial q}(q, q_0, t) - \frac{\partial S_2}{\partial q_0}(q, q_0, t)$ over the time interval $[0, 100]$ using the symplectic midpoint scheme with fixed time step, a symplectic Gauss implicit Runge-Kutta algorithm of order 8 with fixed time step and a non-symplectic Runge-Kutta integrator of order 8 to integrate Eqns. (4.40). We remark that only symplectic integrators allow us to recover the generating functions because the exactness condition is exactly verified. We point out that even over a short time span, numerical integrators fail to satisfy the exactness condition.

4.4 Energy conservation

Symplectic integrators do not conserve energy and in general induce bounded energy error. There are several works that analyze the energy error, we refer to Hairer and Lubich [44] and Hairer, Lubich and Wanner [45] and references therein for more details. The end of this chapter is devoted to the study of the conservation of energy. In this section, we enhance DVPII so that energy conservation is imposed. By considering the time as a coordinate and by adding an independent parameter τ, DVPII yields symplectic energy conserving algorithms. For certain problems, such algorithms may provide better performance[6], but the contrary may also happen [43, 88]. The method we develop in this section is variational and allows us to recover Shibberu's algorithm [87] for Hamiltonian systems and is equivalent to the Kane, Marsden and Ortiz [56] method for Lagrangian systems. Then, in the next section, we develop a discrete Hamilton-Jacobi theory that defines a class of coordinate transformations that leaves the energy error invariant.

4.4.1 Generalized variational principles

Generalized Hamilton's principle

Let us first recall Hamilton's principle for dynamical systems for which time is considered as a generalized coordinate. Such a formulation is typically used in relativity where the time coordinate is equivalent to the space coordinates.

Consider a Lagrangian $L(q, \dot{q})$ and define the *parametric* Lagrangian

$$\bar{L}(q, t, q', t') = t' L(q, \frac{q'}{t'}, t) \,,$$

where $' = \frac{d}{d\tau}$ and τ is an independent parameter that parameterizes the trajectory and the time. Then the generalized Hamilton's principle reads:

Definition IV.10. *Critical points of* $\int_{t_0}^{t_f} \bar{L}(q, \frac{q'}{t'}, t) d\tau$ *in the class of curves* $(q(\tau), t(\tau))$ *with endpoints* (τ_0, q_0, t_0) *and* (τ_f, q_f, t_f) *correspond to trajectories of the Lagrangian systems going from* (q_0, t_0) *to* (q_f, t_f).

The generalized Hamilton's principle yields the following set of equations:

$$\begin{cases} \frac{\partial \bar{L}}{\partial t} - \frac{d}{d\tau}\frac{\partial \bar{L}}{\partial t'} &= 0 \,, \\ \frac{\partial \bar{L}}{\partial q} - \frac{d}{d\tau}\frac{\partial \bar{L}}{\partial q'} &= 0 \,. \end{cases}$$

[6] To quantify the performance of an algorithm, not only we look at its accuracy but we also evaluate its ability to predict the qualitative behavior of the system. In that sense, symplectic-energy conserving algorithms may not predict qualitative behavior better that symplectic algorithms.

Replacing the parametric Lagrangian by the Lagrangian of the system simplifies the above equations to:

$$t'\frac{\partial L}{\partial t} - \frac{d}{d\tau}L + \frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{q}}\frac{q'}{t'}\right) = 0\,, \tag{4.41}$$

$$t'\frac{\partial L}{\partial q} - \frac{d}{d\tau}\frac{\partial L}{\partial \dot{q}} = 0\,. \tag{4.42}$$

These $n+1$ equations should be compared to the n equations obtained when the trajectory is parameterized by the time:

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0\,. \tag{4.43}$$

Since $\frac{d}{d\tau} = t'\frac{d}{dt}$, we conclude that the space components of the generalized Euler-Lagrange equations (Eq. (4.42)) are a multiple by t' of the original Euler-Lagrange equation (Eq. (4.43)). Also, their time component (Eq. (4.41)) is a linear combination of the components of Eq. (4.43) (the sum of each component multiplied by q'). All $n+1$ generalized Euler-Lagrange equations are thus consistent with the original equations but there is no unique solution because they are satisfied by any parameterization. To get a unique solution, it is necessary to add to the generalized Hamilton's principle an additional condition fixing the parameterization. As we see in the next section, in discrete settings we no longer have this freedom. The discrete counterpart of Eq. (4.41) corresponds to an energy constraint that fully specifies the time parameterization, i.e., the time step.

Generalized discrete Hamilton's principle (GDHM)

In contrast with the variational principles introduced in the first part of this chapter, we do not set the time step, i.e., we let the time act as a variable by adding an independent parameter τ_k such that $t_k = t(\tau_k)$ and $\tau_{k+1} - \tau_k = \tau$, τ being a constant. t_k is now a coordinate that plays the same role as q_k, M_k is the extended configuration space (q_k, t_k),

$\mathcal{M} = \bigcup M_k$ and $\mathcal{T} = \{(\tau_k)_{k\in[1,n]}\}$. Define the modified discrete Lagrangian $\bar{L}_d$:

$$\bar{L}_d(q_k^d, t_k^d, \Delta_\tau^d q_k^d, \Delta_\tau^d t_k^d) = \Delta_\tau^d t_k^d L_d(q_k^d, \frac{\Delta_\tau^d q_k^d}{\Delta_\tau^d t_k^d}, t_k^d)\,,$$

where L_d is the discrete Lagrangian previously defined. In addition, since we are interested in conservation of energy, we only consider systems that are time independent (Prop. II.24). As a consequence, L_d does not depend on time and $\frac{\partial \bar{L}_d}{\partial t_k^d} = 0$.

Definition IV.11 (Generalized Discrete Hamilton's Principle (GDHP)). *Critical points of the discrete action*

$$S_d^L = \sum_{k=0}^{n-1} \bar{L}_d(q_k^d, t_k^d, \Delta_\tau^d q_k^d, \Delta_\tau^d t_k^d)\tau\,,$$

in the class of discrete curves $(q_k^d, t_k^d)_k$ with endpoints (τ_0, t_0, q_0) and (τ_n, t_n, q_n) correspond to discrete trajectories of the Lagrangian system going from (t_0, q_0) to (t_n, q_n).

Again, to proceed to the derivation of the equations of motion we need to specify the derivative operator Δ_τ^d.

Generalized discrete modified Hamilton's principle

Definition IV.12. *Let $\bar{L}_d$ be a discrete Lagrangian on $T\mathcal{M}$ and define the discrete Legendre transform (or discrete fiber derivative) $\mathbb{F}L : T\mathcal{M} \to T^*\mathcal{M}$ which maps the discrete extended phase space $T\mathcal{M}$ to $T^*\mathcal{M}$ by*

$$(q_k^d, t_k^d, \Delta_\tau^d q_k, \Delta_\tau^d t_k^d) \mapsto (q_k^d, t_k^d, p_k^d, e_k^d)\,,$$

where

$$p_k^d = \frac{\partial \bar{L}_d(q_k^d, t_k^d, \Delta_\tau^d q_k^d, \Delta_\tau^d t_k^d)}{\partial \Delta_\tau^d q_k^d}\,, \quad e_k^d = \frac{\partial \bar{L}_d(q_k^d, t_k^d, \Delta_\tau^d q_k^d, \Delta_\tau^d t_k^d)}{\partial \Delta_\tau^d t_k^d}\,. \tag{4.44}$$

The Legendre transform as defined by Eqns. (4.44) is equivalent to the previous definition (Eq. (4.5)). Indeed,

$$\frac{\partial \bar{L}_d(q_k^d, t_k^d, \Delta_\tau^d q_k^d, \Delta_\tau^d t_k^d)}{\partial \Delta_\tau^d q_k^d} = \frac{\partial L_d(q_k^d, \frac{\Delta_\tau^d q_k^d}{\Delta_\tau^d t_k^d})}{\partial \Delta_\tau^d q_k^d} = D_2 L_d(q_k^d, \Delta_t^d q_k^d)\,,$$

where $\Delta_t^d = \frac{\Delta_\tau^d}{\Delta_\tau^d t_k^d}$ represents the discrete derivative with respect to time.

If the discrete fiber derivative is a local isomorphism, $\bar{L}_d$ is called regular and if it is a global isomorphism we say that $\bar{L}_d$ is hyperregular. If $\bar{L}_d$ is hyperregular, we define the corresponding discrete Hamiltonian function on $T^*\mathcal{M}$ by

$$\bar{H}_d(q_k^d, t_k^d, p_k^d, e_k^d) = \langle p_k^d, \Delta_\tau^d q_k^d \rangle - \bar{L}_d(q_k^d, \Delta_\tau^d q_k^d)\,, \tag{4.45}$$

where $\Delta_\tau^d q_k$ is defined implicitly as a function of (q_k^d, p_k^d) through Eq. (4.44). $\bar{H}_d$ is related to the previously defined Hamiltonian function by the following relationship:

$$\bar{H}_d(q_k^d, p_k^d) = \Delta_\tau^d t_k^d H_d(q_k^d, p_k^d)\,.$$

In addition, we have: $e_k^d = -H_d(q_k^d, p_k^d)$, that is, the momentum associated with the time is the opposite of the Hamiltonian.

Let S_d^H be the discrete action summation:

$$\begin{aligned} S_d^H &= \tau \sum_{k=0}^{n-1} \langle p_k^d, \Delta_\tau^d q_k^d \rangle - \bar{H}_d(q_k^d, p_k^d) \\ &= \tau \sum_{k=0}^{n-1} \langle p_k^d, \Delta_\tau^d q_k^d \rangle + \langle e_k^d, \Delta_\tau^d t_k^d \rangle\,. \end{aligned}$$

Before stating the generalized discrete modified Hamilton's principle, we need to remark that all the coordinates are not independent since the *holonomic* constraint $e_k^d = -H(q_k^d, p_k^d)$ holds. There are two ways to handle this situation: one can either replace e_k^d by $-H(q_k^d, p_k^d)$ in the action and then take the variations or one can use Lagrange multipliers to append the constraint $e_k^d + H(q_k^d, p_k^d) = 0$ to the integral. Therefore we can give two equivalent formulations of the GDMHP:

Definition IV.13 (Generalized discrete modified Hamilton's principle). *Critical points of the discrete action*

$$S_d^H = \tau \sum_{k=0}^{n-1} \langle p_k^d, \Delta_\tau^d q_k^d \rangle + \langle e_k^d, \Delta_\tau^d t_k^d \rangle$$

in the class of discrete curves $(q_k^d, t_k^d, p_k^d, e_k^d)$ *with endpoints* (τ_0, t_0, q_0) *and* (τ_n, t_n, q_n), *subject to the constraint* $e_k^d + H_d(q_k^d, p_k^d) = 0$*, correspond to trajectories of the discrete Hamiltonian system going from* (t_0, q_0) *to* (t_n, q_n).

Definition IV.14 (Generalized discrete modified Hamilton's principle). *Critical points of the discrete action*

$$S_d^H = \tau \sum_{k=0}^{n-1} \langle p_k^d, \Delta_\tau^d q_k^d \rangle - H_d(q_k^d, p_k^d) \Delta_\tau^d t_k^d$$

in the class of discrete curves (q_k^d, t_k^d, p_k^d) *with endpoints* (τ_0, t_0, q_0) *and* (τ_n, t_n, q_n) *correspond to trajectories of the discrete Hamiltonian system going from* (t_0, q_0) *to* (t_n, q_n).

Remark IV.15. To prove that these two formulations of the generalized discrete modified Hamilton's principle are equivalent, we only need to remark that the constraint is holonomic. We refer to Bloch et al. [14] for more details.

To derive the equations of motion we need to specify the discrete derivative operator, Δ_τ^d, and its associated Leibnitz law.

4.4.2 Examples

Störmer type of algorithm

Lagrangian approach Consider a Lagrangian function $L(q, \dot{q})$ and define the discrete Lagrangian map trivially by $L_d(q_k, \Delta_\tau q_k) = L(q_k, \Delta_\tau q_k)$. Discrete equations of motion are obtained from the generalized discrete Hamilton's principle:

$$\begin{aligned}
\delta S_d^L &= \tau \sum_{k=0}^{n-1} \delta \bar{L}_d(q_k, t_k, \Delta_\tau q_k, \Delta_\tau t_k) \\
&= \tau \sum_{k=0}^{n-1} \delta(\Delta_\tau t_k L_d(q_k, \frac{\Delta_\tau q_k}{\Delta_\tau t_k})) \\
&= \tau \sum_{k=0}^{n-1} (\delta \Delta_\tau t_k) L_d^k + \Delta_\tau t_k \langle D_1 L_d^k, \delta q_k \rangle \\
&\quad + \Delta_\tau t_k \langle D_2 L_d^k, \frac{\Delta_\tau \delta q_k}{\Delta_\tau t_k} - \frac{\Delta_\tau q_k}{(\Delta_\tau t_k)^2} \delta \Delta_\tau t_k \rangle \,,
\end{aligned}$$

where $L_d^k = L_d(q_k, \frac{\Delta_\tau q_k}{\Delta_\tau t_k})$. Using the Leibnitz law (Eq. (4.1)) and the fixed end point constraint, we obtain:

$$\delta S_d^L = \tau \sum_{k=1}^{n-1} -\Delta_\tau e_k \delta t_k + \langle \Delta_\tau t_k D_1 L_d^k - \Delta_\tau D_2 L_d^{k-1}, \delta q_k \rangle \,, \tag{4.46}$$

where

$$e_{k+1} = \frac{\partial \bar{L}_d^k}{\partial \Delta_\tau t_k} = L_d(q_k, \frac{\Delta_\tau q_k}{\Delta_\tau t_k}) - \langle D_2 L_d(q_k, \frac{\Delta_\tau q_k}{\Delta_\tau t_k}), \frac{\Delta_\tau q_k}{\Delta_\tau t_k} \rangle \,.$$

Finally we obtain the modified Euler-Lagrange equations by setting the variations to zero:

$$\begin{aligned}
e_{k+1} - e_k &= 0 \,, \\
\Delta_\tau t_k D_1 L_d(q_k, \tfrac{\Delta_\tau q_k}{\Delta_\tau t_k}) - \Delta_\tau D_2 L_d(q_{k-1}, \tfrac{\Delta_\tau q_{k-1}}{\Delta_\tau t_{k-1}}) &= 0 \,.
\end{aligned} \tag{4.47}$$

Lemma IV.16. *The algorithm defined by Eqns. (4.47) preserves the Lagrangian two-form and the energy.*

Proof. The first equation of the algorithm proves energy conservation. To show that the Lagrangian two-form is preserved, we compute dS_d^L along a discrete trajectory:

$$\begin{aligned}
dS_d^L &= \tau \sum_{k=1}^{n-1} \Delta_\tau (L_d^{k-1} dt_k) + \Delta_\tau (D_2 L_d^{k-1} dq_k) - \Delta_\tau (\frac{D_2 L_d^{k-1}}{\Delta_\tau t_{k-1}} \Delta_\tau q_{k-1} dt_k) \\
&= \tau \sum_{k=1}^{n-1} \Delta_\tau (e_k dt_k + D_2 L_d^{k-1} dq_k) \\
&= \tau \sum_{k=1}^{n-1} \Delta_\tau \theta_k^L \,,
\end{aligned} \tag{4.48}$$

where $\theta_k^L = e_k dt_k + D_2 L_d^{k-1} dq_k$. Since $d^2 = 0$, we obtain that the symplectic two-forms $\omega_k^L = d\theta_k^L$ is preserved along the trajectory. □

The proof of this lemma only involves the modified Leibnitz law and does not depend on the definition of the discrete Lagrangian function. As a consequence, it also applies to the modified velocity Verlet and Newmark algorithms.

Hamiltonian approach Let the Lagrangian function be $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} - V(q)$. Then

$$\bar{L}_d = \Delta_\tau t_k \left(\frac{1}{2} \frac{\Delta_\tau q_k}{\Delta_\tau t_k} M \frac{\Delta_\tau q_k}{\Delta_\tau t_k} - V(q_k) \right) , \tag{4.49}$$

and the associated momenta are:

$$\begin{aligned} p_{k+1} &= M \tfrac{\Delta_\tau q_k}{\Delta_\tau t_k} , \\ e_{k+1} &= -\tfrac{1}{2} \tfrac{\Delta_\tau q_k}{\Delta_\tau t_k} M \tfrac{\Delta_\tau q_k}{\Delta_\tau t_k} - V(q_k) . \end{aligned}$$

The discrete Hamiltonian function is then:

$$\bar{H}_d = \Delta_\tau t_k (\frac{1}{2} p_{k+1}^T M^{-1} p_{k+1} + V(q_k)) = \Delta_\tau t_k H_d(q_k, p_{k+1}) . \tag{4.50}$$

One can readily verify that $H_d(q_k, p_{k+1}) = -e_{k+1}$.

Let us now derive the modified discrete equations of motion by applying the GDMHP (Thm. (IV.14)). We skip a few steps in the evaluation of the variations of S_d^H to finally find:

$$\begin{aligned} \delta S_d^H &= \tau \delta \sum_{k=0}^{n-1} \langle p_{k+1}, \Delta_\tau q_k \rangle - \bar{H}_d(q_k, p_{k+1}) \\ &= \tau \sum_{k=0}^{n-1} \langle \Delta_\tau q_k - \Delta_\tau t_k D_2 H_d(q_k, p_{k+1}), \delta p_{k+1} \rangle \\ &\quad - \langle \Delta_\tau p_k + \Delta_\tau t_k D_1 H_d(q_k, p_{k+1}), \delta q_k \rangle + \tau \sum_{k=1}^{n-1} \Delta_\tau e_k \delta t_k . \end{aligned}$$

The variations of $(\delta q_k, \delta p_{k+1}, \delta t_k)$ being independent, we obtain:

$$\begin{cases} \Delta_\tau q_k &= \Delta_\tau t_k p_{k+1}\,, \\ \Delta_\tau p_k &= -\Delta_\tau t_k \nabla V(q_k)\,, \\ \Delta_\tau e_k &= 0\,. \end{cases} \tag{4.51}$$

Lemma IV.17. *The algorithm defined by Eqns. (4.51) preserves the symplectic two-form and the energy.*

Proof. The proof proceeds as the previous ones: we compute dS_d^H along a discrete trajectory. We skip the details of the computation:

$$dS_d^H = \tau \sum_{k=0}^{n-1} \Delta_\tau \langle p_k, dq_k \rangle + e_k dt_k\,.$$

Define $\theta_k^H = \langle p_k, dq_k \rangle + e_k dt_k$ and $\omega_k^H = d\theta_k^H$. Since $d^2 = 0$, we obtain that $\Delta_\tau \omega_k^H = 0$. □

Remark IV.18. The one-form θ_k^H corresponds to the contact 1-form θ encountered in continuous time dynamics (Thm. II.21). Indeed, if one remembers that $e_k = -H_d(q_{k-1}, p_k)$, then we have:

$$\begin{aligned} \theta &= pdq - Hdt\,, \\ \theta_k^H &= p_k dq_{k-1} - H_d(q_{k-1}, p_k) dt_k\,. \end{aligned}$$

Midpoint discretization

In the same manner, we can apply the modified variational principle to other discretizations. For the midpoint scheme we have $q_k^d = \frac{q_{k+1}+q_k}{2}$ and the modified Leibnitz rule is defined by Eq. (4.29). Let us define the generalized momenta:

$$\begin{aligned} \frac{p_{k+1}+p_k}{2} = \quad p_k^d \quad &= \frac{\partial \bar{L}_d}{\partial \Delta_\tau^d q_k^d}\,, \\ \frac{e_{k+1}+e_k}{2} = \quad e_k^d \quad &= \frac{\partial \bar{L}_d}{\partial \Delta_\tau^d t_k^d}\,. \end{aligned}$$

Then applying the modified discrete Hamilton's principle (Def. (IV.14)) yields (after a few simplifications):

$$\delta S_d^H = \tau \sum_{k=0}^{n-1} \langle \Delta_\tau^d t_k^d D_1 L_d^k - \Delta_\tau^d p_k^d, \delta q_k^d \rangle - \Delta_\tau^d e_k^d \delta t_k^d \,, \tag{4.52}$$

where $L_d^k = L_d(q_k^d, \frac{\Delta_\tau^d q_k^d}{\Delta_\tau^d t_k^d})$. The variations $(\delta q_k^d, \delta t_k^d)$ being independent, we obtain:

$$\begin{aligned}
\frac{p_{k+1} - p_k}{\tau} &= \frac{t_{k+1} - t_k}{\tau} D_1 L_d(\frac{q_{k+1} + q_k}{2}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k}) \,, \\
e_{k+1} &= e_k \,, \\
\frac{p_{k+1} + p_k}{2} &= \frac{t_{k+1} - t_k}{\tau} D_2 L_d(\frac{q_{k+1} + q_k}{2}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k}) \,, \\
\frac{e_{k+1} + e_k}{2} &= L_d(\frac{q_{k+1} + q_k}{2}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k}) \\
&\quad - \langle D_2 L_d(\frac{q_{k+1} + q_k}{2}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k}), \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \rangle \,.
\end{aligned} \tag{4.53}$$

Lemma IV.19. *The algorithm defined by Eqns. (4.53) preserves the Lagrangian two-form as well as the energy.*

Proof. We omit the proof since it proceeds as before. □

Now define the discrete Hamiltonian function $H_d(q_k^d, p_k^d) = H(\frac{q_{k+1}+q_k}{2}, \frac{p_{k+1}+p_k}{2})$ and the modified Hamiltonian function $\bar{H}_d = \Delta_\tau^d t_k^d H_d(q_k^d, p_k^d)$. Then applying the generalized discrete modified Hamilton's principle yields:

$$\left\{ \begin{aligned}
\tfrac{q_{k+1}-q_k}{\tau} &= \tfrac{t_{k+1}-t_k}{\tau} D_2 H_d(\tfrac{q_{k+1}+q_k}{2}, \tfrac{p_{k+1}+p_k}{2}) \,, \\
\tfrac{p_{k+1}-p_k}{\tau} &= \tfrac{t_{k+1}-t_k}{\tau} D_1 H_d(\tfrac{q_{k+1}+q_k}{2}, \tfrac{p_{k+1}+p_k}{2}) \,, \\
e_{k+1} - e_k &= 0 \,, \\
\tfrac{e_{k+1}+e_k}{2} &= -H_d(\tfrac{q_{k+1}+q_k}{2}, \tfrac{p_{k+1}+p_k}{2}) \,.
\end{aligned} \right. \tag{4.54}$$

Lemma IV.20. *The algorithm defined by Eqns. (4.54) preserves the symplectic two-form as well as the energy.*

Proof. We omit the proof since it proceeds as before. □

The algorithm defined by Eqns. (4.54) is the same as the one developed by Shibberu [87]. Shibberu's approach is a particular case of the first formulation of the generalized discrete modified Hamilton's principle (Def. (IV.13)) for the midpoint rule but he used a different discrete variational principle from DVPII.

One other work on symplectic energy preserving algorithms is that of Kane, Marsden and Ortiz [56]. They developed a generalized discrete modified Hamilton's principle that is based on DVPI. Their approach is different from ours: they assume a different time step at each iteration, and then take the variation of the discrete action without varying the time step (i.e., in an n-dimensional space). As a consequence they only obtain n equations for the $n + 1$ variables (q_k, h_k) where h_k is the time step at the k^{th} step. They then add an energy constraint to obtain $n+1$ equations. Their definition of the energy is similar to ours and therefore both methods provide the same algorithm. However, there are fundamental differences between the two methods. First, our method is fully variational. Second, all the differences between DVPI and DVPII that we emphasize at the beginning of this chapter still remain because their work is based on DVPI whereas our is based on DVPII.

4.5 Discrete Hamilton-Jacobi theory

So far we have developed two variational principles that are the discrete counterparts of Hamilton's principle on the tangent bundle and on the cotangent bundle. Through several examples we have observed that both variational principles are equivalent and that they allow us to recover classical variational symplectic integrators. We have also shown that they can be modified so that energy conservation is assured. In this section, we concentrate on discrete Hamilton-Jacobi theory. We define discrete canonical transformations (DCT), discrete generating functions (DGF) and derive a discrete Hamilton-Jacobi equation that allows us to show that the energy error for a certain class of scheme is invariant under discrete canonical transformations.

4.5.1 Discrete symplectic geometry

We consider again a discretization of the time t into n instants $\mathcal{T} = \{(t_k)_{k\in[1,n]}\}$ but we restrict to the case where M_k is a n-dimensional vector space. We still define $\mathcal{M} = \bigcup M_k$.

Definition IV.21. *A discrete symplectic form ω on $\mathcal{M}$ is one such that at t_k, $\omega = \omega_k^d$, where ω_k^d is a non-degenerate, closed, two-form on $M_k^d = M_k \cup M_{k+1}$.*
A discrete canonical one-form θ on $\mathcal{M}$ is such that at t_k, $\theta = \theta_k^d$, and $\omega_k^d = -d\theta_k^d$.
A discrete symplectic vector space $(\mathcal{M}, \omega)$ is a vector space $\mathcal{M} = \bigcup M_k$ together with a discrete symplectic two form on $\mathcal{M}$.

Using a symplectic chart, a discrete symplectic form on $\mathcal{M}$ at t_k can be written as:

$$\omega_k^d = dq_k^d \wedge dp_k^d\,,$$

and the canonical one-form as $\theta_k^d = p_k^d dq_k^d$.

In the remainder of this section we consider the geometry associated with the midpoint scheme, that is, we define $z_k^d = (q_k^d, p_k^d)$ as $z_k^d = \frac{z_k+z_{k+1}}{2}$ and use the modified Leibnitz law defined by Eq. (4.29). However, the content of this section can be applied to any scheme as long as one can define a discrete Hamiltonian vector field from the discrete Hamiltonian function and the discrete symplectic two-form (see Def. 4.55). In particular, it is clear that the theory herein can be adapted to systems for which the action integral involves a term of the form $H_d(z_k^d)$, where z_k^d is a linear combination of z_k and z_{k+1}, but it is not clear if it can be adapted to the Störmer rule for instance ($z_k^d = (q_k, p_{k+1})$ cannot be written as a linear combination of z_{k+1} and z_k so the next definition does not apply). We do not know how to modify this approach so that a discrete Hamiltonian vector field can be defined from the Hamiltonian function $H_d(q_k, p_{k+1})$ corresponding to the Störmer scheme).

Definition IV.22. *Let $(\mathcal{M},\omega)$ be a discrete symplectic vector space, and $H_d : \mathcal{M} \to \mathbb{R}$ a smooth function. Define the discrete vector field X_H^d such that at t_k, $X_H^d = X_k^d$, where X_k^d is of the form*

$$X_H^d = \Delta_\tau^d q_k^d \frac{\partial}{\partial q_k^d} + \Delta_\tau^d p_k^d \frac{\partial}{\partial p_k^d}\,,$$

and verifies:

$$i_{X_k^d}\omega_k^d = dH_d\,. \tag{4.55}$$

The discrete vector field X_H^d is called the discrete Hamiltonian vector field.

$(\mathcal{M},\omega,X_H^d)$ is called a discrete Hamiltonian system.

Proposition IV.23. *Using the canonical coordinates, a Hamiltonian vector field is of the form:*

$$X_H^d = J \cdot dH_d\,, \ \ \textit{where} \ \ J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{4.56}$$

Proof. Eq. (4.55) is expressed in local coordinates as:

$$i_{X_H^d}(dq_k^d \wedge dp_k^d) = D_1 H_d(q_k^d, p_k^d) dq_k^d + D_2 H_d(q_k^d, p_k^d) dp_k^d\,. \tag{4.57}$$

Let X_H^d be:

$$X_H^d = \Delta_\tau^d q_k^d \frac{\partial}{\partial q_k^d} + \Delta_\tau^d p_k^d \frac{\partial}{\partial p_k^d}\,,$$

then

$$\begin{aligned} i_{X_H^d}(dq_k^d \wedge dp_k^d) &= (i_{X_H^d} dq_k^d) dp_k^d - dq_k^d \wedge (i_{X_H^d} dp_k^d) \\ &= \Delta_\tau^d q_k^d dp_k^d - \Delta_\tau^d p_k^d dq_k^d\,. \end{aligned}$$

Identifying this last equation with Eq. (4.57) leads to Eq. (4.56). □

4.5.2 Discrete canonical transformation

We now define the class of discrete canonical transformations. The definition given here is restricted to linear with respect to the phase space variables, time-dependent maps. We believe a larger class of transformations may be considered if one works with discretization of the spacetime [65]. Let $(\mathcal{M}_1, \omega_1)$ and $(\mathcal{M}_2, \omega_2)$ be discrete symplectic vector spaces and $\mathcal{F}$ be the set of maps $f : \mathcal{T} \times \mathcal{M}_1 \to \mathcal{T} \times \mathcal{M}_2$ that are linear with respect to the phase space variables. Consider a map $f \in \mathcal{F}$ such that $\forall t_k \in \mathcal{T}$, $f(t_k, \cdot) = f_k(\cdot)$ where f_k is the following linear map:

$$\begin{aligned} f_k : \ M^d_{1,k} &\to M^d_{2,k} \\ z_k = (q_k, p_k) &\mapsto Z_k = (Q_k, P_k) = A_k z_k + B_k \,. \end{aligned}$$

Since f_k is linear, we have:

$$f_k(z^d_k) = \frac{1}{2}(f_k(z_k) + f_k(z_{k+1}))\,, \tag{4.58}$$

$$f_k(\Delta^d_\tau z^d_k) = A_k \Delta^d_\tau z^d_k \,. \tag{4.59}$$

Definition IV.24. *A linear, time-dependent map f is called a discrete canonical transformation (DCT) (or a discrete symplectic map) if and only if $f^*\omega_2 = \omega_1$, or equivalently, $\forall k \in [1, n]$, $f^*_k \omega^d_{2,k} = \omega^d_{1,k}$.*

Proposition IV.25. *If f is a DCT then A_k is invertible for all $k \in [1, n]$.*

Proof. Suppose there exists a k such that A_k is not invertible. Then

$$\exists z^d_k \in M^d_{1,k} \; \exists v_1 \in T_{z^d_k} M^d_{1,k} \mid Tf_k \cdot v_1 = 0 \,.$$

Since f is symplectic, $\forall v_2 \in T_{z^d_k} M^d_{1,k} \mid v_2 \neq 0$, $\omega^d_{1,k}(v_1, v_2) = \omega^d_{2,k}(Tf_k \cdot v_1, Tf_k \cdot v_2)$. The right hand side is zero but the left hand side is not. This is a contradiction and therefore A_k is invertible. □

Lemma IV.26. *Let f be a discrete canonical transformation. Then ${f_k}^*\omega^d_{2,k} = \omega^d_{1,k}$ can be written in the matrix form $A_k J A_k^T = J$. In addition, f preserves the form of the discrete Hamilton's equations.*

Proof. The statement $A_k J A_k^T = J$ is just the matrix statement of ${f_k}^*\omega^d_{2,k} = \omega^d_{1,k}$. Let us prove that f preserves the form of the discrete Hamilton's equations. Define the function K_d such that $f^* K_d = H_d$.

On one hand, using Eq. (4.59) we have:

$$\begin{aligned}\Delta^d_\tau Z^d_k &= \frac{f_k(z_{k+1}) - f_k(z_k)}{\tau} \\ &= A_k \Delta^d_\tau z^d_k \,.\end{aligned}$$

On the other hand:

$$\begin{aligned}J\nabla H_d(z^d_k) &= J\nabla(K_d \circ f_k(z^d_k)) \\ &= J A_k^T \nabla K_d(z^d_k)\,.\end{aligned}$$

Since $A_k J A_k^T = J$, we obtain $\Delta^d_\tau Z^d_k = J\nabla K_d(z^d_k)$ □

This last result can be summarized as follows:

Proposition IV.27. *Let X^d_H be a discrete Hamiltonian vector field with Hamiltonian function H_d and f a discrete symplectic map. Then $f_* X^d_H$ is a discrete Hamiltonian vector field with Hamiltonian function $f_* H_d$.*

4.5.3 Discrete generating functions

Proposition IV.28. *Let $(\mathcal{M}_1, \omega_1)$ and $(\mathcal{M}_2, \omega_2)$ be two discrete symplectic vector spaces, $\pi_i : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{M}_i$ the projection onto $\mathcal{M}_i$ and define*

$$\Omega = \pi_1^*\omega_1 - \pi_2^*\omega_2\,.$$

Then,

1. *Ω is a discrete symplectic form on $\mathcal{M}_1 \times \mathcal{M}_2$,*

2. *a map $f : \mathcal{M}_1 \to \mathcal{M}_2$ is a discrete symplectic map if and only if $i_f^*\Omega = 0$, where $i_f : \Gamma_f \to \mathcal{M}_1 \times \mathcal{M}_2$ is the inclusion map and Γ_f is the graph of f.*

Proof. We recall that at t_k, $\Omega = \Omega_k^d$ where $\Omega_k^d = \pi_1{}^*\omega_{1,k}^d - \pi_2{}^*\omega_{2,k}^d$. To prove that Ω is a discrete symplectic form, we need to prove that Ω_k^d is a symplectic form on $M_{1,k}^d \times M_{2,k}^d$ for all $k \in [1, n]$.

$$\begin{aligned} d\Omega_k^d &= d(\pi_1^*\omega_{1,k}^d - \pi_2^*\omega_{2,k}^d) \\ &= \pi_1^* d\omega_{1,k}^d - \pi_2^* d\omega_{2,k}^d \\ &= 0\,, \end{aligned}$$

since $\omega_{i,k}^d$ is closed and d commutes with the pull back operator.

Now let $z_k^d = (z_{1,k}^d, z_{2,k}^d) \in M_{1,k}^d \times M_{2,k}^d$ and $v = (v_1, v_2) \in T_{z_k^d}(M_{1,k}^d \times M_{2,k}^d) \sim T_{z_{1,k}^d} M_{1,k}^d \times T_{z_{2,k}^d} M_{2,k}^d$ such that

$$\forall w = (w_1, w_2) \in T_{z_k^d}(M_{1,k}^d \times M_{2,k}^d)\,,\ \Omega_k^d(v, w) = 0$$

and let us prove that v is zero. We have

$$\begin{aligned} \Omega_k^d(v, w) &= \omega_{1,k}^d(\pi_1(z_k^d))(T\pi_1 \cdot v, T\pi_1 \cdot w) - \omega_{2,k}^d(\pi_2(z_k^d))(T\pi_2 \cdot v, T\pi_2 \cdot w) \\ &= \omega_{1,k}^d(z_{1,k}^d)(v_1, w_1) - \omega_{2,k}^d(z_{2,k}^d)(v_2, w_2)\,. \end{aligned} \tag{4.60}$$

The right hand side of Eq. (4.60) is zero for all w if and only if both terms are zero, that is,

$$\omega_{1,k}^d(z_{1,k}^d)(v_1, w_1) = 0\,,\ \omega_{2,k}^d(z_{2,k}^d)(v_2, w_2) = 0\,.$$

Since $\omega_{i,k}^d$ is non-degenerate, $v_1 = v_2 = 0$. Thus, Ω_k^d is closed and non-degenerate for all k, i.e, Ω a discrete symplectic two-form.

We now prove the second statement of the proposition. We first notice that f_k induces a diffeomorphism of $M^d_{1,k}$ to Γ_{f_k}, so we can write

$$T_{(z^d_k, f_k(z^d_k))} = \left\{(v, Tf_k \cdot v) | v \in T_{z^d_k} M^d_{1,k}\right\} .$$

Then,

$$\begin{aligned} i^* \Omega^d_k((v_1, Tf_k \cdot v_1), (v_2, Tf_k \cdot v_2)) &= \omega^d_{1,k}(v_1, v_2) - \omega^d_{1,k}(Tf_k \cdot v_1, Tf_k \cdot v_2) \\ &= (\omega^d_{1,k} - {f_k}^* \omega^d_{2,k})(v_1, v_2) . \end{aligned}$$

Hence, f_k is symplectic if and only if $i^*\Omega^d_k = 0$, i.e., f is a discrete symplectic map if and only if $i^*\Omega = 0$. □

Using the Poincaré lemma, we may write $\Omega^d_k = -d\Theta^d_k$ and the previous proposition says that $i^*_{f_k}\Theta^d_k$ is closed if and only if f is a discrete symplectic map. Using again the Poincaré lemma, we conclude that if f is a discrete symplectic map then there exists a function $S : \Gamma_f \longrightarrow \mathbb{R}$ such that $i^*_f \Theta = dS$, i.e., $\forall k \in [1, n]$, $i^*_{f_k}\Theta^d_k = dS_k$

Definition IV.29. *Such a function S is called a discrete generating function for the discrete symplectic map f. S is locally defined and depends on the choice of Θ.*

- Let $\theta^d_{1,k} = p^d_k dq^d_k$ and $\theta^d_{2,k} = P^d_k dQ^d_k$, then

$$\begin{aligned} i^*_{f_k}\Theta^d_k &= (\pi_1 \circ i_{f_k})^* p^d_k dq^d_k - (\pi_2 \circ i_{f_k})^* P^d_k dQ^d_k , \\ dS &= \frac{\partial S}{\partial q}(q^d_k, Q^d_k) dq^d_k + \frac{\partial S}{\partial Q}(q^d_k, Q^d_k) dQ^d_k , \end{aligned}$$

  that is,

$$p^d_k = \frac{\partial S}{\partial q}(q^d_k, Q^d_k) , \; P^d_k = -\frac{\partial S}{\partial Q}(q^d_k, Q^d_k) .$$

  S as defined corresponds to a discrete generating function of the first kind.

- Let $\theta^d_{1,k} = p^d_k dq^d_k$ and $\theta^d_{2,k} = -Q^d_k dP^d_k$, then

  $$\begin{aligned} i^*_{f_k}\Theta^d_k &= (\pi_1 \circ i_{f_k})^* p^d_k dq^d_k + (\pi_2 \circ i_{f_k})^* Q^d_k dP^d_k \,, \\ dS &= \frac{\partial S}{\partial q}(q^d_k, P^d_k) dq^d_k + \frac{\partial S}{\partial Q}(q^d_k, P^d_k) dP^d_k \,, \end{aligned}$$

  that is,

  $$p^d_k = \frac{\partial S}{\partial q}(q^d_k, P^d_k)\,, \; Q^d_k = \frac{\partial S}{\partial P}(q^d_k, P^d_k)\,.$$

  S as defined corresponds to a discrete generating function of the second kind.

In the same way, one can define 4^n generating functions as in the continuous case. Note that since f is linear with respect to its spatial variables (Def. IV.24), S is also linear with respect to its spatial variables. At t_k, $S = S_k$, where $S_k(\cdot) = T_k(\cdot) + U_k$ is an affine map, T_k is a $2n \times 2n$ matrix and U_k is a $2n \times 1$ matrix.

4.5.4 Discrete Hamilton-Jacobi theory

In this section we use the notions introduced previously to develop a discrete Hamilton-Jacobi theory. Let f be a discrete symplectic map, $M^d_{i,k} = T^*\mathcal{Q}^d_{i,k}$ and S be an associated discrete generating function such that $S = S^d_k$ at t_k, where $S_k(\cdot) = T_k(\cdot) + U_k$

Theorem IV.30. *Define*

$$\tilde{p}^d_k(q^d_k, Q^d_k) = D_1 S_k(q^d_k, Q^d_k)\,, \quad \tilde{P}^d_k(q^d_k, Q^d_k) = -D_2 S_k(q^d_k, Q^d_k)\,.$$

Then the following two conditions are equivalent:

1. *S is a discrete generating function associated with f;*

2. - *For every curve $(c_k)_k$ in $\mathcal{Q}_1 = \bigcup \mathcal{Q}_{1,k}$ satisfying:*

     $$\Delta^d_\tau c^d_k = T\pi^*_{\mathcal{Q}^d_{1,k}} X^d_H(c^d_k, \tilde{p}^d_k)\,,$$

the curve $k \mapsto (c_k^d, \tilde{p}_k^d)$ is a discrete integral curve of X_H^d, where $\pi^*_{\mathcal{Q}^d_{1,k}}$ is the cotangent bundle projection onto the configuration space.

- *For every curve $(c_k)_k$ in $\mathcal{Q}_2 = \bigcup \mathcal{Q}_{2,k}$ satisfying:*

$$\Delta_\tau^d c_k^d = T\pi^*_{\mathcal{Q}^d_{2,k}} X_K^d(c_k^d, \tilde{P}_k^d)\,,$$

the curve $k \mapsto (c_k^d, \tilde{P}_k^d)$ is a discrete integral curve of X_K^d, where $\pi^*_{\mathcal{Q}^d_{2,k}}$ is the cotangent bundle projection onto the configuration space.

Proof. Suppose S is a discrete generating function, let Q_k^d be fixed and consider a curve $(c_k)_k$ verifying

$$\Delta_\tau^d c_k^d = T\pi^*_{\mathcal{Q}^d_{1,k}} X_H^d(c_k^d, \tilde{p}_k^d)\,,$$

In other words, c_k verifies:

$$\Delta_\tau^d c_k^d = D_2 H(c_k^d, \tilde{p}_k^d)\,,$$

Since S is a generating function, $\tilde{p}_k^d$ is the momentum associated with c_k^d and verifies:

$$\Delta_\tau^d \tilde{p}_k^d = -D_1 H(c_k^d, \tilde{p}_k^d)\,.$$

These last two equations are exactly a restatement of: $k \mapsto (c_k^d, \tilde{p}_k^d)$ is a discrete integral curve of X_H^d. To derive the second item we proceed in the same manner, but this time q_k^d is fixed.

Now we suppose 2. and we show that S is a discrete generating function for f. The statements $k \mapsto (c_k^d, \tilde{p}_k^d)$ is a discrete integral curve of X_H^d and $k \mapsto (c_k^d, \tilde{P}_k^d)$ is a discrete integral curve of X_K^d are equivalent to saying that $\tilde{p}_k^d$ and $\tilde{P}_k^d$ are the momenta associated with the generalized coordinates, and therefore, S is a generating function for f. □

Theorem IV.31. *We consider again a time-dependent function S which is linear with respect to the spatial variables. Then the following two statements are equivalent:*

1. *S is a discrete generating function associated with f;*

2. *For every H there is a function K such that*

$$H(q^d_k, D_1S(q^d_k, Q^d_k)) = K(Q^d_k, D_2S(q^d_k, Q^d_k))\,.$$

Proof. Suppose S is a discrete generating function. Then from the previous theorem, for every curve (c_k, C_k) in $\mathcal{Q}_1 \times \mathcal{Q}_2$ satisfying $\Delta^d_\tau c^d_k = T\pi^*_{\mathcal{Q}^d_{1,k}} X^d_H(c^d_k, \tilde{p}^d_k)$ and $\Delta^d_\tau C^d_k = T\pi^*_{\mathcal{Q}^d_{2,k}} X^d_K(C^d_k, \tilde{P}^d_k)$, the curves $k \mapsto (c^d_k, \tilde{p}^d_k)$ and $k \mapsto (C^d_k, \tilde{P}^d_k)$ are discrete integral curves of X^d_H and X^d_K respectively. Then, using the symplectic identity (see e.g. Abraham and Marsden [1] page 382) that holds for any function S:

$$\omega^d_{1,k}(T(D_1S \circ \pi^*_{\mathcal{Q}^d_{1,k}}) \cdot v, w) = \omega^d_{1,k}(v, w - T(D_1S \circ \pi^*_{\mathcal{Q}^d_{1,k}}) \cdot w)\,,$$

we obtain:

$$\begin{aligned}&\omega^d_{1,k}(T(D_1S \circ \pi^*_{\mathcal{Q}^d_{1,k}}) \cdot X^d_H(c_k, D_1S_k), w) = \\ &\qquad\qquad \omega^d_{1,k}(X^d_H(c_k, D_1S_k), w) - dH_d(c_k, D_1S_k) \cdot TD_1S(c_k, D_1S_k) \cdot w\,,\end{aligned} \tag{4.61}$$

$$\begin{aligned}&\omega^d_{2,k}(T(-D_2S \circ \pi^*_{\mathcal{Q}^d_{2,k}}) \cdot X^d_K(C_k, -D_2S_k), w) = \\ &\qquad \omega^d_{2,k}(X^d_K(C_k, -D_2S_k), w) - dK_d(C_k, -D_2S_k) \cdot T - D_2S(C_k, -D_2S_k) \cdot w\,.\end{aligned} \tag{4.62}$$

In addition, since $p^d_k = D_1S(c^d_k, C^d_k)$ and $P^d_k = -D_1S(c^d_k, C^d_k)$,

$$\begin{aligned}\Delta^d_\tau p^d_k &= TD_1S(c^d_k, C^d_k)\Delta^d_\tau c^d_k = T(D_1S \circ \pi^*_{\mathcal{Q}^d_{1,k}}) \cdot X^d_H(c_k, D_1S_k)\,, \\ \Delta^d_\tau P^d_k &= T(-D_2S \circ \pi^*_{\mathcal{Q}^d_{2,k}}) \cdot X^d_K(C_k, -D_2S_k)\,.\end{aligned}$$

f being a discrete canonical map, $Tf_k(\Delta^d_\tau p^d_k) = \Delta^d_\tau P^d_k$ so the left hand side of Eq. (4.62) is the image under f of the left hand side of Eq. (4.62). Using Prop. (IV.27), we conclude that:

$$Tf_k \cdot dH_d(c_k, D_1S_k) \cdot TD_1S(c_k, D_1S_k) = -dK_d(C_k, -D_2S_k) \cdot TD_2S(C_k, -D_2S_k)\,,$$

which is equivalent to the discrete Hamilton-Jacobi equation.

The proof that $2.$ implies $1.$ follows from these arguments. □

4.5.5 Applications of the discrete Hamilton-Jacobi theory

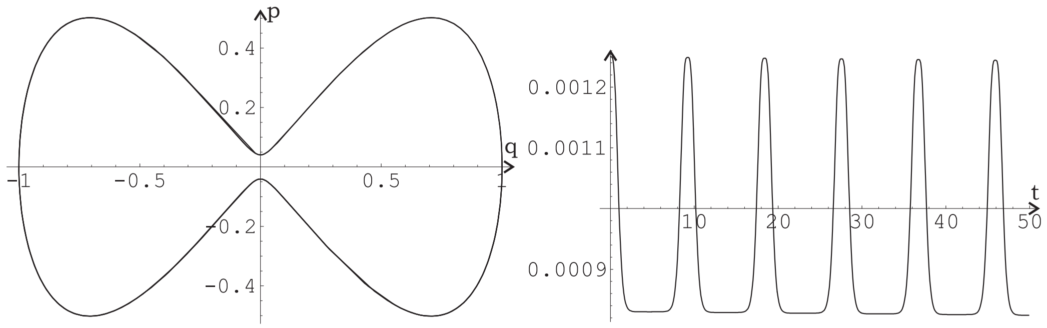
The goal of this section is to highlight the benefit of having a discrete Hamilton-Jacobi theory. First, we have proved the invariance of the discrete Hamilton's equations under a certain class of coordinate transformations. Second, we have shown in Thm. IV.31 that changing coordinates using a discrete symplectic map does not improve the performance of the algorithm in terms of energy conservation. As a consequence we have the following lemma:

Lemma IV.32. *The midpoint scheme preserves the energy for linear systems.*

Proof. The discrete phase flow for linear systems is piecewise linear continuous and the map $(q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$ is symplectic (the midpoint scheme is a symplectic algorithm). Therefore, the discrete phase flow is a discrete symplectic map that maps H into a constant K that can be chosen to be 0 (the discrete flow maps (q_k, p_k) into (q_0, p_0)). Integration of the Hamiltonian system defined by K is trivial since the system is in equilibrium and it obviously preserves the energy. As a consequence, the integration of the Hamiltonian system defined by H also preserves the energy (Thm. IV.31). □

Finally, we illustrate the use of the above material with a nonlinear example. We study the energy error in the integration of the equations of motion of a particle in a double well potential using different sets of canonical coordinates.

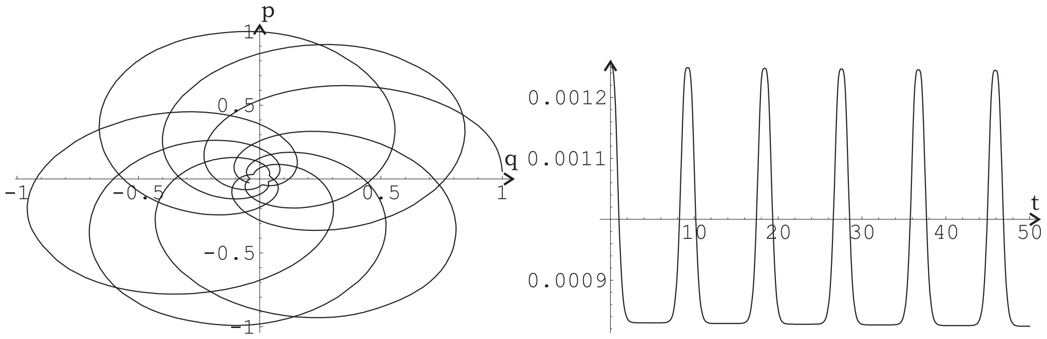
Example IV.33. Consider a particle in a double well potential, i.e., $H = \frac{1}{2}p^2 + \frac{1}{2}(q^4 - q^2)$. As shown in Fig. (4.2), the midpoint scheme does not preserve the energy. The following time-dependent discrete canonical transformation (at each step the transformation is a different expression) $Z_k = A_k z_k + B_k$ where $A_k = \begin{pmatrix} \cos(k\theta) & -\sin(k\theta) \\ \sin(k\theta) & \cos(k\theta) \end{pmatrix}$ and $B_k = 0$, rotates the system by $k\theta = k \arccos 0.99$ at the k^{th} step. In Fig. (4.3) we plot the same trajectory in the new system of coordinates. As predicted by the discrete Hamilton-Jacobi theory, the energy error is exactly the same. In other words, the energy error is invariant under discrete canonical maps.



(a) Trajectory in the $q - p$ plane

(b) Energy error for constant time step midpoint scheme as a function of time.

Figure 4.2: Particle in a double well potential with initial conditions $(q, p) = (1, 0.05)$



(a) Trajectory in the $q - p$ plane

(b) Energy error for constant time step midpoint scheme as a function of time.

Figure 4.3: Particle in the vector field $f_* X_H^d$ with initial conditions $f_0(1, 0.05)$.

CHAPTER V

COMPUTING THE GENERATING FUNCTIONS

The Hamilton-Jacobi equation (Eq. (2.39)) was first encountered by Hamilton [46] in geometric optics as a partial differential equation that the characteristic function had to satisfy. A year later, he introduced Hamilton's principal function [47] for studying dynamical systems and found that this also satisfies the Hamilton-Jacobi equation. Since then, this equation has been regularly encountered in many different fields.

- In quantum mechanics the phase of the wave function verifies the Schrödinger equation which is a Hamilton-Jacobi equation for Hamiltonian systems of the form $H = T + V$.

- In optimal control the Hamilton-Jacobi equation arises from the sufficiency conditions for optimality and is called the Hamilton-Jacobi-Bellman equation.

- In the present research, the generating functions for the phase flow transformation verify the Hamilton-Jacobi equation.

Early on, researchers proved the existence of solutions to the Hamilton-Jacobi equation (Lions [61] and Aubin [8]). Meanwhile, analytical methods were developed to solve this partial differential equation. Many of them can be found in textbooks (see e.g. Greenwood [28], Goldstein [27] and Arnold [4]). Since then, during the last two decades, numerical techniques have been explored either based on geometric (multi-symplectic) integrators or on properties of a particular system [64, 70, 17]. However, none of these methods and algorithms allows us to solve the Hamilton-Jacobi equation for the generating functions because of three main difficulties: 1) The boundary conditions for integration are specified in terms of functions with parameters. 2) Generating functions may develop singularities that prevent the integration from going forward (some algorithms have been developed to compute multiple solutions, see e.g. Benamou [11] and references therein). 3) We want to apply our theory to non-trivial systems and so analytical methods fail due to the complexity of the system. The purpose of this chapter is to develop a robust algorithm that addresses these challenges. Specifically, the algorithm we present approximates solutions to the Hamilton-Jacobi equation locally in space and globally in time. It allows one to use a variety of boundary conditions and can avoid singularities in the functions during the integration. Most important, our algorithm is independent of the complexity of the dynamical system.

5.1 Initial conditions for the generating functions

To compute the generating functions, one needs boundary conditions to solve the Hamilton-Jacobi partial differential equation. At the initial time, the flow induces the identity transformation, and thus the generating functions should also do so. In other words, at the initial time,

$$q(t_0) = q_0\,,\ p(t_0) = p_0\,. \tag{5.1}$$

that is, $(q(t_0), p_0)$ and $(p(t_0), q_0)$ are the only sets of independent variables that contain n initial conditions and n components of the state vector at the initial time. As a consequence, all the generating functions save F_2 and F_3 are singular at the initial time (we already saw this result for linear generating functions in Section 3.2.2).

Example V.1. Let us look, for example, at the generating function of the first kind, $F_1(q, q_0, t)$. At the initial time, q is equal to q_0 whatever values the associated momenta p and p_0 take. Therefore, there are multiple solutions to the boundary value problem that consists of going from q_0 to $q = q_0$ in 0 units of time. From Prop. III.5, we conclude that F_1 is singular.

We now focus on the boundary conditions for the F_2 and F_3 generating functions. At the initial time we must have:

$$\left\{\begin{array}{rcl} p_0 & = & \frac{\partial F_2}{\partial q}(q = q_0, p_0, t_0)\,, \\ q_0 & = & \frac{\partial F_2}{\partial p_0}(q = q_0, p_0, t_0)\,, \end{array}\right. \qquad \left\{\begin{array}{rcl} q_0 & = & -\frac{\partial F_3}{\partial p}(p = p_0, q_0, t_0)\,, \\ p_0 & = & -\frac{\partial F_3}{\partial q_0}(p = p_0, q_0, t_0)\,. \end{array}\right.$$

Due to the non-commutativity of the derivative operator and the operator that assigns the value t_0 at t, solutions to these equations are not unique. As a result, the boundary conditions verified by F_2 and F_3 are not uniquely defined as well. For instance, they may be chosen to be:

$$F_2(q, p_0, t) = \langle q, p_0 \rangle\,,\ F_3(p, q_0, t) = -\langle p, q_0 \rangle\,, \tag{5.2}$$

or

$$F_2(q, p_0, t) = \frac{1}{t - t_0} e^{(t-t_0)\langle q, p_0 \rangle}\,,\ F_3(p, q_0, t) = -\frac{1}{t - t_0} e^{(t-t_0)\langle p, q_0 \rangle}\,, \tag{5.3}$$

where $\langle , \rangle$ is the inner product. One can readily verify that Eqns. (5.2) and (5.3) generate the identity transformation (5.1) at the initial time $t = t_0$.

The singularity at the initial time of all but two generating functions is a major issue: it prevents us from initializing the integration, i.e., from solving the Hamilton-Jacobi equation. In Section 5.3.1 we present a technique to circumvent this problem, namely we are able to specify boundary value conditions for all generating functions at a later time.

5.2 The use of partial differential equation solvers

In the previous section, we saw that, *a priori*, only F_2 and F_3 may be found because the other generating functions have singular boundary conditions at the initial time. In this section, we use standard partial differential equation solvers to compute the F_2 generating function. In particular, we show that they impose drastic restrictions on the boundary value problem solved by F_2.

The Hamilton-Jacobi equation verified by F_2 reads:

$$\frac{\partial F_2}{\partial t}(q, p_0, t) + H(q, \frac{\partial F_2}{\partial q}(q, p_0, t), t) = 0\,,\ F_2(q, p_0, 0) = \langle q, p_0 \rangle\,.$$

In this partial differential equation p_0 does not appear explicitly. Therefore it may be viewed as a parameter, in which case the Hamilton-Jacobi equation simplifies to:

$$\frac{\partial F_2}{\partial t}(q, t) + H(q, \frac{\partial F_2}{\partial q}(q, t), t) = 0\,,\ F_2(q, 0) = \langle q, p_0 \rangle\,, \tag{5.4}$$

where p_0 is a parameter that specifies the initial conditions. Classical numerical partial differential equation solvers do not accept symbolic boundary conditions and so we need to specify p_0. Once p_0 is set to a value α, we can solve the Hamilton-Jacobi equation on the interval $[q_{min}, q_{max}] \times [t_0, t_f]$ as long as no singularities are encountered. The resulting function corresponds to the generating function $F_2(q, p_0 = \alpha, t)$. Since $p_0 = \alpha$, F_2 only solves two-point boundary value problems that consists of going to q in t units of time with a given initial momentum $p_0 = \alpha$. We loose the freedom to choose p_0.

Example V.2 (Weakly perturbed pendulum). To illustrate the use of classical partial differential equation solvers, let us compute F_2 for a weakly perturbed pendulum. The Hamiltonian for this system is given by:

$$H(q, p) = \frac{1}{2}p^2 + \frac{0.01}{2}q^2 - \cos(q)\,,$$

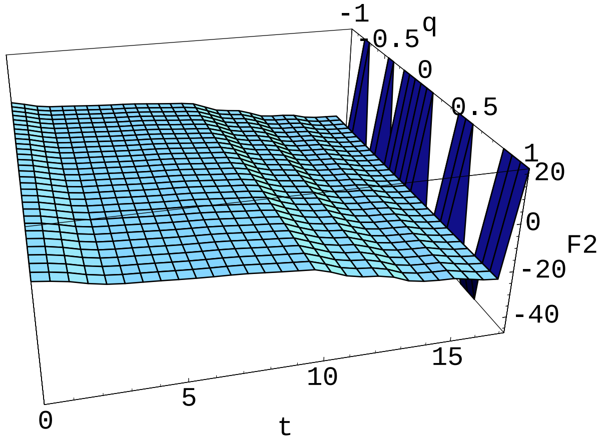


Figure 5.1: F_2 computed using the method of lines

and the Hamilton-Jacobi equation reads:

$$\frac{\partial F_2}{\partial t}(q,t) + \frac{1}{2}\left(\frac{\partial F_2}{\partial q}(q,t)\right) + \frac{0.01}{2}q^2 - \cos(q) = 0\,, \quad F_2(q,0) = qp_0\,,$$

where p_0 is chosen to be 2.1. Using the built-in $Mathematica^{©}$ function $NDSolve$[1] we solve the Hamilton-Jacobi for F_2 over the interval $(q,t) \in [-1,1] \times [0, 17.215]$. In Fig. V.2 we plot this solution. In order to solve a boundary value problem that consists of going to $q \in [-1,1]$ in $t \in [0, 17.215]$ units of time with initial momentum $p_0 = 2.1$, we approximate $\frac{\partial F_2}{\partial q}(q,t)$ at the point (q,t). We point out that at $t = 17.215$, F_2 becomes singular and the integration stops. Therefore, we cannot solve any problems involving transfer times that are larger than 17.215.

Through the weakly perturbed pendulum, we illustrated the restriction imposed by partial differential equation solvers on boundary value problems. We showed that the initial state must be partially known *a priori*. Moreover, the integration of the Hamilton-Jacobi equation stops as soon as a singularity is encountered. Most importantly, only two types of boundary value problems can be solved because all but two generating functions

[1] $NDSolve$ uses the method of lines. It consists of discretizing all but one variable so that at every node, the partial differential equation reduces to an ordinary differential equation.

are singular at the initial time. These issues are important ones and must be overcome in order to take full advantage of the theory we introduced in Chapter III. In the remainder of this chapter, we present a new algorithm that addresses this difficulty. Specifically, our algorithm can approximate locally in the spatial domain any kind of generating functions over an arbitrary large time interval while avoiding singularities.

5.3 A new algorithm to compute the generating functions

In this section we introduce an algorithm that computes an approximation to the generating functions locally in the spatial domain and globally in the time domain. By locally in the spatial domain, we mean that we are able to compute the generating functions in a domain in which the Hamiltonian function may be expressed as a convergent Taylor series in the q's and p's.

5.3.1 Local solution of the Hamilton-Jacobi equation

We consider the general case of Hamiltonian systems whose Hamiltonian function H can be written as a power series in its spatial variables with time-dependent coefficients. This case obviously includes systems with polynomial Hamiltonians such as the harmonic oscillator, and the double well potential. It also includes systems describing the relative motion of two particles moving in a Hamiltonian vector field (see Appendix A for an expression of the Hamiltonian) and more generally, the motion of a particle in the vicinity of an equilibrium or of a known trajectory. Recall the Hamilton-Jacobi equation (Eq. (3.6)):

$$H(q_{I_p}, -\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}}, \frac{\partial F_{I_p,K_r}}{\partial q_{I_p}}, p_{\bar{I}_p}, t) + \frac{\partial F_{I_p,K_r}}{\partial t} = 0\,. \tag{5.5}$$

Since H is a Taylor series in its spatial variables, we look for a solution of the same form, that is, we assume that generating functions are Taylor series as well:

$$F_{I_p,K_r}(y,t) = \sum_{q=2}^{\infty} \sum_{\substack{i_1,\cdots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}!} f^{p,r}_{q,i_1,\cdots,i_{2n}}(t) y_1^{i_1}\cdots y_{2n}^{i_{2n}}\,, \tag{5.6}$$

where $y = (q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$. We substitute this expression into Eq. (5.5). The resulting equation is an ordinary differential equation that has the following structure:

$$P(y, f^{p,r}_{q,i_1,\cdots,i_{2n}}(t), \dot{f}^{p,r}_{q,i_1,\cdots,i_{2n}}(t)) = 0\,, \tag{5.7}$$

where P is a series in y with time-dependent coefficients. An explicit expression of P up to order 3 is given in Appendix B. Eq. (5.7) holds for all y if and only if all the coefficients of P are zero. In this manner, we transform the ordinary differential equation (Eq. (5.7)) into a set of ordinary differential equations whose solutions are the coefficients of the generating function F_{I_p,K_r}.

Now it remains to specify initial conditions for the integration. We have seen before that only F_2 and F_3 can generate the identity transformation, the other generating functions being singular. Let us look more closely at F_2 and F_3, and especially at the coefficients[2] $f^2_{q,i_1,\cdots,i_{2n}}(t_0)$ and $f^3_{q,i_1,\cdots,i_{2n}}(t_0)$. At the initial time we have:

$$\begin{aligned} p_0 &= p \\ &= \frac{\partial F_2}{\partial q}\,, \end{aligned}$$

and

$$\begin{aligned} q &= q_0 \\ &= \frac{\partial F_2}{\partial p_0}\,. \end{aligned}$$

[2] We change our notation for convenience: f^2 stands for $f^{n,0}$, i.e., represents the coefficients of the Taylor series of F_2. We do the same for all four kinds of generating functions F_1, F_2, F_3 and F_4.

Within the radius of convergence, the Taylor series defining the generating functions (Eq. (5.6)) converge normally, therefore, we can invert the summation and the derivative operator. We obtain:

$$f^2_{q,i_1,\cdots,i_{2n}}(t_0) = \begin{cases} 1 & \text{if } q = 2, i_k = i_{k+n} = 1, i_{l \neq \{k,k+n\}} = 0, \forall (k,l) \in [1,n] \times [1,2n], \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we obtain for F_3:

$$f^3_{q,i_1,\cdots,i_{2n}}(t_0) = \begin{cases} -1 & \text{if } q = 2, i_k = i_{k+n} = 1, i_{l \neq \{k,k+n\}} = 0, \forall (k,l) \in [1,n] \times [1,2n], \\ 0 & \text{otherwise.} \end{cases}$$

These initial conditions allow one to integrate two generating functions among the 4^n, but what about the other ones? This issue on singular initial conditions is similar to the one on singularity avoidance during the integration. In the next section we propose a technique to handle these problems based on the Legendre transformation. But before going further, one remark needs to be made. After we proceed with the integration, one must always verify that the series converge and that they describe the true dynamics[3] in some open set. If these two conditions are verified we can identify the generating functions with their Taylor series within the radius of convergence.

Singularity avoidance

We have seen that most of the generating functions are singular at the initial time. Moreover solutions to the Hamilton-Jacobi equations often develop caustics (Chapter III). These two issues prevent numerical integration, and the goal of this section is to introduce a technique to overcome this difficulty.

[3] Remember that even if a function is C^∞ and has a converging Taylor series, it may not equal its Taylor series. As an example take $f(x) = \exp(1/x^2)$ if $x \neq 0$, $f(0) = 0$, it is C^∞ and its Taylor series at $x = 0$ is 0, and therefore converges. However, f is not identically zero.

We first need to recall the Legendre transformation, which allows one to derive one generating function from another (Eq. (2.28)). Suppose F_2 is known, then we can find F_1 from:

$$F_1(q, q_0, t) = F_2(q, p_0, t) - \langle q_0, p_0 \rangle \,, \tag{5.8}$$

where p_0 is viewed as a function of (q, q_0). Obviously, the difficulty in proceeding with a Legendre transformation lies in finding p_0 as a function of (q, q_0). To find such an expression we use Eq. (3.11):

$$q_0 = \frac{\partial F_2}{\partial p_0}(q, p_0, t) \,, \tag{5.9}$$

and then solve for $p_0(q, q_0)$.

For the class of problems we consider, F_2 is a Taylor series. Therefore we need to perform a series inversion to eventually find p_0 as a Taylor series of (q, q_0). Series inversion is a classical problem and we can use the method developed by Moulton [72] (see also Chapter III). We first suppose that there exists a series expansion of p_0 as a function of q and q_0, then insert this expression into Eq. (5.9) and balance terms of the same order. We obtain a set of linear equations, whose solution is found at the cost of $n \times n$ matrix inversion (an example of series inversion can be found in Ex. III.10).

Let us return to the problem of singularity avoidance. So far, we were able to integrate generating functions of the second and third kinds since they have well-defined initial conditions. If we want to find F_1, then we perform a Legendre transformation at $t_1 > 0$ to find the value of F_1 at this instant from the value of F_2. This value can in turn be used to initialize the integration of the Hamilton-Jacobi equation for F_1.

Now suppose F_2 is singular at t_2, let us see how we can take advantage of the Legendre transformation to integrate F_2 for $t > t_2$.

Prop. II.31 tells us that at least one of the generating functions is non-singular at t_2. Without loss of generality, suppose F_1 is non-singular at t_2. At $t_1 < t_2$ we carry out a Legendre transformation to find F_1 from F_2, then we integrate F_1 over $[t_1, t_3 > t_2]$ and carry out another Legendre transformation to recover F_2 at t_3. Once the value of F_2 is found at t_3, the integration of the Hamilton-Jacobi equation can be continued.

We have described an algorithm for solving the Hamilton-Jacobi equation and developed techniques to continue the integration despite singularities. In the next section, we introduce an indirect approach to compute the generating functions based on the initial value problem. This approach naturally avoids singularities but requires more computations (Section 5.3.3).

5.3.2 An indirect approach

By definition, generating functions implicitly define the canonical transformation they are associated with. Hence, we may compute the generating functions from the canonical transformation, that is, compute the generating functions for the phase flow transformation from knowledge of the phase flow. In this section, we develop an algorithm based on these remarks.

Recall Hamilton's equations of motion for relative motion:

$$\begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = J \nabla H^h(q, p, t) \, . \tag{5.10}$$

Suppose that $q(q_0, p_0, t)$ and $p(q_0, p_0, t)$ can be expressed as series in the initial conditions (q_0, p_0) with time dependent coefficients, truncate the series to order N and substitute these into Eq. (5.10). Hamilton's equations reduce to an ordinary differential equation of a form that is polynomial in (q_0, p_0). As before, we balance terms of the same order and transform Hamilton's equations into a set of ordinary differential equations whose variables are the time dependent coefficients defining q and p as series of q_0 and p_0. Using $q(q_0, p_0, t_0) = q_0$ and $p(q_0, p_0, t_0) = p_0$ as initial conditions for the integration, we are able to compute an approximation of order N of the phase flow. Once the flow is known, we recover the generating functions by performing a series inversion.

Example V.3. Suppose F_1 is needed. From $q = q(q_0, p_0, t)$ we carry out a series inversion to eventually find $p_0 = p_0(q, q_0, t)$. Then $p_0 = p_0(q, q_0, t)$ together with $p = p(q_0, p_0, t)$ defines the gradient of F_1:

$$\begin{aligned} \frac{\partial F_1}{\partial q}(q, q_0, t) &= p \\ &= p(q_0, p_0(q, q_0, t))\,, \qquad (5.11) \\ \frac{\partial F_1}{\partial q_0}(q, q_0, t) &= -p_0 \\ &= -p_0(q, q_0, t)\,, \qquad (5.12) \end{aligned}$$

We recover F_1 from its gradient by performing two quadratures over the polynomial terms. We point out that the inversion has multiple solutions if and only if F_1 is singular at t. In addition, if one uses traditional numerical integrators to integrate the phase flow, Eqns. (5.11) and (5.12) are not integrable due to numerical round off $\left(\frac{\partial p(q_0, p_0(q, q_0, t))}{\partial q_0} \neq -\frac{\partial p_0(q, q_0, t)}{\partial q}\right)$. Using symplectic algorithms to compute the approximate phase flow, we preserve the Hamiltonian structure of the flow. Therefore we are assured that Eqns. (5.11) and (5.12) are integrable. This issue was discussed and illustrated in Section 4.3.3.

5.3.3 A comparison of the direct and indirect approach

We have introduced two algorithms that compute the generating functions associated with the phase flow. In this section we highlight the advantages and drawbacks of each method. In addition, we show that by combining them we obtain a robust and powerful algorithm.

Method specifications

The direct approach The direct approach provides us with a closed form approximation of the generating functions over a given time interval. However, there are inherent difficulties as generating functions may develop singularities which prevent the integration from going further in time. The technique we developed to bypass this problem results in additional computations. It requires us to first identify the times at which generating functions become singular, and then to find a non-singular generating function at each of these times. Over a long time simulation, this method reaches its limits as many singularities may need to be avoided.

The indirect approach The main advantage of the indirect method is that it never encounters singularities, as the flow is always non-singular. On the other hand, this method requires us to solve many more equations than the direct approach (see below). Such trade offs between dimensionality and singularities are well known to engineers. For instance, to describe the attitude of a rigid body, one may use Eulerian angles or quaternions. Eulerian angles allow one to describe the attitude with only 3 coordinates, but may become singular. In contrast, the quaternions are never singular but have an additional component. Furthermore, a major drawback of the indirect approach is that it computes an expression for the generating functions at a given time only, the time at which the series inversion is performed. Finally, as mentioned earlier, we need to use symplectic integrators to run the indirect approach. Therefore, we believe (but have not proven yet) that the solution found is symplectic. This is very valuable, especially if we want to find the generating functions over a large time span on which classical integrators fail to preserve the geometric properties of the system.

The curse of dimensionality In this paragraph, we point out a difficulty inherent to both methods, namely the "curse of dimensionality". As we solve the generating functions to higher and higher orders, the number of variables grows dramatically. This problem is the limiting factor for computation: typically on a $2GHz$ Linux computer with $1G$ RAM, we have trouble solving the generating functions to order 7 and up for a 6-dimensional Hamiltonian system.

Computation of the generating functions using the direct approach requires us to find all the coefficients of a $2n$-dimensional series with no linear terms. At order N, a $2n$-dimensional Taylor series has M terms, where

$$M = \begin{pmatrix} 2n-1+N \\ N \end{pmatrix} = \frac{(2n-1+N)!}{N!(2n-1)!}\,.$$

In the indirect approach we express the $2n$-dimensional state vector as Taylor series with respect to the $2n$ initial conditions. Therefore, we need to compute the coefficients of $2n$ $2n$-dimensional Taylor series.

To summarize, an approximation of order N of the generating functions is found by solving:

- $\displaystyle\sum_{n=2}^{N} \frac{(2n-1+N)!}{N!(2n-1)!}$ ordinary differential equations using the direct approach,
- $\displaystyle 2n\sum_{n=1}^{N-1} \frac{(2n-1+N)!}{N!(2n-1)!}$ ordinary differential equations using the indirect approach[4].

In Fig. 5.3.3, the solid line and dotted line indicate the numbers of equations that needs to be solved with the direct and indirect methods for a 6-dimensional Hamiltonian system.

[4] The summation goes from 1 to $N-1$ because the indirect approach computes the gradient of the generating functions.

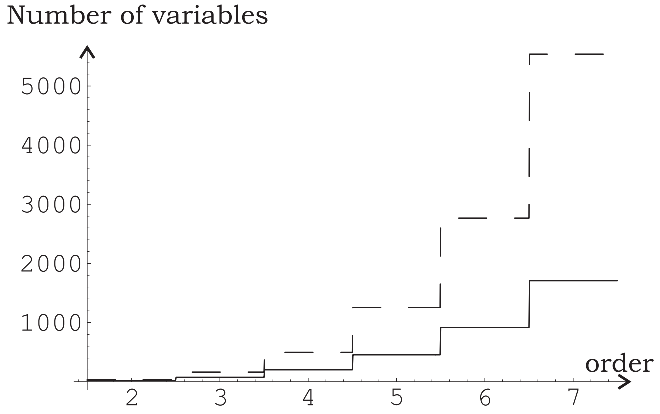


Figure 5.2: Number of variables in the indirect (dashed) and direct (solid) methods.

A combined algorithm

In practice, to solve boundary value problems over a long time span it is most convenient to combine both methods. Typically, we first solve the initial value problem (indirect method) up to a time of interest, say t_1. Then we solve the Hamilton-Jacobi equation (direct approach) about t_1, with initial conditions equal to the values of the generating functions at t_1 found using the indirect approach.

5.4 Convergence and existence of solutions

We now study the convergence properties of our algorithm. In particular, we provide a criterion to evaluate the domain in which the approximation of order N of the generating functions is valid. An example to illustrate this criterion is given.

5.4.1 Theoretical considerations

Recall the general form of a generating function (Eq. (5.6)):

$$F_{I_p,K_r}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t) = \sum_{q=0}^{\infty} \sum_{\substack{i_1,\cdots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1! \cdots i_{2n}!} f^{p,r}_{q,i_1,\cdots,i_{2n}}(t) y_1^{i_1} \cdots y_{2n}^{i_{2n}} \,.$$

Definition V.4 (Radius of convergence). *The radius of convergence of the multi-variable series defining F_{I_p,K_r} at t is the real number R_t such that:*

$$\forall\eta\,,\ 0<\eta<R_t,\ \sum_{q=0}^{\infty}\left(\sum_{\substack{i_1,\cdots,i_{2n}=0\\ i_1+\cdots+i_{2n}=q}}^{q}\frac{1}{i_1!\cdots i_{2n}!}f^{p,r}_{q,i_1,\cdots,i_{2n}}(t)\right)\eta^q$$

converges absolutely and

$$\forall\eta>R_t\,,\ \sum_{q=0}^{\infty}\left(\sum_{\substack{i_1,\cdots,i_{2n}=0\\ i_1+\cdots+i_{2n}=q}}^{q}\frac{1}{i_1!\cdots i_{2n}!}f^{p,r}_{q,i_1,\cdots,i_{2n}}(t)\right)\eta^q\quad \textit{diverges.}$$

The following proposition whose proof can be found in many textbooks concerns the normal convergence of the series. Earlier, we used this result for finding the initial conditions to integrate the Hamilton-Jacobi equation.

Proposition V.5. *Let R_t be the radius of convergence of the multi-variable series defining F_{I_p,K_r} at the time t. Then for all $\eta<R_t$ the series converges normally in $\{y\in\mathbb{R}^{2n}:\|y\|\leq\eta\}$ at t.*

The radius of convergence is not appropriate for studying series of functions as it is a function of time. To remove the time dependency, we define the domain of convergence, a domain $\mathcal{D}$ in $\mathbb{R}\times\mathbb{R}^{2n}$ in which the series converge uniformly.

Definition V.6 (Domain of convergence). *The domain of convergence $\mathcal{D}$ is a region in $\mathbb{R}\times\mathbb{R}^{2n}$ in which the series*

$$\sum_{q=0}^{\infty}\sum_{\substack{i_1,\cdots,i_{2n}=0\\ i_1+\cdots+i_{2n}=q}}^{q}\frac{1}{i_1!\cdots i_{2n}!}f^{p,r}_{q,i_1,\cdots,i_{2n}}(t)y_1^{i_1}\cdots y_{2n}^{i_{2n}}$$

converges uniformly.

In contrast with the radius of convergence, the domain of convergence is not uniquely defined. The spatial domain depends on the time interval and vice versa. For instance,

$\sum_n t^n y^n$ converges if and only if $ty < 1$. $\mathcal{D} = \{[0,2] \times [0,0.5]\}$ and $\mathcal{D} = \{[0,0.5] \times [0,2]\}$ are two well-defined domains of convergence.

In Def. V.6, the uniform convergence of the series is of prime importance. It allows one to bound the error between the true series and its truncation. Indeed, by definition we have:

$$\begin{aligned}&\forall \epsilon > 0\,,\ \exists N > 0\,,\ \forall (t,y) \in \mathcal{D}\,,\\ &F_{I_p,K_r}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t) - \sum_{q=0}^{N} \sum_{\substack{i_1,\cdots,i_{2n}=0\\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}!} f^{p,r}_{q,i_1,\cdots,i_{2n}}(t) y_1^{i_1} \cdots y_{2n}^{i_{2n}} < \epsilon\,.\end{aligned} \tag{5.13}$$

In other words, given a domain of convergence and a precision goal ϵ, there exists a positive integer N such that the truncated Taylor series of order N approximates the true function within ϵ in the domain.

5.4.2 Practical considerations

In practice, for most of the problems we are interested in, we are only able to compute finitely many terms in the series. As a result, it is impossible to estimate a domain of convergence. Worst, we cannot theoretically guaranty that the generating functions can be expressed as Taylor series. In fact, we have seen earlier that even if the Taylor series of F_{I_p,K_r} converges on some open set and F_{I_p,K_r} is smooth, then F_{I_p,K_r} may not be equal to its Taylor series. One can readily verify that the function $f(x) = \exp(1/x^2)$ if $x \neq 0$, $f(0) = 0$ is smooth and has a converging Taylor series at 0. However, f is not equal to its Taylor series. In the following we make two realistic assumptions in order to develop a practical tool for estimating a domain of convergence.

We first assume that the flow may be expressed as a Taylor series in some open set. This is a very common assumption when studying dynamical systems. For example, we make this hypothesis when we approximate the flow by the state transition matrix at linear order. We noticed in the indirect approach that the generating functions may be computed from the flow at the cost of a series inversion. From the series inversion theory (see e.g. Moulton [72]), we conclude that the generating functions can also be expressed as Taylor series (when they are not singular). Thus, for almost every t, there exists a non-zero radius of convergence. In addition, the concept of domain of convergence is well-defined.

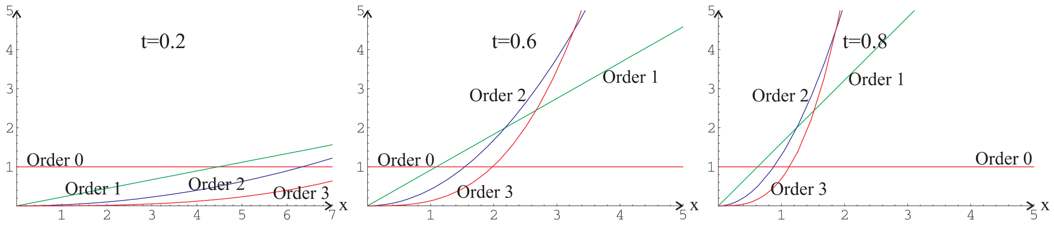
The second assumption we make is also reasonable. We assume that there exists a domain in which the first order terms of the series defining F_{I_p,K_r} are dominant. In other words, we assume that there exists a domain in which the linear order is the largest, followed by the second order, third order etc... This is again a very common assumption for dynamical systems. When approximating the flow with the state transition matrix, we implicitly assume that the linear term is dominant. However, in the present case, there is a subtlety due to the presence of singularities. We observe that this assumption no longer holds as we get closer to a singularity. Let us look at an example to illustrate this phenomenon.

Example V.7. The Taylor series in x of $f(x,t) = (1-t)^x$ for $t \in (0,1)$ is

$$\sum_{r=0}^{\infty} a_n x^n \ , \text{ where } a_n = \frac{\log(1-t)^n}{n!} \ .$$

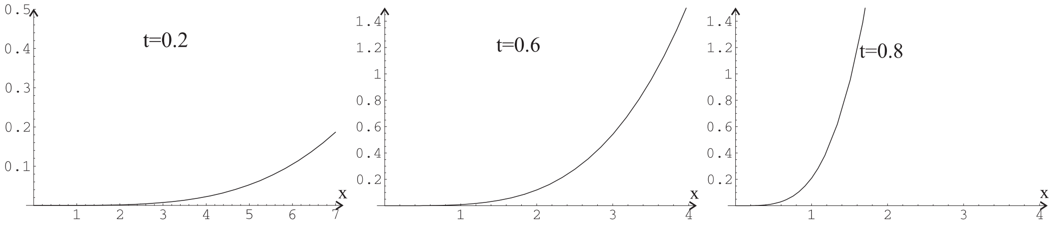
Its radius of convergence is $R_t = \infty$ for all $t \in (0,1)$ and it is singular at $t = 1$. In figure 5.3, we plot the first four terms of the series as a function of x for different times. Clearly, as t gets closer to 1, the first order terms are less and less dominant. Equivalently, the x-interval in which the first order terms are dominant shrinks as t goes to 1. In figure 5.4, we plot $(1-t)^x - \sum_{r=0}^{3} \frac{\log(1-t)^n}{n!} x^n$. One can readily verify that given a prescribed error margin, the domain in which the order 4 approximates f within this margin shrinks as t gets closer to 1. This is a very common behavior that motivates the need for a new criterion.

Suppose that the fourth order approximation of f is to be used for solving a given problem where the time evolves from 0 to 0.6. We know that such an approximation is relevant if the firsts order terms are dominant, i.e., $a_0 > a_1 > a_2 > a_3$. From figure 5.3, we infer that this condition is satisfied if and only if $\|x\| \leq 1$. We call the domain $\mathcal{D}_u = \{[0,1],[0,0.6]\}$ the domain of use.



(a) $t = 0.2$ (b) $t = 0.6$ (c) $t = 0.8$

Figure 5.3: Contribution of the first four terms in the Taylor series of $(1-t)^x$



(a) $t = 0.2$ (b) $t = 0.6$ (c) $t = 0.8$

Figure 5.4: $(1-t)^x - \sum_{n=0}^{3} \frac{\log(1-t)^n}{n!} x^n$

Let us formalize the concept of *domain of use*.

Definition V.8 (Domain of use). *The domain of use $\mathcal{D}_u$ is a domain in $\mathbb{R} \times \mathbb{R}^{2n}$ in which*

$$\left(| \sum_{\substack{i_1,\cdots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1! \cdots i_{2n}} f_{q,i_1,\cdots,i_{2n}}^{p,r}(t) y_1^{i_1} \cdots y_{2n}^{i_{2n}} \right)_q$$

is a decreasing sequence.

This definition is very conservative but very easy to work with. For a given problem, we identify a time interval (or a spatial domain) in which we want to use the generating functions. Then we compute the spatial domain (or the time interval) in which our solution is valid. Once we have identified the domain of use, one can safely work with the solution within this domain. Let us illustrate the use of the above tool with an example.

5.4.3 Examples

We consider the following fictional space mission: A formation of spacecraft is flying about the Libration point L_2 in the nondimensionalized Hill three-body problem (Appendix C) and we wish to use F_1 for solving position to position boundary value problems in order to reconfigure the formation. The mission specifications restrict the spacecraft to stay within 0.05 units of length from L_2 (i.e., $107,500\ km$ in the Earth-Sun system).

We expand the Hamiltonian describing the Hill three-body problem about the equilibrium point L_2 and use the algorithm described previously to solve F_1 up to order 5 in the time interval $(0, 3.5)$. We encounter a number of singularities for F_1 at $t = 0$, $t = 1.68$, and $t = 3.19$ (these were predicted in Section 3.2.3). In Fig. 5.5, we plot the maximum value of $\|y\|$ so that the first five terms are in decreasing order[5]. We notice that as we get closer to the singularity, the maximum value of $\|y\|$ goes to 0. To find the domain of use, we only need to intersect this plot with $\|y\| = 0.05$ and check that we are within the radius of convergence R_t. From Fig. 5.5, we infer that the domain of use is

$$D = \left\{(y,t) \in [-0.05, 0.05]^{2n}, (0.01, 1.32) \cup (1.84, 3.12) \cup (3.12, 3.5)\right\}. \tag{5.14}$$

Error in the approximation We can verify *a posteriori* that the Taylor series expansion found for the generating function F_1 approximates the true dynamics. To do so, we set

[5] Some terms may change sign and therefore may be very small. In that case we ignore these terms so that the decreasing condition can be satisfied (For instance if the order 2 term goes to 0, it will be smaller than any other terms and therefore must be ignored).

$q(T) = q_1$ and q_0, and find $p(T) = p_1$ and p_0 from Eqns. (3.7)-(3.8). Then we integrate the trajectory whose initial condition is (q_0, p_0) to find $(q(T), p(T)) = (q_2, p_2)$. The error in the approximation is defined as the norm of $(q_2 - q_1, p_2 - p_1)$. In Fig. 5.6 we plot this error for $q_0 = 0$ and q_1 that takes values on the circle centered at L_2 of radius 0.05 for different values of t. We observe that the truncated series provide a good approximation of the true dynamics.

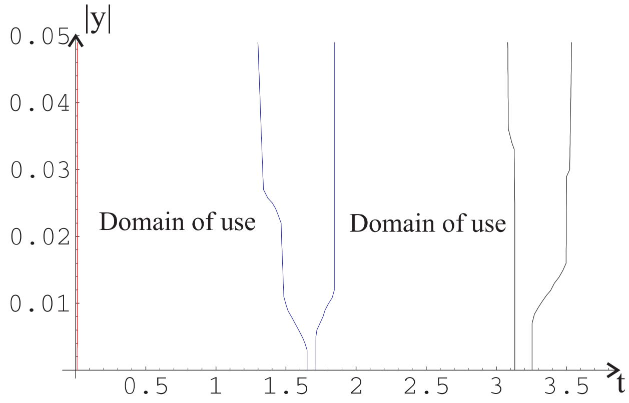
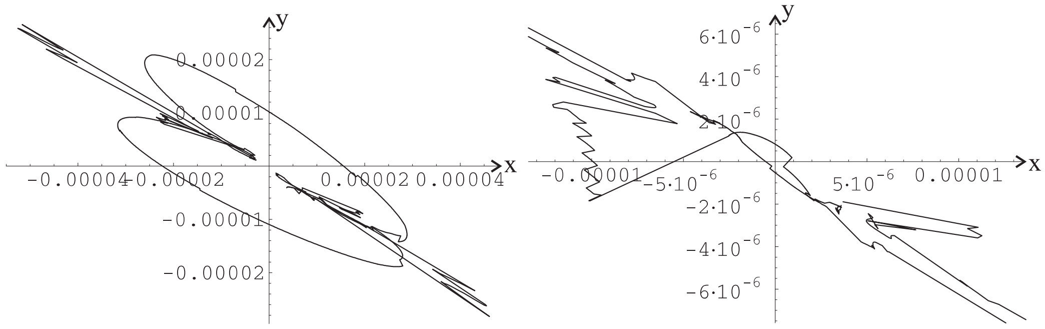


Figure 5.5: Domain of use

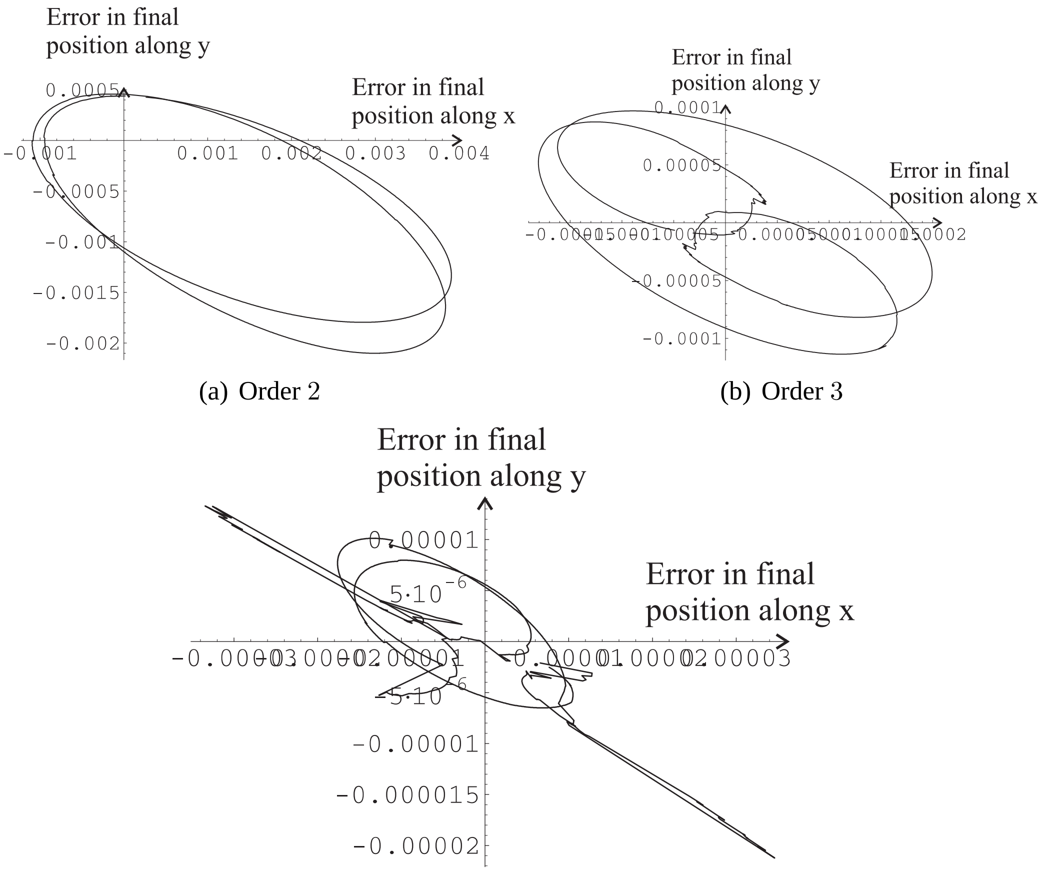


(a) $t = 1.3$

(b) $t = 1$

Figure 5.6: Difference between the true and the approximate dynamics

We also point out that since the series is converging and the magnitude of each order decreases in the domain of use, the accuracy must always increase if an additional order is taken into account. In Fig. 5.7, we observe that the order two solution provides a poor approximation to the initial momentum because the error ranges up to $4.5 \cdot 10^{-3}$ units of length (i.e., $9615\ km$ in the Earth-Sun system). Order three and four give order of magnitude improvements, the error is less than $2.2 \cdot 10^{-4}$ units of length ($480\ km$) for order three and less than $3.5 \cdot 10^{-5}$ units of length ($77\ km$) for order four, over two orders of magnitude better than the order two solution.



(a) Order 2

(b) Order 3

(c) Order 4

Figure 5.7: Error in the normalized final position for $t = 0.9$

CHAPTER VI

THE NUMERICS OF OPTIMAL CONTROL PROBLEMS AND A NOVEL METHOD TO SOLVE OPTIMAL CONTROL PROBLEMS

For a general optimal control problem, necessary conditions for optimality may be derived from the Pontryagin maximum principle. These conditions often take the form of a two-point boundary value problem and are therefore difficult to solve in general. There has been much work on solving this type of problem, some analytical and others numerical. We will not attempt to survey this literature in any systematic fashion as the literature is simply too large, but we can confidently say that numerical techniques almost always require (there are a few exceptions such as methods that consists of solving the Hamilton-Jacobi-Bellman equation) integration of some if not all of the ordinary differential equations given by the Pontryagin maximum principle. To perform this integration, one uses numerical integrators that take an initial condition and move objects in the direction specified by the differential equations. As discussed in Chapter IV, these methods do not exactly satisfy all the physical conservation laws associated with the system. An alternative approach to integration, the theory of geometric integrators (Chapter IV), has been developed. However, integration of the necessary conditions using geometric integrators is usually not possible because they are often coupled with nonlinear equations that would need to be discretized in such a manner that the algorithm keeps its properties. For instance, under some smoothness conditions, the Pontryagin maximum principle yields the following conditions:

$$\begin{aligned} \dot{x} &= D_2 H(x,p,u)\,, && (6.1)\\ \dot{p} &= -D_1 H(x,p,u)\,, && (6.2)\\ 0 &= D_3 H(x,p,u)\,. && (6.3) \end{aligned}$$

To solve this set of equations, one needs to simultaneously solve the ordinary differential equations (Eqns. (6.1) and (6.2)) as well as the (nonlinear) equation (Eq. (6.3)). In Section 6.2, we extend the discrete geometric framework introduced in Chapter IV to overcome this difficulty. Specifically, we are able to state a discrete maximum principle that yields discrete necessary conditions for optimality. Most importantly, these conditions are in agreement with the ones obtained from the Pontryagin maximum principle and define symplectic algorithms. The approach adopted here allows one to recover as a particular case earlier works on symplectic integrators in optimal control such as [16] and to adapt most of the classical symplectic integrators used in dynamics.

Furthermore, if Eq. (6.3) can be solved for the optimal feedback control law, then the boundary value problem defined by the necessary conditions for optimality (Eqns. (6.1)-(6.3)) reduces to a *Hamiltonian* boundary value problem. In Section 6.3, we show that our approach for solving two-point boundary value problems directly applies. In particular, using the generating functions, we obtain an estimate of the initial adjoint variables without an initial guess and therefore solve the optimal control problem. Most important, our approach overcomes some of the barriers to truly reconfigurable control. Specifically, at the cost of algebraic manipulations, we can solve optimal control problems with different boundary conditions as long as the cost function and the dynamics are unchanged.

In the first section of this chapter, we review the maximum principle and derive necessary conditions for optimality. Then, we introduce a discrete maximum principle and show with a few examples how it yields necessary conditions that define symplectic algorithms. Finally, we apply the theory developed in Chapter III to solve optimal control problems for which the optimal feedback control law may be expressed as a function of the state and the adjoint variables. To illustrate this approach, we analyze the linear quadratic controller problem and then a targeting problem using the Hill three-body dynamics.

6.1 Necessary and sufficient conditions for optimality

6.1.1 Problem Statement

Let $J = \int_0^{t_f} g(x,u)dt$ be a performance index (also called a cost function) and consider the following optimal control problem:

$$\min_u \int_{t_0}^{t_f} g(x,u)dt\,, \tag{6.4}$$

subject to the dynamics

$$\dot{x} = f(x,u)\,, \tag{6.5}$$

and to r_i initial and r_f final constraints:

$$\phi_i(x(t_0),t_0) = 0\,, \quad \phi_f(x(t_f),t_f) = 0\,, \tag{6.6}$$

where f and g are functions from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}$ of class C^1, $\phi_i : (R)^n \times \mathbb{R} \to \mathbb{R}^{r_i}$ and $\phi_f : (R)^n \times \mathbb{R} \to \mathbb{R}^{r_f}$

Remark VI.1. Although this optimal control problem is written using the Lagrange formulation, the following readily applies to the Bolza or the Mayer formulations.

$$\min_u K(x(t_f)) + \int_{t_0}^{t_f} L(x,u,t)dt \qquad \text{Bolza formulation} \tag{6.7}$$

$$\min_u K(x(t_f)) \qquad \text{Mayer formulation} \tag{6.8}$$

Remark VI.2. The boundary conditions defined by Eq. (6.6) are in a very general form. They include hard constraint problems (HCP), as well as soft constraint problems (SCP). For HCP, the initial and terminal boundary conditions are fully specified, i.e., $r_i = r_f = n$ whereas $r_i = n$ and $r_f = 0$ for SCP.

6.1.2 Maximum principle

To solve the optimal control problem defined by Eqns. (6.4), (6.5) and (6.6), we apply the Pontryagin principle.

Theorem VI.3 (Maximum principle). *Solutions to the optimal control problem defined by Eqns.* (6.4), (6.5) *and* (6.6) *correspond to critical points of the cost function J in the class of curves $\gamma = (x(t), u(t)) \in \Gamma$ where Γ is the set of curves satisfying Eqns. (6.5) and (6.6).*

Proof. To find critical points of the functional J under the non-holonomic constraints defined by Eqns. (6.5) and (6.6), we must impose the constraints on the velocity vectors of the class of allowable curves (details on non-holonomic variational principle may be gleaned in Bloch and Crouch [15] and Bloch, Bailleul, Crouch and Marsden [14] for instance). Therefore, before taking the variations of the cost function J, we must append the constraints using the Lagrange multipliers. The new function, J_a, is often called the augmented cost function:

$$\begin{aligned} J_a &= \int_{t_0}^{t_f} g(x,u) - \langle p, \dot{x} - f(x,u)\rangle dt + \langle \lambda_i, \phi_i(x(t_0), t_0)\rangle + \langle \lambda_f, \phi_f(x(t_f), t_f)\rangle \\ &= \int_{t_0}^{t_f} H(x,p,u) - \langle p, \dot{x}\rangle dt + \langle \lambda_i, \phi_i(x(t_0), t_0)\rangle + \langle \lambda_f, \phi_f(x(t_f), t_f)\rangle , \end{aligned}$$

where the p's, the λ_i's and the λ_f's are Lagrange multipliers and $H(x,p,u) = g(x,u) + \langle p, f(x,u)\rangle$. Taking variations of the augmented cost function assuming fixed initial and final times yields:

$$\begin{aligned}\delta J_a &= \delta\left(\int_{t_0}^{t_f} H(x,p,u) - \langle p, \dot{x}\rangle dt\right) + \delta\langle \lambda_i, \phi_i(x(t_0), t_0)\rangle \\ &\quad + \delta\langle \lambda_f, \phi_f(x(t_f), t_f)\rangle \\ &= \int_{t_0}^{t_f} \langle D_2 H(x,p,u) - \dot{x}, \delta p\rangle + \langle D_1 H(x,p,u) + \dot{p}, \delta x\rangle \\ &\quad + \langle D_3 H(x,p,u), \delta u\rangle dt + \langle -p(t_f) + D_1\phi_f^T \lambda_f, \delta x_f\rangle \\ &\quad + \langle p(t_i) + D_1 \phi_i^T \lambda_i, \delta x_i\rangle \,.\end{aligned}$$

We now let the variations of J_a be zero to obtain necessary conditions for optimality:

$$\dot{x} = D_2 H(x,p,u)\,, \tag{6.9}$$

$$\dot{p} = -D_1 H(x,p,u)\,, \tag{6.10}$$

$$0 = D_3 H(x,p,u)\,, \tag{6.11}$$

as well as transversality conditions:

$$p(t_i) = -D_1\phi_i(x(t_0), t_0)^T \lambda_i\,,\ p(t_f) = D_1 \phi_f(x(t_f), t_f)^T \lambda_f\,. \tag{6.12}$$

Eqns. (6.9)-(6.12) define the necessary conditions for optimality. □

Eqns. (6.9) and (6.10) are $2n$ ordinary differential equations coupled with m (nonlinear) equations defined by Eq. (6.11). To solve these equations we need $2n$ boundary conditions. On one hand, r_i initial and r_f final conditions are given in the problem statement. On the other hand, the transversality conditions yield n initial and n final conditions but introduce r_i unknowns λ_i and r_f unknows λ_f. Thus, we obtain $2n$ boundary conditions as well as $r_i + r_f$ equations that allows us to solve for (λ_i, λ_f). As a result, the necessary conditions obtained by the maximum principle define a well-posed problem.

Remark VI.4. This formulation of the necessary conditions differs from the one given by Pontryagin [78] but the main point of the Pontryagin principle is that it yields necessary conditions for optimality under far less severe regularity conditions. The above formulation is based on the equivalence between the Pontryagin principle and the calculus of variations in the case where the control region is an open set in a finite dimensional vector space (see [78] chapter V for more details). It is therefore equivalent to classical variational formulations given in Bloch et al. [14, 15] and Gregory and Lin [29] for instance.

The necessary conditions are of the same form as Hamilton's equations but are coupled with a nonlinear equation (Eq. (6.11)). We have seen previously that Hamiltonian systems, i.e., Hamilton's equations, can be integrated using symplectic integrators. However, if Hamilton's equations are coupled with algebraic nonlinear equations the theory no longer applies. What is the correct discretization of the algebraic equation? In the next section, we develop a discrete maximum principle that tackles this problem and provides a unified view on solving optimal control problems using symplectic integrators.

6.2 Discretization of optimal control problems

We propose two methods to discretize the necessary conditions for optimality. The first, most intuitive one, has several inherent drawbacks that we point out. The second method requires the use of a discrete maximum principle that we present. It is more general and we show, with a few examples, that it yields necessary conditions that define symplectic algorithms. Furthermore, this second approach can be enhanced to yield symplectic-energy conserving algorithms. Finally we prove that the discrete necessary conditions are in agreement with the necessary conditions obtained from the Pontryagin maximum principle. We illustrate this equivalence with an example from sub-Riemannian optimal control problems.

6.2.1 Solving the necessary conditions for optimality

The first method we propose to discretize the necessary conditions assumes that we can find the optimal feedback control law as a function of the x's and the p's. More precisely, suppose Eq. (6.11) allows one to solve for u as a function of (x, p) and define the Hamiltonian function

$$\bar{H}(x, p) = H(x, p, u(x, p))\,. \tag{6.13}$$

Then the necessary conditions (6.9) and (6.10) simplify to:

$$\dot{x} \;=\; D_2\bar{H}(x, p)\,, \tag{6.14}$$

$$\dot{p} \;=\; -D_1\bar{H}(x, p)\,. \tag{6.15}$$

Equations (6.14) and (6.15) are of the same form as the Hamilton equations. Therefore, the system defined by $\bar{H}$ is Hamiltonian, and is better simulated using symplectic algorithms (see Chap. IV). We point out that this Hamiltonian system has no physical meaning in general and may even not be Lagrangian. For example, we show later that $\bar{H}$ is not hyperregular for sub-Riemannian optimal control problems. As a result, the Legendre transform is ill-defined and we cannot define a Lagrangian function associated with the Hamiltonian $\bar{H}$. This fact has many consequences, for instance DVPI (DVPI and DVPII are defined in Section 4.2) cannot be used to discretize such systems whereas one could use DVPII (DVPI acts on the tangent bundle only whereas DVPII has two formulations, one on the tangent bundle for Lagrangian systems and one the co-tangent bundle for Hamiltonian systems (Def. IV.4)).

Example VI.5 (Midpoint scheme). To integrate the necessary conditions using the midpoint scheme, we apply the modified discrete Hamilton's principle (Def. IV.4) to the

Hamiltonian system defined by $\bar{H}$ using the midpoint Leibnitz law (Eq. (4.29)). We obtain:

$$\frac{x_{k+1}-x_k}{h} = D_2\bar{H}(\frac{x_{k+1}+x_k}{2}, \frac{p_{k+1}+p_k}{2})\,, \tag{6.16}$$

$$\frac{p_{k+1}-p_k}{h} = -D_1\bar{H}((\frac{x_{k+1}+x_k}{2}, \frac{p_{k+1}+p_k}{2})\,. \tag{6.17}$$

Lemma IV.7 guarantees the symplectic nature of this implicit algorithm.

6.2.2 Discrete maximum principle

If the feedback control law cannot be solved from Eq. (6.11), then the above method to discretize the necessary conditions no longer applies. In this section we address this issue. Specifically, we introduce a discrete maximum principle that allows us to derive discrete necessary conditions for optimality that are in agreement with the one obtained from the maximum principle.

Problem statement

We assume the same geometric framework than in Chapter IV, that is, we consider a discretization of the time t into n instants $\mathcal{T} = \{(t_k)_{k\in[1,n]}\}$. $t_{k+1} - t_k$ may not be equal to $t_k - t_{k-1}$ in general but for sake of simplicity, we assume in the following that $t_{k+1} - t_k = \tau$, $\forall k \in [1, n]$. At t_k, x_k lies in the n-dimensional vector space $M_k = \mathbb{R}^n$, u_k lies in $U_k = \mathbb{R}^m$ and we set $\mathcal{M} = \bigcup M_k$ and $\mathcal{U} = \bigcup U_k$. On $\mathcal{T}$, we define a discrete time derivative operator Δ^d_τ. Δ^d_τ may not verify the usual Leibnitz law but a modified one. We denote by x^d_k and u^d_k two points in $\mathcal{M}$ and $\mathcal{U}$ respectively. Later we give an explicit definition of these points but so far we only need to know that x^d_k can be expressed as a function of x_k and x_{k+1} (and u^d_k can be expressed as a function of u_k and u_{k+1}).

In discrete settings, the cost function is

$$J = \sum_{k=0}^{n-1} g_d(x^d_k, u^d_k)\tau\,,$$

and the optimal control problem (6.4) is formulated as:

$$\min_{u_k^d} \sum_{k=0}^{n-1} g_d(x_k^d, u_k^d)\tau \,, \tag{6.18}$$

subject to the dynamics

$$\Delta_\tau^d x_k^d = f_d(x_k^d, u_k^d)\,, \tag{6.19}$$

and to $r_i + r_f$ boundary conditions:

$$\phi_i(x_0, t_0) = 0\,, \quad \phi_f(x_n, t_n) = 0\,, \tag{6.20}$$

where f_d and g_d are functions from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}$ of class C^1 that correspond to discretization of the continuous time functions f and g, $\phi_i : (R)^n \times \mathbb{R} \to \mathbb{R}^{r_i}$ and $\phi_f : (R)^n \times \mathbb{R} \to \mathbb{R}^{r_f}$

Discrete maximum principle

To obtain necessary conditions for optimality, we define the following discrete maximum principle, the discrete counterpart of the Pontryagin maximum principle:

Definition VI.6 (Discrete maximum principle). *Solutions to the discrete optimal control problem correspond to critical points of the cost function J in the class of discrete curves $\gamma \in \Gamma$, where Γ is the set of all discrete curves $(x_k, u_k)_{k\in[1,n]}$ that verify Eqns. (6.19) and (6.20).*

Remark VI.7. The above definition is the discrete counterpart of Thm. VI.3. It compares to previous works on discrete optimal control theory that extend the Pontryagin maximum principle to discrete systems such as Jordan and Polak [55] as Thm. VI.3 compares to the Pontryagin maximum principle. In other words, in contrast with Jordan and Polak [55], we restrict the class of discrete optimal control problems so that we can derive necessary conditions that define symplectic algorithms.

As in the continuous case, to find critical points of J under the non-holonomic constraints defined by Eqns. (6.19) and (6.20), we must append the constraints to J using the Lagrange multipliers. The resulting function is called the augmented cost function:

$$\begin{aligned} J_a &= \sum_{k=0}^{n-1} (g_d(x_k^d, u_k^d) - \langle p_k^d, \Delta_\tau^d x_k^d - f_d(x_k^d, u_k^d)\rangle)\tau + \langle \lambda_0, \phi_0 \rangle + \langle \lambda_n, \phi_n \rangle \qquad (6.21) \\ &= \sum_{k=0}^{n-1} (H_d(x_k^d, p_k^d, u_k^d) - \langle p_k^d, \Delta_\tau^d x_k^d \rangle)\tau + \langle \lambda_0, \phi_0 \rangle + \langle \lambda_n, \phi_n \rangle \,, \qquad (6.22) \end{aligned}$$

where the p_k's, the λ_0's and the λ_n's are Lagrange multipliers and $H_d(x_k^d, p_k^d, u_k^d) = g_d(x_k^d, u_k^d) + \langle p_k^d, f_d(x_k^d, u_k^d)\rangle$. To apply the discrete maximum principle, one needs to specify the discrete derivative operator as well as the expressions of x_k^d, u_k^d and p_k^d as a function of (x_{k+1}, x_k), (u_{k+1}, u_k) and (p_{k+1}, p_k) respectively.

Examples

Störmer's rule If we choose Δ_τ^d to be the forward difference Δ_τ and $x_k^d = x_k$, $p_k^d = p_{k+1}$, $u_k^d = u_k$, then we recover the discrete maximum principle developed by Bloch, Crouch, Marsden and Ratiu [16].

$$\begin{aligned} \delta J_a &= \delta \left(\sum_{k=0}^{n-1} (H_d(x_k^d, p_k^d, u_k^d) - \langle p_k^d, \Delta_\tau^d x_k^d \rangle)\tau \right) + \delta\langle \lambda_0, \phi_0 \rangle + \delta \langle \lambda_n, \phi_n \rangle \\ &= \sum_{k=0}^{n-1} \langle D_2 H_d(x_k, p_{k+1}, u_k) - \Delta_\tau x_k, \delta p_{k+1} \rangle \tau \\ &\quad + \langle D_1 H_d(x_k, p_{k+1}, u_k) + \Delta_\tau p_k, \delta x_k \rangle \tau + \langle D_3 H_d(x_k, p_{k+1}, u_k), \delta u_k \rangle \tau \\ &\quad + \langle \phi_0, \delta\lambda_0 \rangle + \langle \phi_n, \delta \lambda_n \rangle + \langle -p_n + D_1 \phi_n^T \lambda_n, \delta x_n \rangle + \langle p_0 + D_1 \phi_0^T \lambda_0, \delta x_0 \rangle \,, \end{aligned}$$

where the modified Leibnitz law (Eq. (4.1)) has been used. We impose that the variation of the augmented cost function be zero to obtain the discrete necessary conditions for optimality and the transversality conditions:

$$\Delta_\tau x_k = D_2 H_d(x_k, p_{k+1}, u_k)\,, \tag{6.23}$$

$$\Delta_\tau p_k = -D_1 H_d(x_k, p_{k+1}, u_k)\,, \tag{6.24}$$

$$0 = D_3 H_d(x_k, p_{k+1}, u_k)\,, \tag{6.25}$$

$$p_0 = -D_1\phi_0(x_0, t_0)^T\lambda_0\,, \qquad p_n = D_1\phi_n(x_n, t_n)^T\lambda_n\,. \tag{6.26}$$

The algorithm defined by Eqns. (6.23), (6.24) and (6.25) is equivalent to the one derived by Bloch, Crouch, Marsden and Ratiu [16] for the symmetric rigid body. Our approach generalizes the discrete varaitional principle developed in [16]. We now prove the symplectic nature of the above algorithm.

Lemma VI.8. *The algorithm defined by Eqns. (6.23), (6.24) and (6.25) is symplectic.*

Proof. Define the cost function $\bar{J}_a$ as:

$$\bar{J}_a = \sum_{k=0}^{n-1} (H_d(x_k, p_{k+1}, u_k) - \langle p_{k+1}, \Delta_\tau x_k \rangle)\tau\,.$$

$\bar{J}_a$ is the augmented cost function from which we have removed the boundary conditions. Boundary conditions yield transversality conditions, that is conditions on the initial and final states of the system. Hence these terms are irrelevant to the study of the advance map $(x_k, p_k, u_k) \mapsto (x_{k+1}, p_{k+1}, u_{k+1})$. As in discrete dynamics, we consider $d^2 J_a$, assuming (x_k, p_k, u_k) verifies the above necessary conditions and we obtain:

$$d\bar{J}_a = \sum_{k=0}^{n-1} \Delta_\tau \langle p_k, dx_k \rangle \tau\,.$$

From $d^2 = 0$, we conclude:

$$0 = \sum_{k=0}^{n-1} \Delta_\tau d\langle p_k, dx_k \rangle \tau\,, \text{ that is, } \forall k \in [0, n-1]\,,\; dp_{k+1} \wedge dx_{k+1} = dp_k \wedge dx_k\,.$$

□

The symplectic nature of the algorithm is obtained directly from the variational principle - there is no need to compute $dp_k \wedge dx_k$ and $dp_{k+1} \wedge dx_{k+1}$.

Midpoint scheme Midpoint discretization may also be obtained by choosing

$$x_k^d = \frac{x_{k+1} + x_k}{2}\,,\; p_k^d = \frac{p_{k+1} + p_k}{2}\,,\; u_k^d = \frac{u_{k+1} + u_k}{2}\,.$$

and $\Delta_\tau^d = R_{\tau/2} - R_{-\tau/2}$. One can readily verify that the discrete maximum principle yields the following necessary conditions for optimality and transversality conditions:

$$\Delta_\tau^d x_k^d = D_2 H_d(x_k^d, p_k^d, u_k^d)\,, \tag{6.27}$$

$$\Delta_\tau^d p_k^d = -D_1 H_d(x_k^d, p_k^d, u_k^d)\,, \tag{6.28}$$

$$0 = D_3 H_d(x_k^d, p_k^d, u_k^d)\,, \tag{6.29}$$

$$p_0 = D_1\phi_0(x_0, t_0)^T\lambda_0\,, \qquad p_n = -D_1\phi_n(x_n, t_n)^T\lambda_n\,. \tag{6.30}$$

Lemma VI.9. *The algorithm defined by Eqns. (6.27), (6.28) and (6.29) is symplectic.*

Proof. We omit the proof since it proceeds as before. □

6.2.3 Discrete maximum principle v.s. discretization of the Pontryagin maximum principle

So far we have considered two methods for obtaining a symplectic algorithm that integrates the necessary conditions for optimality. The first method, which applies only to a certain class of problems, consists of discretizing the necessary conditions obtained from the Pontryagin maximum principle once the control has been expressed as a function of (x, p). The second method consists of using the discrete maximum principle. In this section, we show that under certain assumptions both methods are equivalent, that is, we prove the commutative diagram (6.31).

$$\begin{tikzcd}
\begin{array}{c} \min_u \int_0^T g(x,u)dt \\ \dot{x} = f(x,u) \end{array} \arrow[r] \arrow[d, "(PMP)"] & \begin{array}{c} \min_u \sum_{k=0}^{n-1} g_d(x_k^d, u_k^d) \\ \Delta_\tau^d x_k^d = f_d(x_k^d, u_k^d) \end{array} \arrow[d, "(DMP)"] \\
\begin{array}{c} H(x,p,u) \\ \bar{H}(x,p) \end{array} \arrow[r, "(DMHP)"] & \begin{array}{c} H_d(x_k^d, p_k^d, u_k^d) \\ \bar{H}_d(x_k^d, p_k^d) \end{array}
\end{tikzcd} \tag{6.31}$$

where $\bar{H}$ is defined by (Eq. (6.13)), DMHP stands for the discrete modified Hamilton's principle, PMP stands for the Pontryagin maximum principle, and DMP stands for the discrete maximum principle.

We recall the required assumptions to prove the equivalence of the diagram. We assume that Eq. (6.11) can be solved for u as a function of (x,p) and that the initial and final states $x(t_f) = x_f$ and $x(t_0) = x_i$ are given. In addition, we impose $g_d = g$ and $f_d = f$.

To discretize the Hamiltonian system defined by $\bar{H}$, we use the discrete modified Hamilton's principle:

$$0 = \delta S_d^H = \delta \left(\tau \sum_{k=0}^{n-1} \langle p_k^d, \Delta_\tau^d x_k^d \rangle - \bar{H}(x_k^d, p_k^d) \right) , \tag{6.32}$$

for any variations of (x_k^d, p_k^d) and $\delta x_0 = \delta x_n = 0$. One can readily check that Eq. (6.32) can also be written in an equivalent form as:

$$0 = \delta S_d^H = \delta \left(\tau \sum_{k=0}^{n-1} \langle p_k^d, \Delta_\tau^d x_k^d \rangle - H(x_k^d, p_k^d, u_k^d) \right) ,$$

for any variations of (x_k^d, p_k^d, u_k^d) and $\delta x_0 = \delta x_n = 0$ where u_k^d is now considered as an independent variable. In addition since $f = f_d$ and $g = g_d$, $H = H_d$, and we conclude that the discrete modified Hamilton's principle as formulated and the discrete maximum principle are equivalent.

6.2.4 The Heisenberg optimal control problem

The Heisenberg problem (Brockett [18], Bloch et al. [14]) refers to under actuated optimal control problems which are controllable. For instance, consider a particle that has two actuators in the (x, y)-plane and with velocity in the z direction defined by $\dot{z} = y\dot{x} - x\dot{y}$. This system is controllable, however, to reach a point $(a > 0, 0, 0)$ from the origin $(0, 0, 0)$ requires a non-trivial control vector. In the following, we study the Heisenberg problem to illustrate the approaches we have developed above. For this problem the cost function is given by:

$$J = \min_{u=(u_1,u_2)} \int_{t_0}^{t_f} \langle u, u \rangle dt \,,$$

subject to

$$\begin{aligned} \dot{x} &= u \,, \\ \dot{y} &= v \,, \\ \dot{z} &= uy - vx \,, \end{aligned}$$

and to the boundary conditions:

$$(x(t_0), y(t_0), z(t_0)) = (0, 0, 0) \,, \; (x(t_f), y(t_f), z(t_f)) = (a > 0, 0, 0) \,.$$

Define H as

$$H(q, p, u) = \frac{1}{2}\langle u, u \rangle + \langle p, \dot{q} \rangle \,,$$

where $q = (x, y, z)$ and $p = (p_x, p_y, p_z)$. The Pontryagin maximum principle yields:

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}(q, p, u) \,, && (6.33) \\ \dot{p} &= -\frac{\partial H}{\partial q}(q, p, u) \,, && (6.34) \\ 0 &= \frac{\partial H}{\partial u}(q, p, u) \,. && (6.35) \end{aligned}$$

with boundary conditions

$$(x(t_0), y(t_0), z(t_0)) = (0,0,0)\,,\ (x(t_f), y(t_f), z(t_f)) = (a > 0, 0, 0)\,.$$

Note this is a hard constraint problem, therefore the transversality conditions are of no use; They yield $2n$ equations but introduce $2n$ new variables (λ_i, λ_f). Eq. (6.35) allows us to solve for u as a function of (q,p):

$$u_1 = p_x + p_z y\,,\ u_2 = p_y - p_z x\,.$$

Hence, Eqns. (6.33)-(6.34) become:

$$\dot{q} = \frac{\partial \bar{H}}{\partial p}(q,p)\,, \tag{6.36}$$

$$\dot{p} = -\frac{\partial \bar{H}}{\partial q}(q,p)\,, \tag{6.37}$$

where

$$\begin{aligned} \bar{H}(q,p) &= H(q,p,u(q,p)) \\ &= -\frac{1}{2}(p_x^2 + p_y^2) - p_x p_z y + p_y p_z x\,. \end{aligned} \tag{6.38}$$

Eqns. (6.36) and (6.37) are of the same form as the Hamilton equations. Therefore, the necessary conditions for optimality yield a Hamiltonian system with Hamiltonian function $\bar{H}$. We now prove that $\bar{H}$ is degenerate at the origin, and so is the Legendre transform. The Hessian of $\bar{H}$ is:

$$\left(\frac{\partial \bar{H}}{\partial (q,p)}\right) = \begin{pmatrix} -1 & 0 & -y \\ 0 & -1 & x \\ -y & x & 0 \end{pmatrix}$$

Thus, $\det\left(\frac{\partial \bar{H}}{\partial (q,p)}\right) = x^2 + y^2$, i.e., the determinant of the Hessian of $\bar{H}$ is singular at $(0,0)$. As a result, it is not, *a priori*, possible to define a Lagrangian function associated with the Hamiltonian $\bar{H}$ using the Legendre transform[1]. Therefore, the discrete modified Hamilton's principles (DMHP) must be used to discretize Eqns. (6.36) and (6.37). One cannot use a discrete Hamilton's principles (DHP) for instance because the system is not Lagrangian. This point is of importance. It motivates the need to introduce the variational principles presented in Chapter IV, as previous works on variational principles focused on systems with non-degenerate Lagrangian functions.

To discretize the necessary conditions, we choose the geometry associated with the Störmer rule and use the DMHP (Def. IV.4) to eventually find the following symplectic algorithm:

$$\begin{aligned} \Delta_\tau q_k &= D_2\bar{H}(q_k, p_{k+1})\,, &(6.39)\\ \Delta_\tau p_k &= -D_1\bar{H}(q_k, p_{k+1})\,. &(6.40)\end{aligned}$$

Let us now discretize the Heisenberg problem using the second approach, based on the use of the discrete maximum principle. We first discretize the problem statement:

$$\min_{u_k=(u_{1,k},u_{2,k})} \frac{1}{2}\sum_{k=0}^{n-1}\langle u_k, u_k\rangle\,,$$

subject to

$$\begin{aligned} \Delta_\tau x_k &= u_{1,k}\,,\\ \Delta_\tau y_k &= u_{2,k}\,,\\ \Delta_\tau z_k &= u_{1,k}y_k - u_{2,k}x_k\,. \end{aligned}$$

Define the discrete augmented cost function J_a:

$$J_a = \sum_{k=1}^{n-1} H_d(q_k, p_{k+1}, u_k) - \langle p_{k+1}, \Delta_\tau q_k\rangle\,,$$

[1] Using Lagrange multipliers one can define a Legendre transform and find a Lagrangian function associated with the system. We refer to Bloch [14] for a presentation of this technique that involves variational principles with constraints.

where $H_d(q_k, p_{k+1}, u_k) = \langle u_k, u_k\rangle + \langle p_{k+1}, q_k\rangle$, $u_k = (u_{1,k}, u_{2,k})$ and $q_k = (x_k, y_k, z_k)$. To find discrete necessary conditions for optimality we set the variations of J_a to zero, and we obtain:

$$\Delta_\tau q_k = D_2 H_d(q_k, p_{k+1}, u_k)\,, \tag{6.41}$$

$$\Delta_\tau p_k = -D_1 H_d(q_k, p_{k+1}, u_k)\,, \tag{6.42}$$

$$0 = D_3 H_d(q_k, p_{k+1}, u_k)\,. \tag{6.43}$$

Eq. (6.41) allows us to find u_k as a function of (q_k, p_{k+1}):

$$u_{1,k} = p_{x,k+1} + p_{z,k+1} y_k\ ,\ u_{2,k} = p_{y,k+1} - p_{z,k+1} x_k\ . \tag{6.44}$$

We then substitute these expressions into Eqns. (6.41)-(6.42):

$$\Delta_\tau q_k = D_2 \bar{H}_d(q_k, p_{k+1})\,, \tag{6.45}$$

$$\Delta_\tau p_k = -D_1 \bar{H}_d(q_k, p_{k+1})\,, \tag{6.46}$$

where $\bar{H}_d(q_k, p_{k+1}) = H_d(q_k, p_{k+1}, u_k(q_k, p_{k+1}))$. By virtue of the commutative diagram, Eqns. (6.45) and (6.46) define the same symplectic algorithm as Eqns. (6.36) and (6.37).

In this example, we chose a trivial discretization of the dynamics and of the cost function; $f = f_d$ and $g = g_d$. Other algorithms may be obtained using non-trivial discretizations. In that case the equivalence principle may not hold but the algorithm we obtain will still be symplectic. Finally, as in discrete dynamics, the discrete maximum principle may be modified in order to yield symplectic-energy conserving algorithms. We detail the procedure in the next section.

6.2.5 Energy conservation

We have seen that the discrete maximum principle (Def. VI.6) allows one to derive necessary conditions that define symplectic algorithms. Using different definitions for the derivative operator and for the variables (x_k^d, p_k^d, u_k^d), one is able to adapt classical symplectic algorithms to optimal control problems. In general these algorithms are not energy preserving, and we now show how the discrete maximum principle may be modified so that the discrete necessary conditions yield symplectic-energy preserving algorithms.

Generalized discrete maximum principle

In contrast with the discrete maximum principle (Def. VI.6), we allow the time step to vary, the time now plays the same role as the state vector x and we introduce an independent parameter τ_k such that $t_k = t(\tau_k)$, $x_k = x(\tau_k)$ and $\tau_{k+1} - \tau_k = \tau$. The configuration space M_k is now $\mathbb{R}^n \times \mathbb{R}$ and $\mathcal{T} = \left\{(\tau_k)_{k\in[1,n]}\right\}$. One must pay attention to the definition of the cost function since $t_{k+1} - t_k$ no longer equals the constant τ:

$$J = \sum_{k=0}^{n-1} g_d(x_k^d, u_k^d)(t_{k+1} - t_k) = \sum_{k=0}^{n-1} g_d(x_k^d, u_k^d)\Delta_\tau t_k \tau\,. \tag{6.47}$$

In the same manner, the dynamics of the system becomes (Δ_τ^d is now the derivative operator with respect to τ whereas the dynamics is given as function of the discrete derivative of x with respect to time):

$$\frac{\Delta_\tau^d x_k^d}{\Delta_\tau^d t_k^d} = f_d(x_k^d, u_k^d)\,,$$

or equivalently

$$\Delta_\tau^d x_k^d = \Delta_\tau^d t_k^d f_d(x_k^d, u_k^d)\,. \tag{6.48}$$

The boundary conditions are left unchanged:

$$\phi_0(x_0, t_0) = 0\,, \quad \phi_n(x_n, t_n) = 0\,. \tag{6.49}$$

The generalized discrete maximum principle reads as follows:

Definition VI.10 (Generalized discrete maximum principle). *Solutions to the discrete optimal control problem defined by Eqns.* (6.47), (6.48) *and* (6.49) *correspond to critical* *points of the cost function J in the class of discrete curves $\gamma \in \Gamma$, where Γ is the set of all curves $(x_k, u_k, t_k)_{k\in[1,n]}$ that verify Eqns. (6.48), (6.49), $t_n = t_f$ and $t_0 = t_i$ (i.e., fixed initial and final times).*

Again we need to define the augmented cost function to apply the constraints:

$$\begin{aligned} J_a &= \tau \sum_{k=0}^{n-1} g_d(x_k^d, u_k^d)\Delta_\tau^d t_k^d - \langle p_k^d, \Delta_\tau^d x_k^d - \Delta_\tau^d t_k^d f_d(x_k^d, u_k^d)\rangle + \langle \lambda_0, \phi_0\rangle + \langle \lambda_n, \phi_n\rangle \\ &= \tau \sum_{k=0}^{n-1} \Delta_\tau^d t_k^d H_d(x_k^d, p_k^d, u_k^d) - \langle p_k^d, \Delta_\tau^d x_k^d\rangle + \langle \lambda_0, \phi_0\rangle + \langle \lambda_n, \phi_n\rangle\,, \qquad (6.50) \end{aligned}$$

where the p_k's, the λ_0's and the λ_n's are Lagrange multipliers and $H_d(x_k^d, p_k^d, u_k^d) = g_d(x_k^d, u_k^d) + \langle p_k^d, f_d(x_k^d, u_k^d)\rangle$.

Example: the Störmer rule

We only go through the derivation of Störmer type of algorithm. One can derive other symplectic-energy conserving algorithms using the same methodology. For the Störmer rule, Δ_τ^d is the forward difference operator, $x_k^d = x_k$, $p_k^d = p_{k+1}$ and $u_k^d = u_k$. Variations of the augmented Lagrangian read:

$$\begin{aligned} \delta J_a &= \delta\left(\sum_{k=0}^{n-1}(\Delta_\tau t_k H_d(x_k^d, p_k^d, u_k^d) - \langle p_k^d, \Delta_\tau^d x_k^d\rangle)\tau\right) + \delta\langle \lambda_0, \phi_0\rangle + \delta\langle \lambda_n, \phi_n\rangle \\ &= \sum_{k=0}^{n-1}\langle \Delta_\tau t_k D_2 H_d(x_k, p_{k+1}, u_k) - \Delta_\tau x_k, \delta p_{k+1}\rangle\tau \\ &\quad + \langle \Delta_\tau t_k D_1 H_d(x_k, p_{k+1}, u_k) + \Delta_\tau p_k, \delta x_k\rangle\tau + \langle \Delta_\tau t_k D_3 H_d(x_k, p_{k+1}, u_k), \delta u_k\rangle\tau \\ &\quad - \sum_{k=1}^{n-1}\Delta_\tau H_d(x_{k-1}, p_k, u_{k-1})\delta t_k \tau + H_d(x_{n-1}, p_n, u_{n-1})\delta t_n - H_d(x_0, p_1, u_0)\delta t_0 \\ &\quad + \langle -p_n + D_1\phi_n^T\lambda_n, \delta x_n\rangle + \langle p_0 + D_1\phi_0^T\lambda_0, \delta x_0\rangle + \langle \phi_0, \delta\lambda_0\rangle + \langle \phi_n, \delta\lambda_n\rangle \quad (6.51) \end{aligned}$$

Since the variations of the augmented cost function must be zero for any δt_k, δx_k, δp_k and $\delta t_0 = \delta t_n = 0$, we obtain the discrete necessary conditions for optimality:

$$\begin{aligned} \Delta_\tau x_k &= \Delta_\tau t_k D_2 H_d(x_k, p_{k+1}, u_k)\,, && (6.52)\\ \Delta_\tau p_k &= -\Delta_\tau t_k D_1 H_d(x_k, p_{k+1}, u_k)\,, && (6.53)\\ 0 &= D_3 H_d(x_k, p_{k+1}, u_k)\,, && (6.54)\\ 0 &= \Delta_\tau H_d(x_{k-1}, p_k, u_{k-1})\,, && (6.55) \end{aligned}$$

as well as transversality conditions:

$$\begin{aligned} p_0 &= -D_1\phi_0(x_0, t_0)^T \lambda_0\,, && (6.56)\\ p_n &= D_1\phi_n(x_n, t_n)^T \lambda_n\,. && (6.57) \end{aligned}$$

Let $e_k = -H_d(x_{k-1}, p_k, u_{k-1})$ define the energy at τ_k and

$$\theta_k = \langle p_k, dx_k\rangle - H_d(x_{k-1}, p_k, u_{k-1}) dt_k$$

be the contact one-form.

Theorem VI.11. *The algorithm defined by Eqns. (6.52), (6.53), (6.54) and (6.55) defines a symplectic-energy conserving algorithm, i.e.,* $e_{k+1} = e_k$ *and* $\omega_{k+1} = \omega_k$ *where* $\omega_k = d\theta_k$*.*

Proof. Eq. (6.55) is equivalent to $e_{k+1} = e_k$, so the algorithm is energy conserving. Let us prove that the symplectic two-form is also preserved. Again we define the augmented cost function $\bar{J}_a$:

$$\bar{J}_a = \sum_{k=0}^{n-1} (\Delta_\tau t_k H_d(x_k, p_{k+1}, u_k) - \langle p_{k+1}, \Delta_\tau x_k\rangle)\tau\,. \qquad (6.58)$$

$\bar{J}_a$ is the augmented cost function to which we have withdrawn the boundary conditions. Consider a discrete trajectory (q_k, p_k, u_k, t_k) that satisfies Eqns. (6.52), (6.53), (6.54) and

(6.55) and let us compute the one-form $d\bar{J}_a$:

$$\begin{aligned} d\bar{J}_a &= d\sum_{k=0}^{n-1}(\Delta_\tau t_k H_d(x_k, p_{k+1}, u_k) - \langle p_{k+1}, \Delta_\tau x_k\rangle)\tau \\ &= \sum_{k=0}^{n-1}\langle \Delta_\tau t_k D_2 H_d(x_k, p_{k+1}, u_k) - \Delta_\tau x_k, dp_{k+1}\rangle\tau \\ &\quad + \langle \Delta_\tau t_k D_1 H_d(x_k, p_{k+1}, u_k) + \Delta_\tau p_k, dx_k\rangle\tau + \langle \Delta_\tau t_k D_3 H_d(x_k, p_{k+1}, u_k), du_k\rangle\tau \\ &\quad - \Delta_\tau(e_k t_k) + \Delta_\tau e_k t_k - \Delta_\tau\langle p_k, dx_k\rangle\tau\,, \end{aligned} \tag{6.59}$$

where the modified Leibnitz law (Eq. (4.1)) has been used. Since (q_k, p_k, u_k, t_k) is a solution to the necessary conditions, Eq. (6.59) reduces to:

$$d\bar{J}_a = -\sum_{k=0}^{n-1}\Delta_\tau(\langle p_k, dx_k\rangle + e_k t_k)\tau\,,$$

and from $d^2 = 0$, we conclude

$$\sum_{k=0}^{n-1}\Delta_\tau d(\langle p_k, dx_k\rangle - H_d(x_{k-1}, p_k, u_{k-1})t_k)\tau = 0\,,$$

that is, $\forall k \in [0, n-1]\,,\ d\theta_{k+1} = d\theta_k$. □

In this section we have developed a new approach to solve optimal control problem. Using discrete geometry we have been able to develop a unified theory to solve optimal control problems using symplectic integrators. We have introduced a discrete maximum principle that yields discrete necessary conditions for optimality. These conditions are in agreement with the ones derived from the Pontryagin maximum principle and define symplectic algorithms. We have also shown that the discrete maximum principle can be enhanced to yield symplectic-energy conserving algorithms. Now, we focus on a specific class of optimal control problem, those for which the optimal control law can be expressed as a function of the state and co-state, i.e., using Eq. (6.11), u may be solved as a function of x and p. For this class of problems we saw earlier that the necessary conditions for optimality yield a *Hamiltonian* two-point boundary value problem. In the next section, we show how it can be solved using the theory presented in Chapter III.

6.3 Solving optimal control problems from the Hamilton-Jacobi theory

We saw earlier that if the feedback control law can be found as a function of (x, p) then the necessary conditions for optimality define a Hamiltonian system (Eqns. (6.14) and (6.15)). Together with the transversality conditions, the necessary conditions reduce to a two-point boundary value problem that can be solved using the theory we developed in Chapter III. In this section, we expose this novel approach to solving optimal control problems.

In the following, we make three assumptions.

1. The cost function J is smooth.

2. One can solve for u as a function of (x, p) using Eq. (6.11), that is, we can define a new Hamiltonian function $\bar{H}(x, p, t) = H(x, p, \bar{u}(x, p, t), t)$.

3. One can eliminate the λ_i's and λ_f's in Eq. (6.12), so that Eq. (6.12) becomes

$$p_k(t_f) = p_{f_k}\,,\ \forall k \in (r_i + 1, n)\,, \qquad p_k(t_0) = p_{0_k}\,,\ \forall k \in (r_f + 1, n)\,, \tag{6.60}$$

   and one can transform Eq. (6.6) into:

$$x_k(t_i) = x_{0_k}\,,\ k \in (1 \cdots r_i)\,, \qquad x_k(t_f) = x_{f_k}\,,\ k \in (1 \cdots r_f)\,. \tag{6.61}$$

Under these assumptions, solutions to the optimal control problem correspond to solutions (x, p) of the following conditions:

$$\dot{x} = \frac{\partial \bar{H}}{\partial p}(x, p, t)\,, \tag{6.62}$$

$$\dot{p} = -\frac{\partial \bar{H}}{\partial x}(x, p, t)\,, \tag{6.63}$$

with boundary conditions

$$\begin{aligned} x_k(t_0) &= x_{0_k} \quad \forall k \in (1, \cdots, r_i)\,, \\ p_k(t_0) &= p_{0_k} \quad \forall k \in (r_i+1, \cdots, n)\,, \\ x_k(t_f) &= x_{f_k} \quad \forall k \in (1, \cdots, r_f)\,, \\ p_k(t_f) &= p_{f_k} \quad \forall k \in (r_f+1, \cdots, n)\,. \end{aligned} \tag{6.64}$$

These equations define a two-point boundary value problem. Hence they are usually difficult to solve because they generally require an estimate of the initial (or final) state, which usually has no physical interpretation (we illustrated this point earlier in this chapter with the Heisenberg optimal control problem). However, treating the system defined by these equations as a Hamiltonian system allows us to apply the theory we developed in Chapter III. Define $I_{r_i} = \{1, \cdots, r_i\}$, $K_{r_f} = \{1, \cdots, r_f\}$ and recall the generating function $F_{I_{r_i},K_{r_f}}$ that verifies Eqns. (2.30), (2.31), (2.32) and (2.33):

$$p_{0_{I_{r_i}}} = -\frac{\partial F_{I_{r_i},K_{r_f}}}{\partial x_{0_{I_{r_i}}}}(x_{f_{K_{r_f}}}, p_{f_{\bar{K}_{r_f}}}, x_{0_{I_{r_i}}}, p_{0_{\bar{I}_{r_i}}}, t_f)\,, \tag{6.65}$$

$$x_{0_{\bar{I}_{r_i}}} = \frac{\partial F_{I_{r_i},K_{r_f}}}{\partial p_{0_{\bar{I}_{r_i}}}}(x_{f_{K_{r_f}}}, p_{f_{\bar{K}_{r_f}}}, x_{0_{I_{r_i}}}, p_{0_{\bar{I}_{r_i}}}, t_f)\,, \tag{6.66}$$

$$p_{f_{K_{r_f}}} = \frac{\partial F_{I_{r_i},K_{r_f}}}{\partial x_{K_{r_f}}}(x_{f_{K_{r_f}}}, p_{f_{\bar{K}_{r_f}}}, x_{0_{I_{r_i}}}, p_{0_{\bar{I}_{r_i}}}, t_f)\,, \tag{6.67}$$

$$x_{f_{\bar{K}_{r_f}}} = -\frac{\partial F_{I_{r_i},K_{r_f}}}{\partial p_{\bar{K}_{r_f}}}(x_{f_{K_{r_f}}}, p_{f_{\bar{K}_{r_f}}}, x_{0_{I_{r_i}}}, p_{0_{\bar{I}_{r_i}}}, t_f)\,. \tag{6.68}$$

These equations solve the above two-point boundary value problem and thus the necessary conditions for optimality.

Example VI.12 (Hard and soft constraint problems). Hard constraint problems [76] have their initial and terminal conditions entirely specified by the problem statement, i.e., x_0 and x_f are known ($r_i = n = r_f$). Thus, the F_1 generating function solves the necessary conditions for HCP:

$$p_f = \frac{\partial F_1}{\partial q_f}(x_f, x_0, t_f)\,, \tag{6.69}$$

$$p_0 = -\frac{\partial F_1}{\partial q_0}(x_f, x_0, t_f)\,. \tag{6.70}$$

On the other hand, soft constraint problems for which the initial state is fully determined, are solved using F_3. In fact, the transversality conditions provide us with p_f. Then, x_f and p_0 can be found from Eqns. (3.13) and (3.14):

$$x_f = -\frac{\partial F_3}{\partial p_f}(p_f, x_0, t_f)\,, \tag{6.71}$$

$$p_0 = -\frac{\partial F_3}{\partial x_0}(p_f, x_0, t_f)\,. \tag{6.72}$$

Using the Legendre transformation (Eq. (2.28)), we can transform generating functions into each other. *As a consequence, if the dynamics and the cost function remain unchanged (i.e., the Hamiltonian is the same), we are able to solve the optimal control problem for any boundary conditions at the cost of algebraic manipulations.* This fact is of importance because as different boundary conditions are applied to the system, the nature of the optimal feedback control laws change. This is a fundamental difficulty, and implies that the optimal control law for a given dynamical system must be re-solved as the boundary conditions and targets for the system change. Our approach directly tackles this issue, and allows us to overcome some of the barriers to truly reconfigurable control.

6.3.1 Linear quadratic terminal controller

Only rarely it is feasible to find the *explicit* feedback control laws for nonlinear systems. However, if an extremal path is known, it is usually possible to approximate the optimal control law in its neighborhood. For instance, at linear order this consists of solving an optimal control problem with quadratic performance criteria for time-varying linear systems. This class of problems has been widely studied and a detailed solution procedure is known. For that reason, it is of interest to first analyze the linear quadratic terminal controller as an introduction to our approach. This problem has been studied by Guibout and Scheeres [37] and Park and Scheeres [75].

Consider a linear dynamical system:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)\,,$$

and a quadratic cost function J:

$$J = \frac{1}{2}x(t_f)^T Q_f x(t_f) + \frac{1}{2}\int_{t_0}^{t_f} x^T Q x + u^T R u\,,$$

subject to r_i initial and r_f terminal conditions of the form:

$$x_k(t_0) = x_{0_k}\,,\ \forall k \in [1, r_i]\,, \quad x_k(t_f) = x_{f_k}\,,\ \forall k \in [1, r_f]\,,$$

where Q_f, Q and R are symmetric positive definite matrices. We define the Lagrangian L as $L = x^T Q x + u^T R u$ and the Hamiltonian function H as:

$$H(x, p, u) = p^T \dot{x} + L(x, u)\,.$$

From Eq. (6.11), we obtain

$$\bar{u} = -R^{-1}B^T p\,.$$

Substituting $\bar{u}$ in Eqns. (6.9) and (6.10) implies:

$$\begin{aligned} \bar{H}(x,p) &= H(x, p, -R^{-1}B^T p - R^{-1}N^T x) \\ &= \frac{1}{2}\begin{pmatrix} x \\ \lambda \end{pmatrix}^T \begin{pmatrix} Q & A^T \\ A & -BR^{-1}B^T \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix}\,. \end{aligned}$$

Then the necessary conditions for optimality yield $2n$ ordinary differential equations:

$$\dot{x} = Ax - BR^{-1}B^T p\,, \tag{6.73}$$

$$\dot{p} = -(A^T p + Qx)\,, \tag{6.74}$$

as well as $2n$ boundary conditions. To showcase our method, we consider the two following particular cases.

- HCP: $r_f = r_i = n$, i.e., the initial and final positions are specified. In that case the transversality conditions are void since we already know that:

  $$x(t_0) = x_0\,, \ \text{ and } \ x(t_f) = x_f\,. \tag{6.75}$$

  The necessary conditions can also be solved using F_1. From Eqns. (3.7) and (3.8), we directly find p_0 and p_f:

  $$\begin{cases} p &= \frac{\partial F_1}{\partial x}(x_f, x_0, t_f)\,, \\ p_0 &= \frac{\partial F_1}{\partial x_0}(x_f, x_0, t_f)\,. \end{cases} \tag{6.76}$$

- SCP: Another case that is often treated (see e.g. Bryson [19] and Park and Scheeres [76]), is the soft constraint problem, where only the initial state is given. In that case, the transversality conditions yield:

  $$\begin{aligned} x(t_0) &= x_0\,, \\ p(t_f) &= Q_f x(t_f)\,. \end{aligned} \tag{6.77}$$

  Eqns. (6.73), (6.74) and (6.77) define a linear Hamiltonian two-point boundary value problem. However, it is not defined as usual boundary value problems since the final conditions are expressed as constraints on the state and co-state. This is due to the explicit dependence in the final state of the cost function. If $Q_f = 0$, the necessary conditions for optimality reduce to a classical Hamiltonian two-point boundary value problem and we recover the results found in Ex. VI.12. The generating function can also solve the soft constraint problem. From Lemma III.3, F_1 can be written in the following form:

  $$F_1 = \frac{1}{2} Y^T \begin{pmatrix} F^1_{11}(t) & F^1_{12}(t) \\ F^1_{21}(t) & F^1_{22}(t) \end{pmatrix} Y\,, \tag{6.78}$$

where $Y = (x, x_0)^T$. Then, using Eqns. (3.7) and (3.8), we obtain:

$$\left\{\begin{array}{rcl} p &=& F^1_{11}(t)x + F^1_{12}(t)x_0\,, \\ p_0 &=& -F^1_{21}(t)x - F^1_{22}(t)x_0\,, \\ p_f &=& Q_f x_f\,. \end{array}\right.$$

Solving for p_0 yields:

$$p_0 = -\left[F^1_{21}(t)(Q_f - F^1_{11}(t))^{-1}F^1_{12}(t)\right]\,.$$

This method readily applies to other kinds of boundary conditions. Thus, using the generating functions, we are able to solve the necessary conditions for the linear quadratic terminal controller. This is not surprising since there exist methods to solve this problem based on the state transition matrix, and we showed that state transition matrix and generating functions are closely related.

6.3.2 Targeting problem

To illustrate the use of the generating functions to solve nonlinear optimal control problems we now consider a targeting problem in the two-dimensional Hill three-body problem (Appendix C). We consider a spacecraft away from the Libration point L_2 and want to find the control sequence that moves the spacecraft at the equilibrium point L_2 while minimizing the fuel consumption. Specifically, this optimal control problem formulates as follows:

We want to minimize the cost function $J = \frac{1}{2}\int_{t=0}^{t=t_f}(u_x^2 + u_y^2)dt$ subject to the constraints

$$\left\{\begin{array}{rcl} \dot{x_1} &=& x_3\,, \\ \dot{x_2} &=& x_4\,, \\ \dot{x_3} &=& 2x_4 - \frac{x_1}{(x_1^2+x_2^2)^{3/2}} + 3x_1 + u_x\,, \\ \dot{x_4} &=& -2x_3 - \frac{x_2}{(x_1^2+x_2^2)^{3/2}} + u_y\,, \end{array}\right. \tag{6.79}$$

and the boundary conditions:

$$X(t=0) = X_0\,,\; X(t=t_f) = X_{L_2} = (3^{-1/3},0,0,0)\,, \tag{6.80}$$

where $X = (x_1, x_2, x_3, x_4) = (x, y, \dot{x}, \dot{y})$. Define the Hamiltonian:

$$\begin{aligned} H(X,P,U) = p_1 x_3 + p_2 x_4 + p_3 \left(2x_4 - \frac{x_1}{(x_1^2+x_2^2)^{3/2}} + 3x_1 + u_x\right) \\ + p_4 \left(-2x_3 - \frac{x_2}{(x_1^2+x_2^2)^{3/2}} + u_y\right) + \frac{1}{2}u_x^2 + \frac{1}{2}u_y^2, \end{aligned}$$

where $P = (p_1, p_2, p_3, p_4)$ and $U = (u_x, u_y)$. Then, from $\frac{\partial H}{\partial U} = 0$, we find the optimal control feedback law:

$$u_x = -p_3\,,\; u_y = -p_4\,.$$

Substituting $U = (u_x, u_y)$ into H yields:

$$\begin{aligned} \bar{H}(X,P) = p_1 x_1 + p_2 x_2 + p_3 \left(2x_4 - \frac{x_1}{(x_1^2+x^2)^{3/2}} + 3x_1 - p_3\right) \\ + p_4 \left(-2x_3 - \frac{x_2}{(x_1^2+x^2)^{3/2}} - p_4\right) + \frac{1}{2}p_3^2 + \frac{1}{2}p_4^2\,. \end{aligned} \tag{6.81}$$

We deduce the necessary conditions for optimality:

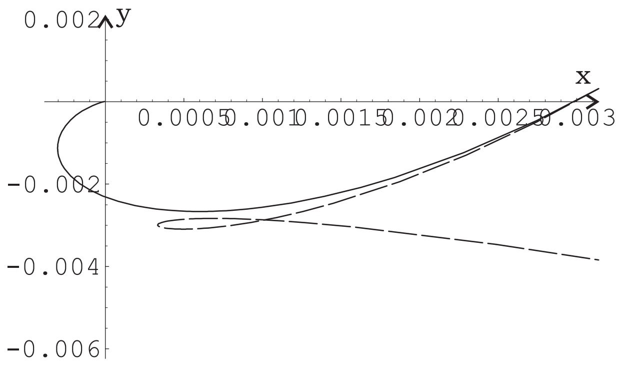
$$\dot{X} = \frac{\partial \bar{H}}{\partial P}\,, \tag{6.82}$$

$$\dot{P} = -\frac{\partial \bar{H}}{\partial X}\,, \tag{6.83}$$

$$X(t=0) = X_0\,, \qquad X(t=t_f) = (3^{-1/3},0,0,0)\,.$$

This is a Hamiltonian two-point boundary value problem that we can solve using the theory developed in Chapter III once the generating functions are known. To compute the generating functions, we can use the algorithm developed in Chapter V by noticing that the solution to the optimal control problem that consists of going from L_2 to L_2 in t_f units of time is the trivial trajectory $X = X_{L_2}$ and $U = (0,0)$ for all t. This trajectory can be taken to be the reference trajectory in the algorithm. In the following we use an approximation of order 4 of F_1.

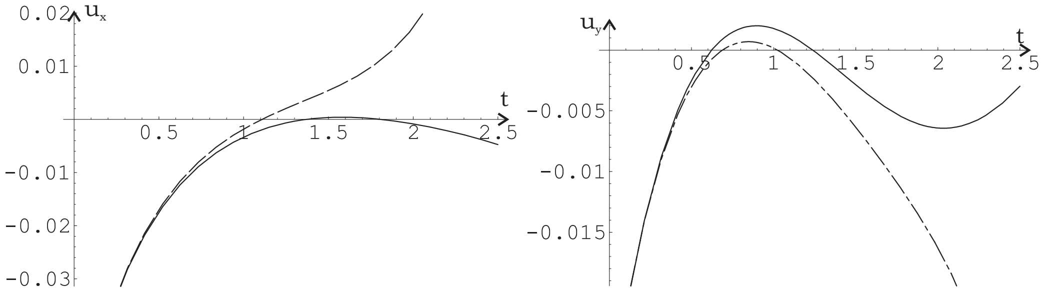
In Fig. 6.1, we plot the optimal trajectory that starts at $X_0 = (10, 700, 10, 700)$ km and reaches L_2 in $t_f = 145$ days. The dotted line corresponds to the solution found using a linear approximation of the dynamics whereas the solid line is the solution computed with an order 4 approximation of the dynamics. We immediately notice that the linear approximation fails to predict a relevant approximation of the control since the trajectory does not reach the Libration point (nor its vicinity). On the other hand, F_1 provides an excellent approximation of the control since the spacecraft is 13 km away from L_2 at t_f. Fig. 6.2 shows the associated control sequence. The dotted line and solid line correspond to the control history computed from the linear model and the fourth order approximated system.



0.01 unit of length $\longleftrightarrow 21,500\ km$

Figure 6.1: Optimal trajectory of the spacecraft

Furthermore, it should be clear that once the F_1 generating function is known, we can *instantaneously* solve this hard constraint problem for any boundary conditions and any final times. In Fig. 6.3 we illustrate this key point by plotting the trajectories for different final times. As t_f increases, the trajectory tends to wrap around the Libration point so that the spacecraft takes advantage of the geometry of the Libration point (Appendix C). On



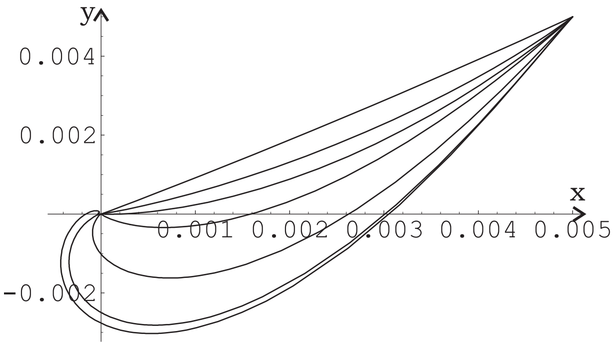
(a) Time history of u_x (b) Time history of u_y

$$\begin{array}{rcl} 1 \text{ unit of time} & \longleftrightarrow & 58\, days\,, \\ 10^{-3} \text{ unit of control} & \longleftrightarrow & 1.36 \cdot 10^{-2} m.s^{-2} \end{array}$$

Figure 6.2: Time history of the control laws

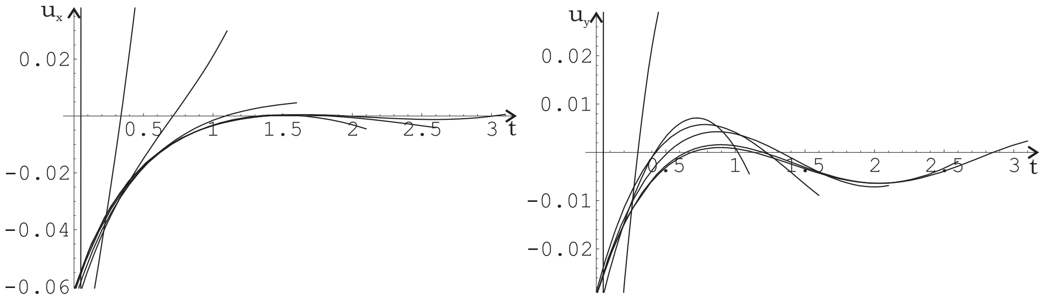
the other hand, if the transfer time is small, the trajectory is almost a straight line and it completely ignores the dynamics. In Fig. 6.4, the associated control laws are represented. As expected, the longer the transfer time is, the smaller the magnitude of the control is. We emphasize that we only need to evaluate the gradient of F_1 (which is a polynomial of order 3) seven times and integrate Eqns. (6.82) and (6.82) seven times to obtain the seven curves in Fig. 6.3. Similarly, in Fig. 6.5, at the cost of sixteen evaluations of the gradient of F_1, we are able to represent the optimal trajectories of spacecraft starting at $X_0 = (r\cos(\theta), r\sin(\theta))$ where $r = 10,700$ km and $\theta = k\pi/8$, and ending at L_2 in 145 days. In Fig. 6.6 the corresponding optimal control laws are represented.

Finally, if different types of boundary conditions are imposed (for instance, the terminal state is free) then we need to perform a Legendre transform to find the generating function that solves this new boundary value problem. We recall that for these types of problems, the Legendre transform is found at the cost of a $n \times n$ matrix inversion.



0.01 unit of length $\longleftrightarrow$ $21,500\ km$

Figure 6.3: Optimal trajectories of the spacecraft for different transfer times.

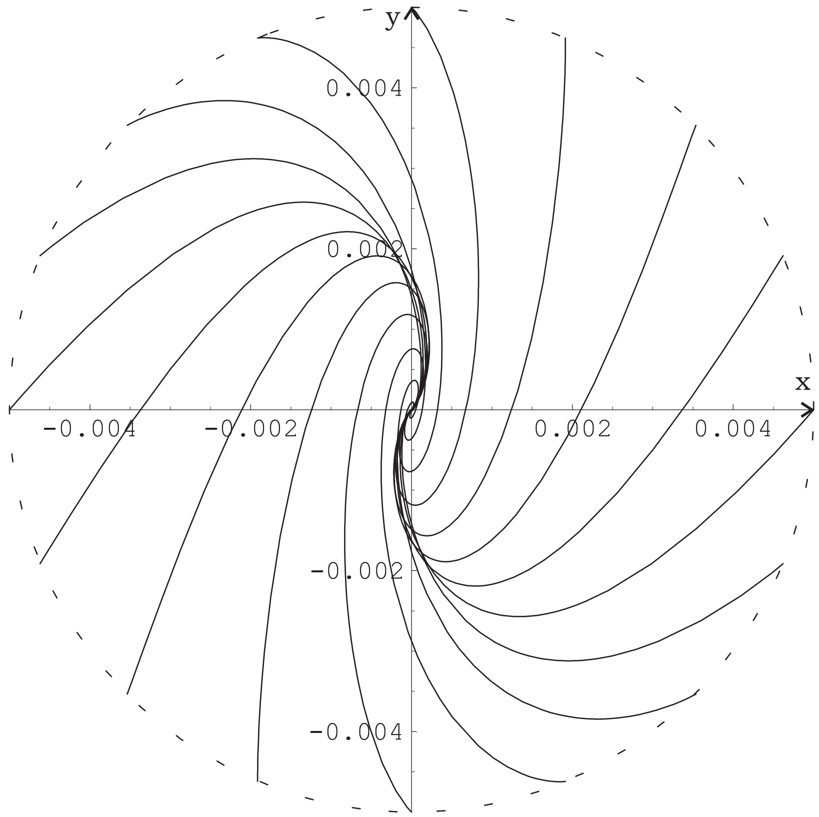


(a) Time history of u_x

(b) Time history of u_y

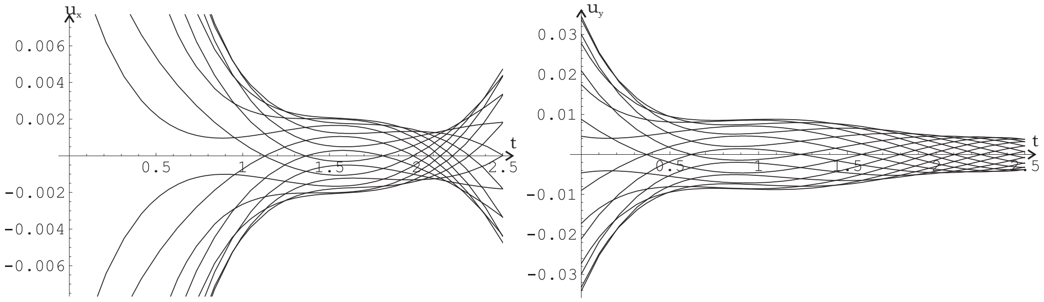
$$\begin{array}{rcl} 1 \text{ unit of time} & \longleftrightarrow & 58\ days\,, \\ 10^{-3} \text{ unit of control} & \longleftrightarrow & 1.36 \cdot 10^{-2} m.s^{-2} \end{array}$$

Figure 6.4: Time history of the control laws



0.01 unit of length $\longleftrightarrow$ $21,500\ km$

Figure 6.5: Optimal trajectories of the spacecraft as a function of the initial position.



(a) Time history of u_x

(b) Time history of u_y

$$\begin{array}{rcl} 1 \text{ unit of time} & \longleftrightarrow & 58\ days\,, \\ 10^{-3} \text{ unit of control} & \longleftrightarrow & 1.36 \cdot 10^{-2} m.s^{-2} \end{array}$$

Figure 6.6: Time history of the control laws

CHAPTER VII

THE SEARCH FOR PERIODIC ORBITS

Periodic orbits have been widely studied over the last century and are still a topic of great interest. Poincaré [77] already realized their importance for understanding the dynamics of non-integrable Hamiltonian systems when he claimed that they are “the only opening through which we can try to penetrate the stronghold”. Indeed, he conjectured that periodic orbits are dense on typical energy surfaces. Though the Poincaré conjecture is not true for every system (e.g., for a product of harmonic oscillator with incommensurate frequencies), many systems have the property predicted by Poincaré. MacKay [63] provides conditions under which the Poincaré conjecture holds.

Many techniques, that we will not attempt to survey in any systematic fashion, have been developed to find periodic orbits. For instance, in the restricted three-body problem one may use perturbation methods (see e.g. Hénon [48]). Such a method allows one to find families of periodic orbits very efficiently once a member of the family is known, but does not provide a systematic procedure to find a periodic orbit of either a given period or going through a given point. By using the theory we developed in Chapter III we can solve such a problem. Specifically, we are able to reduce the search for periodic orbits to either finding the solution to a set of implicit equations, which can often be done graphically, or to finding the roots of an equation of one variable only. Most importantly, the novel approach we develop applies to any Hamiltonian system and therefore is very general. We illustrate its use with two non-trivial examples of finding periodic orbits in the vicinity of other periodic orbits and around the Libration points in the three-body problem.

7.1 Periodic orbits and generating functions

The aim of this section is to transform the search for periodic orbits into a two-point boundary value problem that can be handled with the theory developed in Chapter III.

Periodic orbits in a $2n$-dimensional Hamiltonian dynamical system are characterized by the following equations:

$$q(T) = q_0\,, \tag{7.1}$$

$$p(T) = p_0\,, \tag{7.2}$$

where T is the period of the orbit, (q_0, p_0) are the initial conditions at time $t_0 = 0$ and $(q(t), p(t))$ verifies Hamilton's equations:

$$\dot{q} = \frac{\partial H}{\partial p}(q, p, t)\,,\ \dot{p} = -\frac{\partial H}{\partial q}(q, p, t)\,. \tag{7.3}$$

In the most general case, the search for periodic orbits consists of solving the $2n$ equations (7.1) and (7.2) for the $2n + 1$ unknowns (q_0, p_0, T). Simple methods that solve this problem take a set of initial conditions (q_0, p_0), and integrate Hamilton's equations. If there exists a time $t = T$ such that Eqns. (7.1) and (7.2) are verified, then a periodic orbit is found. Else, other initial conditions need to be guessed. In the approach we propose in this chapter, instead of looking at the initial conditions and the period as the only variables of the problem, we suppose that the period, n initial conditions as well as n components of the state vector at time T are unknowns. Then the search for periodic orbits reduces to solving the $2n$ equations (7.1) - (7.2) for these $2n + 1$ unknowns.

Example VII.1. If $(q(T), q_0, T)$ are taken to be the $2n+1$ unknowns, then the search for periodic orbits consists of solving the $2n$ equations (7.1)-(7.2) for $(q(T), q_0, T)$. Let us now find all periodic orbits of a given period. In other words, T is given and we need to find $(q(T), q_0)$ such that $q(T) = q_0$ and $p(T) = p(0)$. This is a two-point boundary value problem with constraints that can be solved with the generating function F_1. Combining Eqns. (3.7)-(3.8) and Eqns. (7.1)-(7.2) yields:

$$\begin{array}{rclcrcl} p(T) & = & \frac{\partial F_1}{\partial q}(q, q_0, T)\,, & \quad & q(T) & = & q_0\,, \\ p_0 & = & -\frac{\partial F_1}{\partial q_0}(q, q_0, T)\,, & \quad & p(T) & = & p_0\,, \end{array} \tag{7.4}$$

that is:

$$\frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T) = 0\,, \tag{7.5}$$

$$p = p_0 = \frac{\partial F_1}{\partial q}(q = q_0, q_0, T)\,. \tag{7.6}$$

Hence, the search for all periodic orbits of a given period is reduced to solving n equations (7.5) for n variables, the q_0's, and then evaluate n equations (7.6) to compute the corresponding momenta. $2n$ equations still need to be solved, but now n of them are decoupled. Most importantly, once F_1 is known, no additional integration is required. In addition, using the algorithm we developed in Chapter V, solutions of Eq. (7.5) correspond to roots of polynomials and are therefore easily computed.

Similarly, by taking $(p(T), p_0, T)$ as unknowns we can use the F_4 generating function to derive necessary and sufficient conditions. In that case we obtain:

$$\frac{\partial F_4}{\partial p}(p = p_0, p_0, T) + \frac{\partial F_4}{\partial p_0}(p = p_0, p_0, T) = 0\,, \quad q = q_0 = -\frac{\partial F_4}{\partial p}(p = p_0, p_0, T)\,. \tag{7.7}$$

However, there is a difference between these two approaches. Using F_1 we solve the necessary and sufficient conditions defined by Eq. (7.5) in the configuration space (q_0 is the unknown) whereas F_4 yields an equation whose variables are in the momentum space.

Although this difference does not have any importance if one searches for all periodic orbits of a given period, it is crucial if some constraints are imposed on the domain in the phase space in which we search for periodic orbits. For instance, if one wants to find all periodic orbits of period T crossing an axis defined by all but one component of q_0 being non-zero, then one should solve Eq. (7.5) for the only non-zero component of q_0. If one uses F_4 instead, then one needs to solve Eq. (7.7) for the n components of p_0, and then check afterwards which solutions satisfy the constraint.

Finally, we point out that other choices of unknowns may yield more complex necessary and sufficient conditions. Suppose we consider that $(q(T), p_0, T)$ are unknowns, then F_2 must be used. We have:

$$\begin{aligned} p(T) &= \tfrac{\partial F_2}{\partial q}(q, p_0, T)\,, \qquad & q(T) &= q_0\,, \\ q_0 &= \tfrac{\partial F_2}{\partial p_0}(q, p_0, T)\,, \qquad & p(T) &= p_0\,. \end{aligned} \tag{7.8}$$

These equations cannot be decoupled. As a result, we must solve $2n$ coupled equations for $q(T)$ and p_0 to find periodic orbits:

$$p_0 = \frac{\partial F_2}{\partial q}(q(T), p_0, T)\,, \quad q(T) = \frac{\partial F_2}{\partial p_0}(q(T), p_0, T)\,. \tag{7.9}$$

Through this example we have discussed a novel application of the Hamilton-Jacobi theory to find periodic orbits. By considering the period, n initial conditions and n components of the state vector at T as unknowns, we reduced the search for periodic orbits to solving two-point boundary value problems with constraints. Using the generating functions, these boundary value problems simplify into a set of necessary and sufficient conditions that characterize periodic solutions. Furthermore, we proved that there are choices of unknowns that give simpler sets of conditions. Most importantly, once the generating functions are known, solutions of these necessary and sufficient conditions are computed using algebraic manipulations only, no integration is required. In particular, using the algorithm developed in Chapter V, solutions of the necessary and sufficient conditions correspond to roots of polynomials and are therefore easily found.

We now generalize our approach and study its properties. We noticed in the above example that there are choices of unknowns that give simpler sets of necessary and sufficient conditions. For instance, the use of $(q(T), q_0, T)$ and $(p(T), p_0, T)$ allowed us to derive $2n$ necessary and sufficient conditions for periodicity, among which n were decoupled. On the other hand, the use of $(q(T), p_0, T)$ yielded $2n$ *coupled* equations, the reason being that we were unable to simplify the $2n$ equations (3.10) and (3.11). In the general case, an arbitrary generating function F_{I_p,K_r} verifies:

$$\begin{aligned}
p_{I_p} &= \frac{\partial F_{I_p,K_r}}{\partial q_{I_p}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, T)\,, \\
q_{\bar{I}_p} &= -\frac{\partial F_{I_p,K_r}}{\partial p_{\bar{I}_p}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, T)\,, \\
p_{0_{K_r}} &= -\frac{\partial F_{I_p,K_r}}{\partial q_{0_{K_r}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, T)\,, \\
q_{0_{\bar{K}_r}} &= \frac{\partial F_{I_p,K_r}}{\partial p_{0_{\bar{K}_r}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, T)\,.
\end{aligned} \tag{7.10}$$

Assuming $q(T) = q_0$ and $p(T) = p_0$, these equations can be decoupled if and only if $I_p \bigcap K_r \neq \emptyset$. If $dim\,(I_p \bigcap K_r) = m$, then m equations can be decoupled. The case where $m = n$ is optimal as it yields n coupled equations and n decoupled equations. In the following, we restrict ourselves to the case where $m = n$, i.e., $I_p = K_r$, and thus only consider unknowns of the form $(q_{I_p}, p_{\bar{I}_p}, q_{0_{I_P}}, p_{0_{\bar{I}_p}}, T)$.

Let $(i_1, \cdots, i_p)\,(i_{p+1}, \cdots, i_n)$ be a partition of the set $(1, \cdots, n)$ into two non-intersecting parts such that $i_1 < \cdots < i_p$, $i_{p+1} < \cdots < i_n$, and define $I_p = (i_1, \cdots, i_p)$. Let us solve the problem that consists of finding $(q_{I_p}, p_{\bar{I}_p}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)$ such that Eqns. (7.1)-(7.2) are satisfied. This is a two-point boundary value problem with an unknown transfer time T, and constraints defined by Eqns. (7.1)-(7.2). Solutions $(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)$ of this problem are, by definition, periodic orbits of period T that go through the point whose coordinates are partially given by $(q_{0_{I_p}}, p_{0_{\bar{I}_p}})$ at the initial time

$t = 0$ and at $t = T$. For instance, if $p = n$, we recover the above example in which we find all the periodic orbits of period T going though the point q_0 at $t = 0$ and $t = T$.

To find the set of solutions $(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)$ satisfying Eqns. (7.1)-(7.2) we use the generating function F_{I_p,I_p}. Recall the equations satisfied by F_{I_p,I_p}:

$$\begin{aligned}
p_{I_p} &= \frac{\partial F_{I_p,I_p}}{\partial q_{I_p}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)\,,\\
q_{\bar{I}_p} &= -\frac{\partial F_{I_p,I_p}}{\partial p_{\bar{I}_p}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)\,,\\
p_{0_{I_p}} &= -\frac{\partial F_{I_p,I_p}}{\partial q_{0_{I_p}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)\,,\\
q_{0_{\bar{I}_p}} &= \frac{\partial F_{I_p,I_p}}{\partial p_{0_{\bar{I}_p}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)\,.
\end{aligned} \tag{7.11}$$

Combining Eq. (7.11) together with Eqns. (7.1)-(7.2) yields the $2n$ following equations:

$$\frac{\partial F_{I_p,I_p}}{\partial q_{I_p}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) + \frac{\partial F_{I_p,I_p}}{\partial p_{\bar{I}_p}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) = 0\,, \tag{7.12}$$

$$\frac{\partial F_{I_p,I_p}}{\partial q_{0_{I_p}}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) + \frac{\partial F_{I_p,I_p}}{\partial p_{0_{\bar{I}_p}}}(q_{I_p} = q_{0_{I_p}} p_{\bar{I}_p} =, p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) = 0\,, \tag{7.13}$$

$$-\frac{\partial F_{I_p,I_p}}{\partial q_{0_{I_p}}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) = p_{0_{I_p}}\,, \tag{7.14}$$

$$\frac{\partial F_{I_p,I_p}}{\partial p_{0_{\bar{I}_p}}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) = q_{0_{\bar{I}_p}}\,. \tag{7.15}$$

These $2n$ equations are composed of n coupled equations (Eqns. (7.12) and (7.13)) and n decoupled equations (Eqns. (7.14) and (7.15)). Their solutions (q_0, p_0, T) fully determine periodic orbits. In addition, if we use the algorithm presented in Chapter V to approximate the generating functions as polynomials, then the periodic orbits are solutions of $2n$ polynomial equations.

However, these equations cannot be solved as they are because there are $2n + 1$ unknowns, (q_0, p_0, T), and only $2n$ equations. Thus, in general, one needs to take at least one variable as a known parameter. In the above example, we decided to set the period T, but other choices could have been possible. In this chapter, we focus on two choices which are of particular interest in astrodynamics.

1. *Searching in the time domain*: Suppose we are looking at all periodic orbits going through a point in the configuration space that corresponds to the position of a spacecraft. The position q_0 is fixed but the associated momenta and the period are unknowns. This problem requires us to solve $2n$ equations for $n+1$ variables only. Now the problem is over-determined. In the following we show that it can be reduced to solving a single equation for the period T, followed by n function evaluations to find the other n variables. As a result the only variable that is not trivially found is the period T. This motivates our choice to call this class of problem "Searching in the time domain".

2. *Searching in phase space*: The second type of problem we consider corresponds to the one we discussed in the example Ex. VII.1. We set the period and look at all periodic orbits of that period in the phase space. This corresponds to a search for periodic orbits in the *phase space*.

Searching in the time domain

We assume knowledge of n components of a point in the phase space, say $(q_{0_{I_p}}, p_{0_{\bar{I}_p}})$, and search for all periodic orbits going through that point. Recall the conditions (Eqns. (7.12) and (7.13)) derived using F_{I_p,I_p}. Since the coordinates $(q_{0_{I_p}}, p_{0_{\bar{I}_p}})$ are known, these equations are functions of the period T only:

$$\frac{\partial F_{I_p,I_p}}{\partial q_{I_p}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) \\ + \frac{\partial F_{I_p,I_p}}{\partial p_{\bar{I}_p}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) = 0\,,$$

$$\begin{aligned}\frac{\partial F_{I_p,I_p}}{\partial q_{0_{I_p}}}&(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)\\ &\qquad + \frac{\partial F_{I_p,I_p}}{\partial p_{0_{\bar{I}_p}}}(q_{I_p} = q_{0_{I_p}} p_{\bar{I}_p} =, p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) = 0\,.\end{aligned}$$

Solutions of these equations, T, correspond to the periods of periodic orbits that go through $(q_{0_{I_p}}, p_{0_{\bar{I}_p}})$. Instead of solving the n equations for 1 variable we may combine them in the following way:

$$\left\| \left(\frac{\partial F_{I_p,I_p}}{\partial q_{I_p}}(\alpha) + \frac{\partial F_{I_p,I_p}}{\partial p_{\bar{I}_p}}(\alpha), \frac{\partial F_{I_p,I_p}}{\partial q_{0_{I_p}}}(\alpha) + \frac{\partial F_{I_p,I_p}}{\partial p_{0_{\bar{I}_p}}}(\alpha) \right) \right\| = 0\,, \tag{7.16}$$

where $\alpha = (q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T)$ and $\|\|$ is a norm. This equation can be easily solved numerically or even graphically. Finally, to recover the n remaining unknown coordinates, we only need to evaluate the n equations (7.14) and (7.15):

$$\begin{aligned} -\tfrac{\partial F_{I_p,I_p}}{\partial q_{0_{I_p}}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) &= p_{0_{I_p}}\,,\\ \tfrac{\partial F_{I_p,I_p}}{\partial p_{0_{\bar{I}_p}}}(q_{I_p} = q_{0_{I_p}}, p_{\bar{I}_p} = p_{0_{\bar{I}_p}}, q_{0_{I_p}}, p_{0_{\bar{I}_p}}, T) &= q_{0_{\bar{I}_p}}\,. \end{aligned} \tag{7.17}$$

Example VII.2. Suppose $p = n$, so that $F_{I_p,I_p} = F_1$. Then Eq. (7.16) simplifies to:

$$\|\frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T)\| = 0\,. \tag{7.18}$$

Eq. (7.18) is a single equation of one variable that can be solved graphically. To find the corresponding momentum, we can use Eq. (7.6):

$$p(T) = p_0 = \frac{\partial F_1}{\partial q}(q = q_0, q_0, T)\,, \tag{7.19}$$

To conclude, using the Hamilton-Jacobi theory we are able to characterize periodic orbits going through a point in the phase space partially specified by n of its $2n$ coordinates. Solutions to the obtained equations are easily found once the generating functions are known. Indeed, it suffices to solve an equation of one variable to find the periods of the orbits, T. Then the n remaining coordinates are found at the cost of n function evaluations.

Searching in phase space

We now search for all periodic orbits of a given period T. This problem is well-posed since we now have $2n$ equations and $2n$ unknowns. A priori, there are no imposed choices for generating functions to solve this problem. Any generating function of the form F_{I_p,I_p} may be used equivalently. Indeed, Eqns. (7.12)-(7.14) define $2n$ equations of $2n$ variables, n of the equations being decoupled. The difference between conditions derived using different generating functions is mainly the space on which the n coupled equations need to be solved. For instance, F_1 yields conditions whose variables lie in the configuration space whereas F_4 yields conditions whose variables lie in the momentum space. If domain constraints are imposed then some particular choice of generating function may be more appropriate. For instance, if one looks for periodic orbits in the vicinity of an equilibrium point, one should use the F_1 generating function.

To conclude, the approach we propose has many advantages compared to other methods developed in the literature. First, there is no need to integrate the equations of motion once the generating functions are known. Only a system of at most n equations (where $2n$ is the dimension of the phase space) need to be solved. Second, we do not need any initial guess to initialize our algorithm. Finally, and most importantly, our approach is very general and applies to any Hamiltonian system, independent of its complexity.

Let us now illustrate the theory developed above with some examples. First we analyze the necessary and sufficient conditions obtained for linear systems, and then look at some more sophisticated problems such as periodic orbits in the Hill three-body problem and in the restricted three-body problem.

7.2 Linear systems analysis

In this section we focus on finding periodic orbits of linear systems using the generating functions. In particular, we simplify the set of equations (7.12)-(7.14) that characterize periodic orbits. For sake of simplicity, and without loss of generality, we only focus on the F_1 generating function, the content of this section can readily be transported to the other generating functions.

Consider a linear Hamiltonian system with quadratic Hamiltonian function:

$$H(q,p,t) = \frac{1}{2} X^T \begin{pmatrix} H_{qq}(t) & H_{qp}(t) \\ H_{pq}(t) & H_{pp}(t) \end{pmatrix} X\,, \tag{7.20}$$

where $H_{qp} = H_{pq}^T$, H_{qq} and H_{pp} are symmetric and $X = \begin{pmatrix} q \\ p \end{pmatrix}$. If one studies the relative motion of two particles, then $X = X^h = \begin{pmatrix} \Delta q \\ \Delta p \end{pmatrix}$ as previously defined.

For linear systems, the generating function F_1 is also quadratic in its spatial variables without any linear term (Lemma III.3), i.e,

$$F_1(Y,t) = \frac{1}{2} Y^T \begin{pmatrix} F^1_{11}(t) & F^1_{12}(t) \\ F^1_{21}(t) & F^1_{22}(t) \end{pmatrix} Y\,, \tag{7.21}$$

where $F^1_{12} = {F^1_{21}}^T$, F^1_{11} and F^1_{22} are symmetric and $Y = \begin{pmatrix} q \\ q_0 \end{pmatrix}$. Then, Eqns. (7.12)-(7.15) transform into:

$$\begin{aligned} \left[F^1_{11}(T) + F^1_{21}(T) + F^1_{12}(T) + F^1_{22}(T)\right] q_0 &= 0\,, \\ \left[F^1_{21}(T) + F^1_{22}(T)\right] q_0 &= p_0\,. \end{aligned} \tag{7.22}$$

Eq. (7.22) defines two n-dimensional matrix equations that are functions of $2n+1$ variables. As we mentioned before, we need to take at least one variable as a known parameter.

Example VII.3 (Periodic orbits about the Libration point L_2 in the Hill three-body problem). Let us find all the periodic orbits going through a given point q_0 using the linearized equations of the dynamics about the Libration point L_2 in the normalized Hill three-body problem (Appendix C).

We first need to solve the n coupled equations for the time period:

$$\left[F^1_{11}(T) + F^1_{21}(T) + F^1_{12}(T) + F^1_{22}(T)\right] q_0 = 0\,. \tag{7.23}$$

From linear algebra theory, a necessary condition for this equation to have a solution (assuming $q_0 \neq 0$) is that:

$$\det\left[F^1_{11}(T) + F^1_{21}(T) + F^1_{12}(T) + F^1_{22}(T)\right] = 0\,. \tag{7.24}$$

Fig. 7.1 represents this determinant. We notice that there exists one time at which the determinant vanishes. Using Newton iteration we find that it vanishes at[1] $T = 3.0330191$ and that the rank of the matrix $F^1_{11}(T)+F^1_{21}(T)+F^1_{12}(T)+F^1_{22}(T)$ at this T is 0. Therefore, any point q_0 in the configuration space belongs to a periodic orbit of period T. These

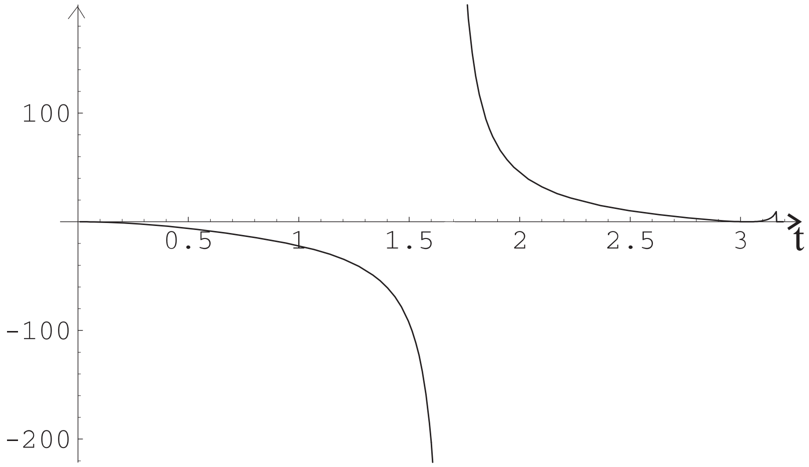


Figure 7.1: Determinant of the matrix defined in Eq. (7.24)

[1] Higher accuracy may be obtained using a smaller time step when solving for F_1

results are in agreement with known results on periodic orbits about the Libration points in the linearized system. Using linear systems theory, we find that the true period of oscillatory motion about L_2 is 3.0330193236451116.

7.3 Nonlinear systems

In this section, we illustrate the power of the proposed method to find periodic orbits of nonlinear systems. We address two non-trivial examples. First, we study periodic orbits about the Libration point L_2 in the normalized Hill three-body problem. Then, we search for periodic orbits in the vicinity of a periodic orbit in the normalized restricted three-body problem.

Study of periodic orbits about L_2

In order to apply our method to finding periodic orbits we need to compute the generating functions. Using the algorithm presented in Chapter V we are able to find a polynomial approximation of the generating function F_1 up to order 5, that is, we use an approximation of order 5 for the dynamics. For the present study, such an approximation is sufficient since we found in Section 5.4.3 that the domain of use is (5.14):

$$D = \{0.05, (0.01, 1.32) \cup (1.84, 3.12) \cup (3.12, 3.5)\}\,.$$

Within the domain of use, we observed that the error is $3.5 \cdot 10^{-5}$ (about $77\ km$ for the Earth-Sun system).

We consider the following two problems:

1. *Searching in the time domain*: Find all periodic orbits going through $q_0 = (0.01, 0)$. To solve such a problem, we use Eq. (7.16):

$$\|\frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T)\| = 0\,. \tag{7.25}$$

In Fig. 7.2 we plot the left-hand side of Eq. (7.25) as a function of the normalized time. We observe that the norm vanishes only at $t = T = 3.03353$. Therefore,

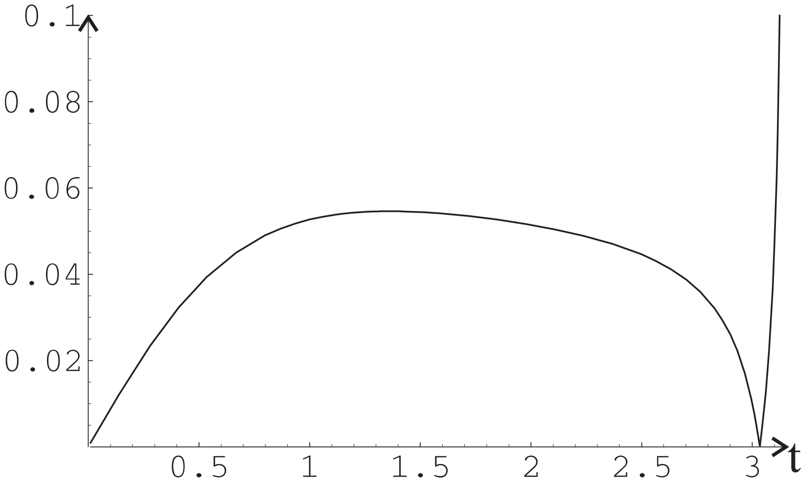
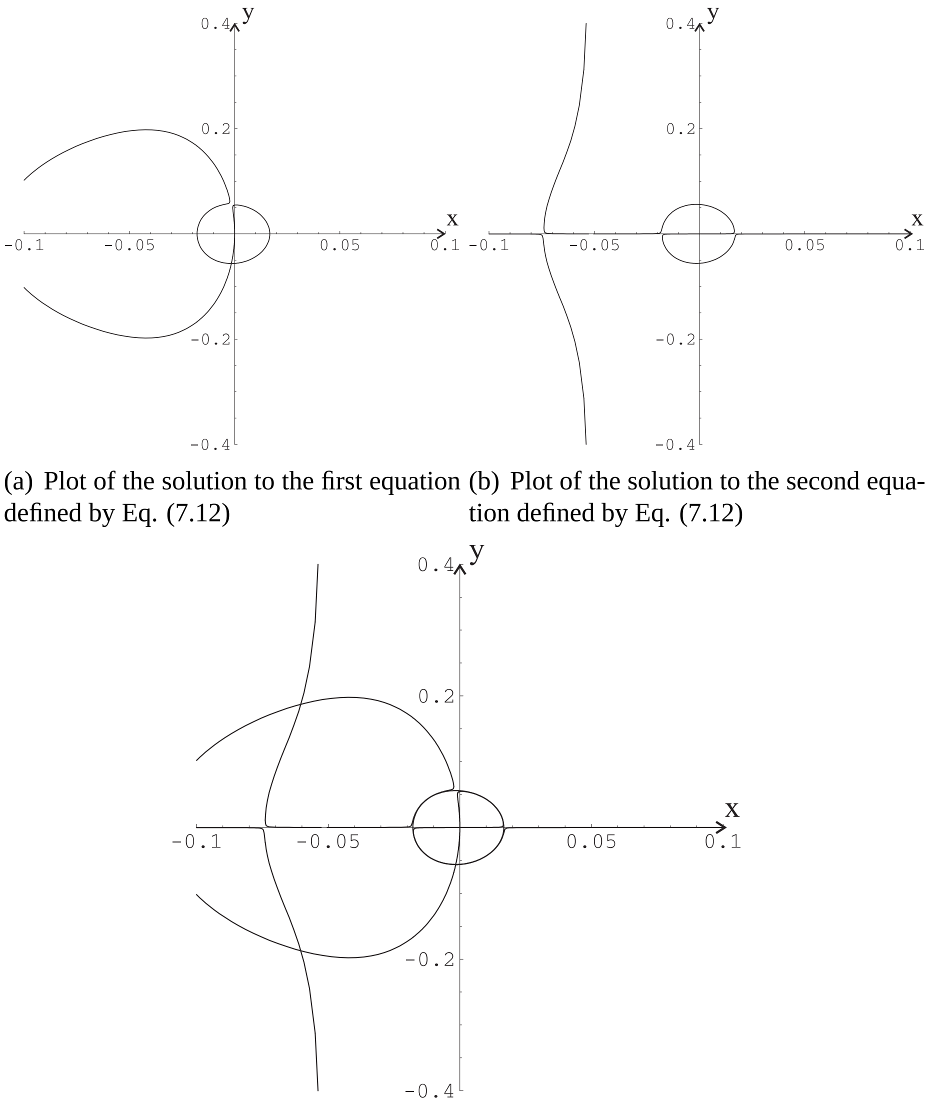


Figure 7.2: Plot of $\| \frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T) \|$ where $q_0 = (0.01, 0)$

there exists only one periodic orbit going through q_0, and its period is T (there may be additional periodic orbits of period $T > 3.2$, but we cannot see them in this figure). Again, these results are in agreement with known results on periodic orbits about L_2. One can show that any point in the vicinity of L_2 belongs to a periodic orbit. The periods of these orbits increase as their distances from L_2 increase. In the limit, as the distance between periodic orbits and L_2 goes to zero, the period goes to $T = T_{linear} = 3.0330193236451116$.

2. *Searching in position space*: Another problem is to find all periodic orbits of a given period $T = 3.0345$. To solve this problem we use Eq. (7.12) which defines two equations with two unknowns that can be solved graphically. In Fig. 7.3, we plot the solutions to each of these two equations and then superimpose them to find their intersection. The intersection corresponds to solutions of Eq. (7.12), that is, to the set of points that belongs to periodic orbits of period T. We observe

that the intersection is composed of a circle and two points whose coordinates are $(q_x, q_y) = (-0.0603795, \pm 0.187281)$. The circle is obviously a periodic orbit but the two points are not equilibrium points, and rather correspond to out-of-plane periodic orbits[2].



(a) Plot of the solution to the first equation defined by Eq. (7.12)

(b) Plot of the solution to the second equation defined by Eq. (7.12)

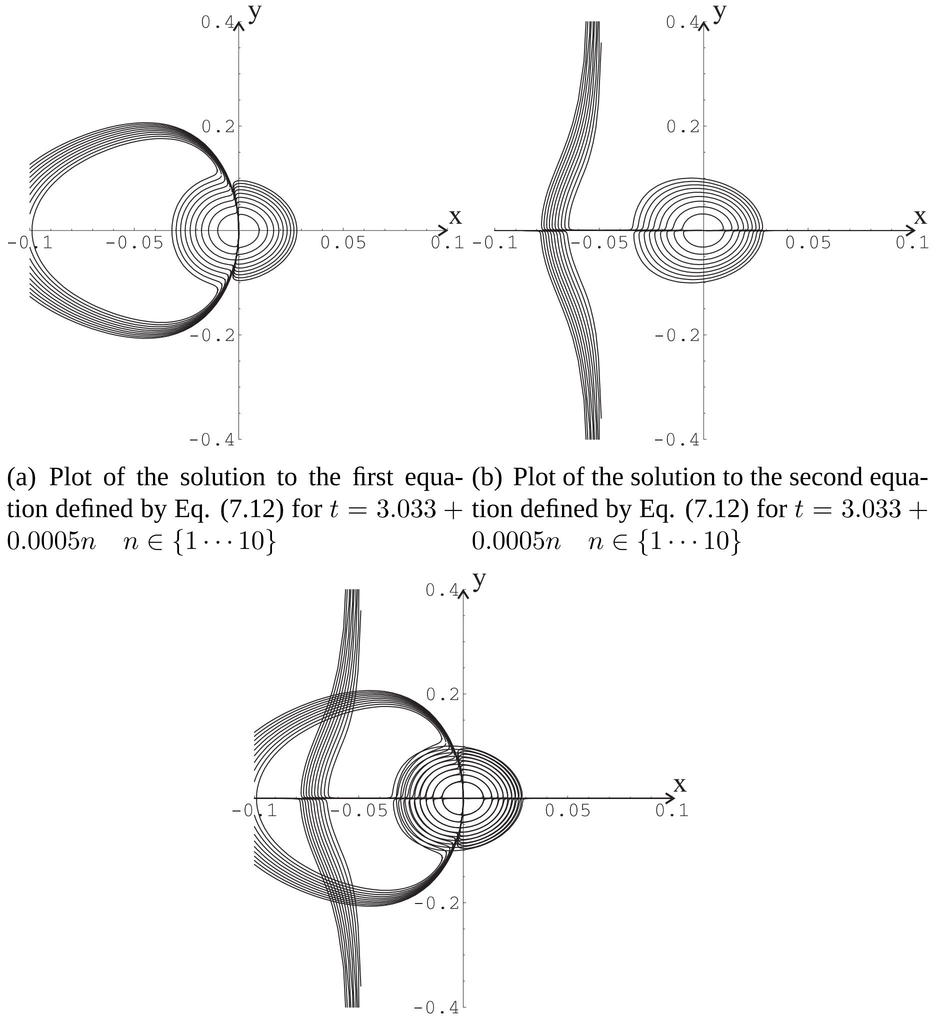
(c) Superposition of the two sets of solutions

Figure 7.3: Periodic orbits for the nonlinear motion about a Libration point

By plotting the intersection for different periods T, we generate a map of a family

[2] We point out that these points do not lie in the domain of use and are only consequences of our approximation of the dynamics.

of periodic orbits around the Libration point. In Fig. 7.4 we represent the solutions to Eq. (7.12) for $t = 3.033 + 0.0005n\,,\; n \in \{1 \cdots 10\}$. For $t = 3.033$ (which is less than the period of periodic orbits in the linearized system), the intersection only contains the origin, which is why there are only 9 periodic orbits shown around the origin. We note that at larger values of $x^2 + y^2$ the curves do not overlay precisely, indicating that higher order terms are needed.



(a) Plot of the solution to the first equation defined by Eq. (7.12) for $t = 3.033 + 0.0005n \quad n \in \{1 \cdots 10\}$

(b) Plot of the solution to the second equation defined by Eq. (7.12) for $t = 3.033 + 0.0005n \quad n \in \{1 \cdots 10\}$

(c) Superposition of the two sets of solutions for $t = 3.033 + 0.0005n \quad n \in \{1 \cdots 10\}$

Figure 7.4: Periodic orbits for the nonlinear motion about a Libration point

Periodic orbits in the vicinity of a given periodic orbit in the restricted three-body problem

We consider the normalized circular restricted three-body problem (Appendix C) with $\mu = 3.0359 \cdot 10^{-6}$ (this value of μ corresponds to the Earth-Sun mass ratio). The periodic orbit of period $T^* = 3.568576$ going through the point $(1.2, 0)$ is chosen to be the reference trajectory. It is represented in Fig. 7.5. In this section, we search for periodic orbits in the vicinity of the reference trajectory. To solve this problem, we use a polynomial approximation of the generating functions of order 5 computed using the algorithm developed in Chapter V. We emphasize the following two problems:

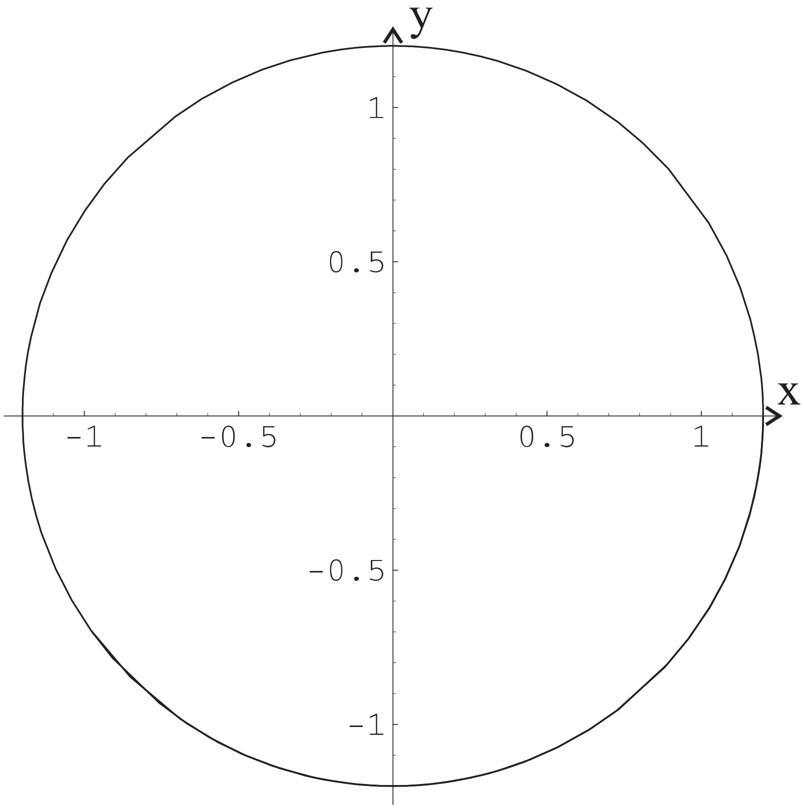


Figure 7.5: Periodic orbit in the restricted three-body problem with period $T = 3.568576$

1. *Searching in the time domain*: Given a point in relative position space $(1.23, 0.03)$, find the periods of all periodic orbits going through this point. In Fig. 7.6 we plot

the left hand side of Eq. (7.18). We notice the existence of two periodic orbits with respective periods T and $2T$ where $T = 3.62613$. In Fig. 7.7, we generate the obtained periodic orbit going through $(1.23, 0.03)$. Note that the approximation of the generating function provides an accurate picture of the true motion since the the periodic orbit repeats itself perfectly. In Fig. 7.7, the orbit repeats itself 30 times.

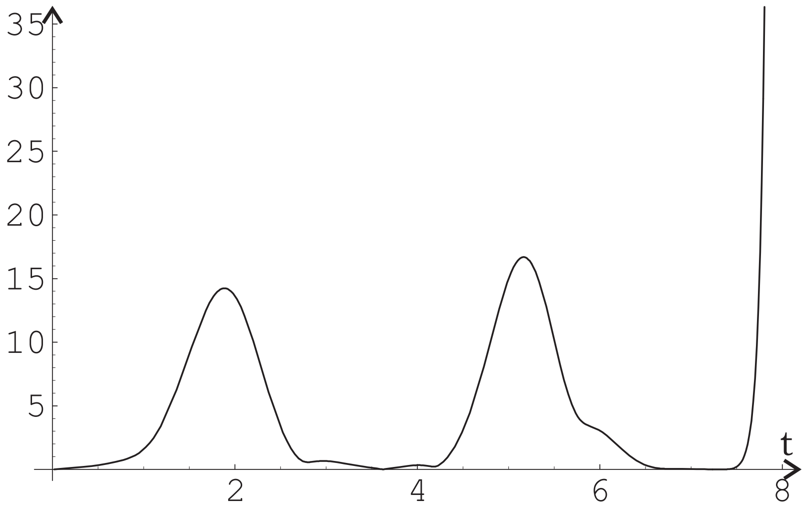


Figure 7.6: Plot of $\|\frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T)\|$ where $q_0 = (0.03, 0.03)$

2. *Searching in position space*: Let us now recover the previous periodic orbit from its period. In this case, we set $T = 3.62613$ and we use Eq. (7.12). Eq. (7.12) defines two equations with two variables. In Fig. 7.8 we have plotted the set of solutions to each of these equations and their superposition. The intersection of the two sets of solutions represent the set of solutions to Eq. (7.12) and is an arc of circle. We verify using Eqns. (7.5) that the intersection corresponds to a periodic orbit (we exactly recover the periodic orbit represented in Fig. 7.6).

   We note that our method does not recover the entire periodic orbit, because the entire orbit does not lie in the domain of convergence. Indeed, we should have found an almost circular trajectory close to the nominal one, that is, of radius larger than 1.2.

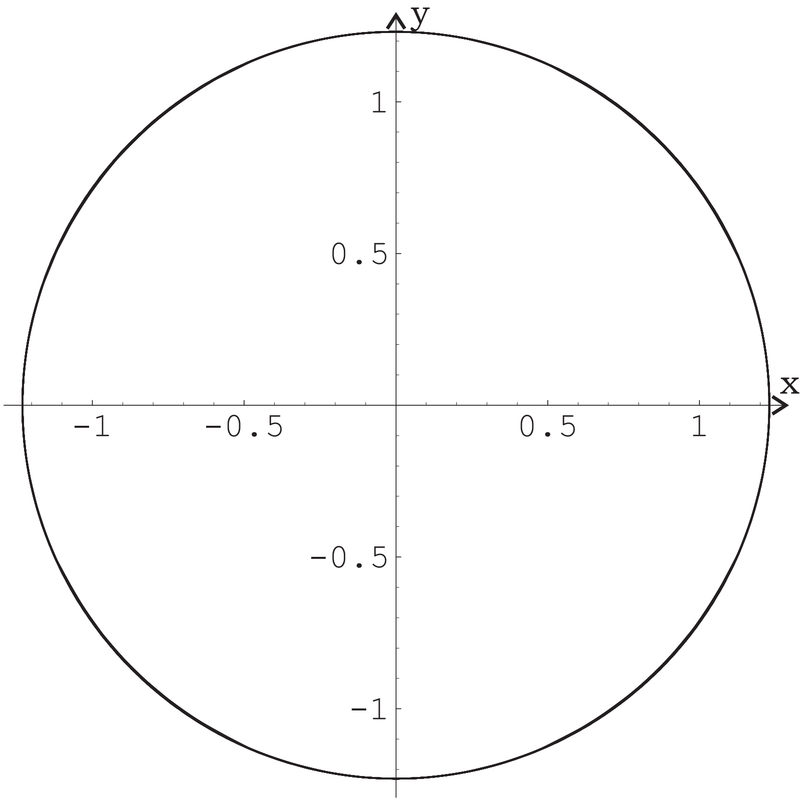
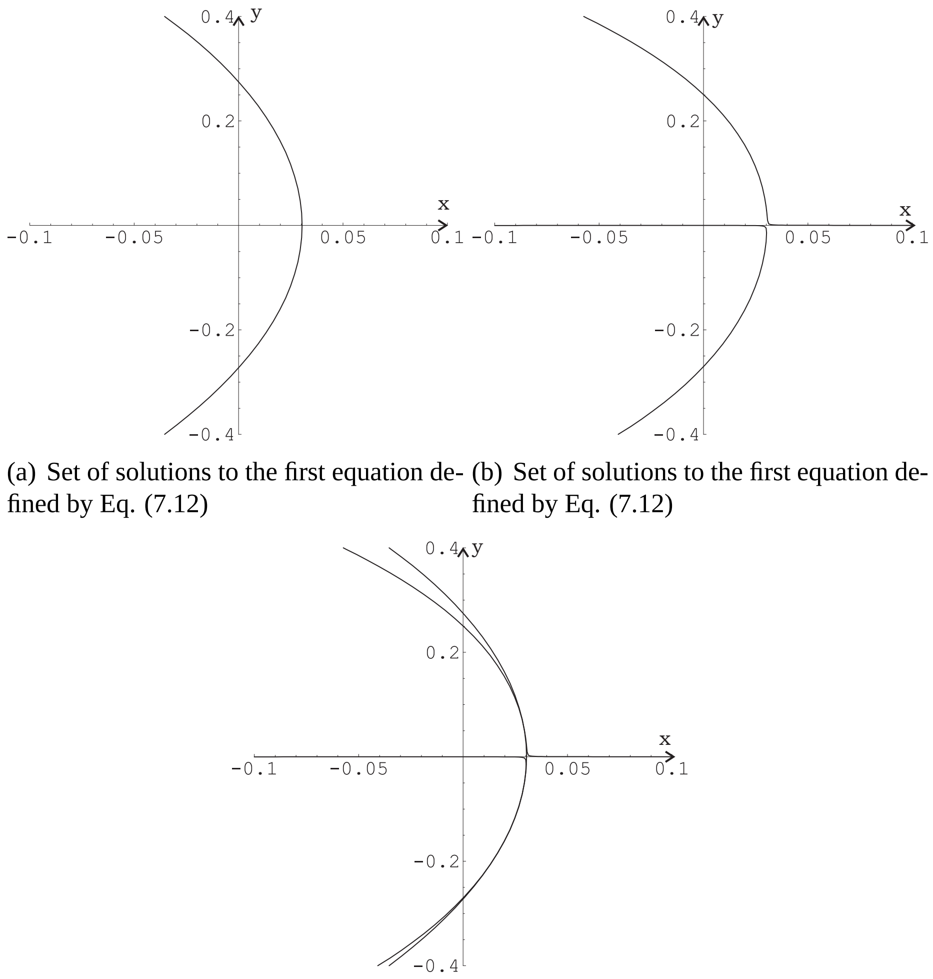


Figure 7.7: Periodic orbits going through the point $(1.23, 0.03)$

To describe such an orbit we must be able to predict relative motions that are as large as 2.4. Nonetheless, the set of intersections we find is enough to recover the whole trajectory using Eq. (3.8) and Hamilton's equations.

Finally, we can let the time vary and obtain a family of periodic orbits. In Fig. 7.9 we set $T = 3.58 + 0.01k\,,\ k \in [0, 7]$

To conclude, we have presented a novel approach for finding periodic orbits. The method we propose allows us to search for periodic orbits in phase space or in the time domain without requiring any initial guess or knowledge of a periodic orbit that belongs to the family. This is a major advantage compared to traditional methods. Most important, we reduce the search for periodic orbits to solving a nonlinear system of equations. Once the generating functions are known, no integration is required to find periodic orbits of



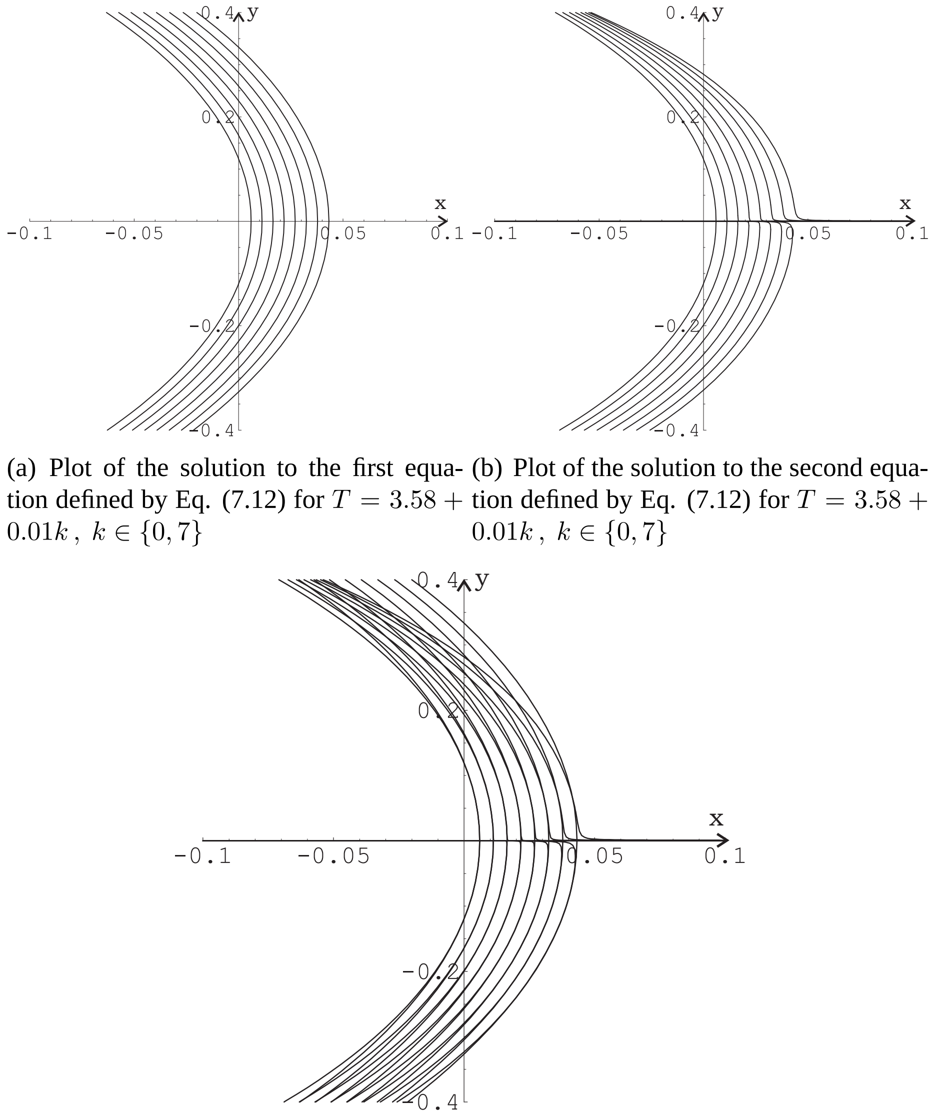
(a) Set of solutions to the first equation defined by Eq. (7.12)

(b) Set of solutions to the first equation defined by Eq. (7.12)

(c) Set of solutions to Eq. (7.12)

Figure 7.8: Periodic orbits of period $T = 3.62613$

different periods and/or going through different points in the phase space. This is a fundamental property of the generating functions that we will use again in the next chapters; once the generating functions are known, any two-point boundary value problem can be solved at the cost of a single function evaluation. Finally, we mention that searching in the time domain may not be as accurate as searching in phase space if one uses the algorithm we developed in Chapter V. Indeed, generating functions are expressed as polynomials with respect to their spatial coordinates with time-dependent coefficients. These coefficients are solutions of ordinary differential equations and are therefore known at certain



(a) Plot of the solution to the first equation defined by Eq. (7.12) for $T = 3.58 + 0.01k\,,\ k \in \{0, 7\}$

(b) Plot of the solution to the second equation defined by Eq. (7.12) for $T = 3.58 + 0.01k\,,\ k \in \{0, 7\}$

(c) Superposition of the two sets of solutions for $T = 3.58 + 0.01k\,,\ k \in \{0, 7\}$

Figure 7.9: Periodic orbits in the three-body problem

times (the nodes) only. As a result, solutions to Eq. (7.16) must be computed using an interpolation of the coefficients between the nodes.

CHAPTER VIII

SPACECRAFT FORMATION DYNAMICS AND DESIGN

Several missions and mission statements have identified formation flying as a means for reducing cost and adding flexibility to space-based programs. However, such missions raise a number of technical challenges as they require accurate dynamic models of the relative motion and control techniques to achieve formation reconfiguration and formation maintenance. There is a large literature on spacecraft formation flight that we will not attempt to survey in a systematic manner. On the one hand we find articles that focus on analytical studies of the relative motion, and on the other hand there are a large class of articles that develop numerical algorithms that solve specific reconfiguration and formation keeping problems. Theoretical studies require a dynamical model for the relative motion that is accurate and tractable. For that reason the Clohessy-Wilshire (CW) equations, Hill's equations or Gauss variational equations have often been used as a starting point. Using the CW equations, Hope and Trask [50] study hover type formation flying about the Earth, Vadali, Vaddi and Alfriend [92] look at periodic relative motion about the Earth, Gurfin and Kasdin[42], and Scheeres, Hsiao and Vinh[85] focus on formation keeping, Howell and Marchand[51], and Vadali, Bae and Alfriend [91] analyze relative motion in the vicinity of the libration points and Vaddi, Alfriend and Vadali [93] study the reconfiguration problem using impulsive thrusts. However, for a large class of orbits these approximations do not hold - J_2 effects as well as non-circular reference trajectory should be taken into account for low Earth orbits and an elliptic orbit for the primary should be considered to study the dynamics at the Libration points. As a result, past researchers have modified the CW equations in order to take the J_2 gravity coefficient into consideration. These improved equations have been widely used; Alfriend and Schaub [2] study periodic relative motion and Lovell, Horneman, Tollefson and Tragesser [62] analyze formation reconfiguration with impulsive thrusts. The non-impulsive thrust problem is usually solved using optimal control theory (although there are some exceptions, for instance F.Y. Hsiao and D.J. Scheeres[52] and I. Hussein, D.J. Scheeres and D. Hyland [53]), and if the dynamical model is tractable then analytical solutions for the feedback control law may be found (see Mishne [69]). These analytical approaches allow one to perform qualitative analysis and provide insight into the dynamics of the relative motion, however they cannot be used for actual mission design (except [3]). Indeed, they have inherent drawbacks: they neglect higher order terms in the dynamics and their domain of validity in phase space is very restricted and difficult to quantify. In addition, methods based on the state transition matrix tend to be valid only over short time spans. On the other hand, numerical algorithms have been developed to design spacecraft formations using the true dynamics. Koon, Marsden, Masdemont and Murray [59] use Routh reduction to reduce the dimensionality of the system and then develop an algorithm based on the use of the Poincaré map to find pseudo-periodic relative motion in the gravitational field of the Earth (including the J_2 gravity coefficient only), Xu and Fitz-Coy[100] and Avanzini, Biamonti and Minisci[9] study formation maintenance as a solution to an optimal control problem that they solve using a genetic algorithm and a multi-objective optimization algorithm respectively. Even though these methods use the exact dynamics and therefore can be used to solve a specific reconfiguration or formation maintenance problem, they fail (except [59]) to provide insight into the dynamics. In addition, as noticed by Wang and Hadaegh [94], formation reconfiguration design is a combinatorial problem. As a result the algorithms mentioned above are not appropriate for reconfiguration design as they require excessive computation (to reconfigure a formation of N spacecraft, there are $N!$ possibilities in general).

The method we expose in this chapter, based on the theory developed in Chapter III and on the algorithm presented in Chapter V, directly tackles these issues and should be viewed as a semi-analytic approach, since it consists of a numerical algorithm whose output is a polynomial approximation of the dynamics. As a consequence, we are able to use a very accurate dynamic model and to obtain tractable expressions describing the relative motion. A fundamental difference with previous studies is that we describe the relative motion, i.e., the phase space in the vicinity of a reference trajectory, as two-point boundary value problems whereas it is usually described as an initial value problem. Such a description of the phase space is very natural and convenient, For instance the reconfiguration problem and the search for periodic formations can be naturally formulated as two-point boundary value problems.

In this chapter, to showcase the strength of our method, we have chosen to study two challenging mission designs.

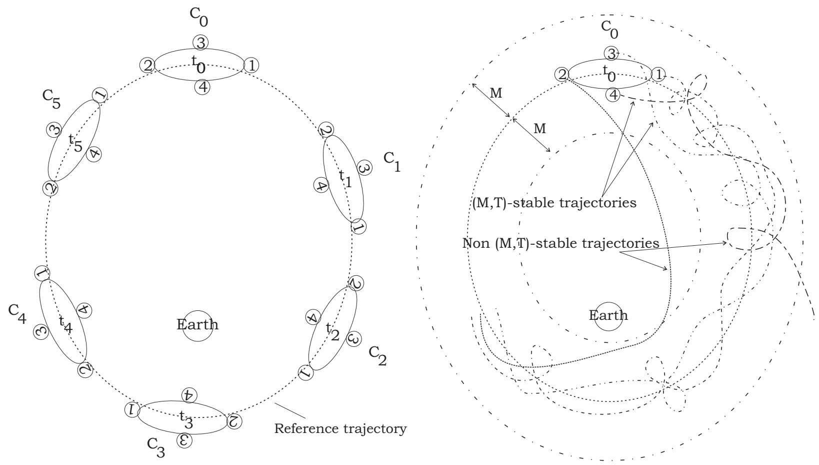
1. We first consider a spacecraft formation about an oblate Earth (the J_2 and J_3 gravity coefficients are taken into account) that must achieve 5 missions over a one month period. For each mission the formation must be in a given configuration C_i that has been specified beforehand, and we wish to minimize the overall fuel expenditure. The configurations C_i are specified as relative positions of the spacecraft with respect to a specified reference trajectory (Fig. 8.1(a)). The C_i's may be fully defined or have one degree of freedom. In our example we require the spacecraft to

be equally spaced on a circle centered on the reference trajectory at several epochs over the time period. The design of such a mission has several challenges:

- the dynamics are non-trivial and non-integrable,
- the reference trajectory has high eccentricity, high inclination and is not periodic,
- missions are planned a month in advance,
- in our specific example discussed here, 4 spacecraft must achieve 5 missions, if one assumes that the C_i are fully defined there are $7,962,624$ ways of satisfying the missions,
- the C_i may be defined by holonomic constraints and have an additional degree of freedom.

2. Next we consider the design of stable formations, the initial deployment of a formation and the redesign of an already deployed formation. For both problems, given a reference trajectory we wish to place the spacecraft in its vicinity and ensure that they remain “close” to each other over an extended period of time (see Fig. 8.1(b)). This design is also very challenging because:

    - the dynamics and the reference trajectory are non-trivial (as before),
    - trajectories must not collide (except at the initial time for the deployment problem),
    - high accuracy in the initial conditions is required for long-term integration.



(a) At each t_i, spacecraft must be in the configuration C_i.

(b) Stable and non-stable trajectories

Figure 8.1: The multi-task mission and the search for stable configurations.

8.1 Problem settings

The motion of a satellite under the influence of the Earth modeled by an oblate sphere (J_2 and J_3 gravity coefficients are taken into account) in the fixed coordinate system (x, y, z) whose origin is the Earth center of mass is described by the following Hamiltonian:

$$
\begin{aligned}
H = & \frac{1}{2}(p_x^2 + p_y^2 + p_z^2) \\
& - \frac{1}{\sqrt{x^2+y^2+z^2}} \left[1 - \frac{R^2}{2r_0^2(x^2+y^2+z^2)} \left(3\frac{z^2}{x^2+y^2+z^2} - 1 \right) J_2 \right. \\
& \left. \qquad - \frac{R^3}{2r_0^3(x^2+y^2+z^2)^2} \left(5\frac{z^3}{x^2+y^2+z^2} - 3z \right) J_3 \right] ,
\end{aligned}
$$

where

$$
\begin{aligned}
GM = 398600.4405\ km^3 s^{-2} , \quad & R = 6378.137\ km , \\
J_2 = 1.082626675 \cdot 10^{-3} , \quad & J_3 = 2.532436 \cdot 10^{-6} ,
\end{aligned}
\tag{8.1}
$$

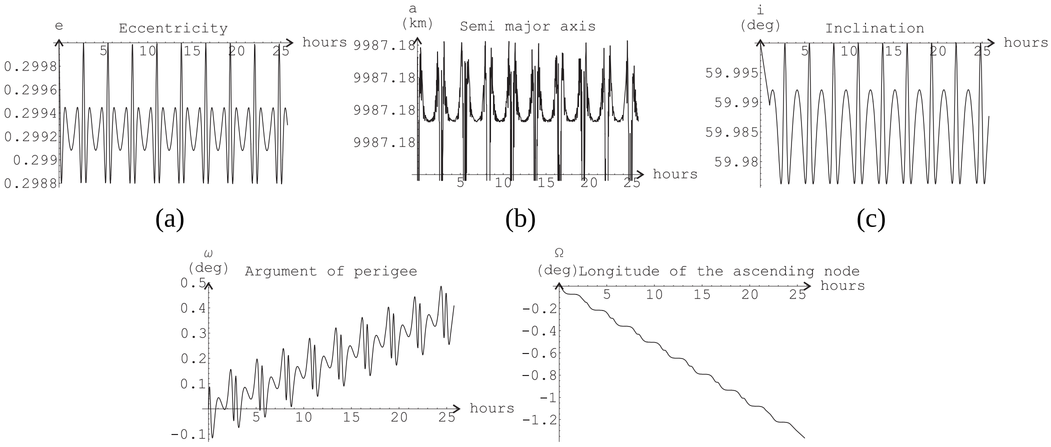
and all the variables are normalized (r_0 is the radius of the trajectory at the initial time):

$$\begin{array}{cccccccccccc} & & & x & \rightarrow & xr_0\,, & y & \rightarrow & yr_0\,, & z & \rightarrow & zr_0\,, \\ t & \rightarrow & t\sqrt{\dfrac{r_0^3}{GM}}\,, & p_x & \rightarrow & p_x\sqrt{\dfrac{GM}{r_0}}\,, & p_y & \rightarrow & p_y\sqrt{\dfrac{GM}{r_0}}\,, & p_z & \rightarrow & p_z\sqrt{\dfrac{GM}{r_0}}\,. \end{array} \tag{8.2}$$

In the following, we consider a reference trajectory whose state is designated by (q^0, p^0) and study the relative motion of spacecraft with respect to it. The reference trajectory is chosen to be highly eccentric and inclined, but any other choice could have been considered. At the initial time its state is:

$$\begin{array}{lll} q_x^0 = r_p\,, & q_y^0 = 0\ km\,, & q_z^0 = 0\ km\,, \\ p_x^0 = 0\ kms^{-1}\,, & p_y^0 = \sqrt{\frac{GMr_a}{\frac{1}{2}r_p(r_a+r_p)}}\cos(\alpha)\ kms^{-1}\,, & p_z^0 = \sqrt{\frac{GMr_a}{\frac{1}{2}r_p(r_a+r_p)}}\sin(\alpha)\ kms^{-1}\,, \\ \alpha = \frac{\pi}{3}rad\,, & r_p = 7,000\ km\,, & r_a = 13,000\ km\,. \end{array} \tag{8.3}$$

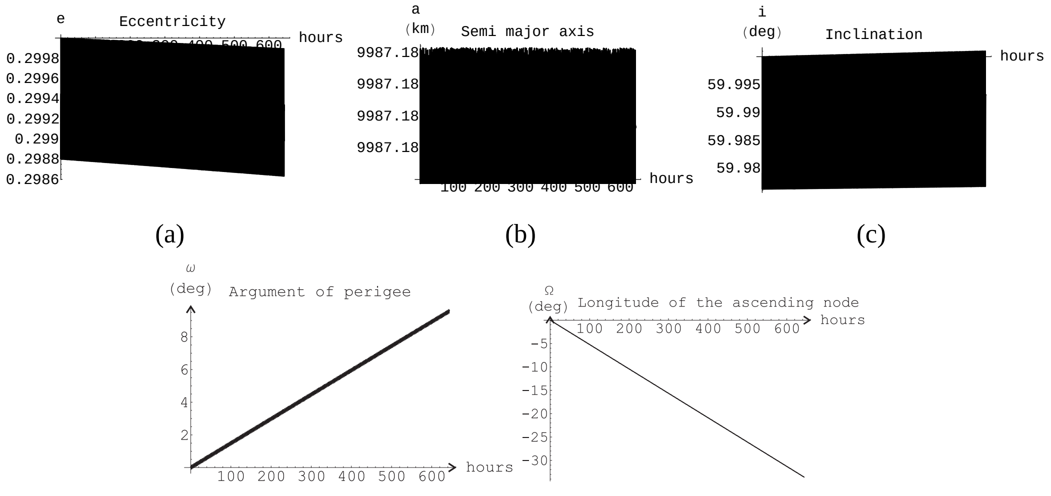
Without the J_2 and J_3 gravity coefficients the reference trajectory would be an elliptic orbit with eccentricity $e = 0.3$, inclination $i = \pi/3\ rad$, argument of perigee $\omega = 0$, longitude of the ascending node $\Omega = 0$, semi-minor axis $r_p = 7,000\ km$, semi-major axis $r_a = 13,000\ km$ and of period $t_p = 2\pi\sqrt{\frac{1}{2^3}\frac{(r_a+r_p)^3}{r_p^3}}\ sec \approx 2\ hours\ 45\ min$. The Earth oblateness perturbation causes (see Chobotov [23] for more details) secular drifts in the eccentricity (due to J_3), in the argument of perigee (due to J_2 and J_3) and in the longitude of the ascending node (due to J_2 and J_3). In addition, all the orbit elements are subject to short and long period oscillations. In Fig. 8.2 and 8.3, we plot the orbit elements for this trajectory as a function of time during a day (about 10 revolutions about the Earth) and over a month period. The symplectic implicit Runge-Kutta integrator built in $Mathematica^{©}$ is used for integration of Hamilton's equations.



(a) (b) (c)

(d) (e)

Figure 8.2: Time history of the orbital elements over a one day period



(a) (b) (c)

(d) (e)

Figure 8.3: Time history of the orbital elements over a one month period

8.2 Formation design

We introduced a dynamical model and defined a reference trajectory. In the previous chapters we presented an algorithm whose outputs are the generating functions associated with the phase flow describing the relative motion. In addition, we explained how these generating functions may be used to solve two-point boundary value problems. We now combine all the above and use it to design spacecraft formations. We first use the "combined" algorithm to find the generating function F_1 up to order 4, that is we need to solve 498 ordinary differential equations in the indirect method, then proceed a series inversion and solve the 203 ordinary differential equations given by the direct method (see appendix for computational times). Once the generating functions are known, we can solve any position to position boundary value problem with only six polynomial evaluations (Eqs. (3.7) and (3.8)).

8.2.1 Multi-task mission

We consider four imaging satellites flying in formation about the reference trajectory. We want to plan spacecraft maneuvers over the next month knowing that they must observe the Earth, i.e., must be in a given configuration C_i at the following instants (chosen arbitrarily for our study:

$$\begin{aligned} t_0 = 0\,, \qquad\qquad & t_1 = 5\ days\ 22\ hours\,, \quad t_2 = 10\ days\ 20\ hours\,, \\ t_3 = 16\ days\ 2\ hours\,, \quad & t_4 = 21\ days\ 14\ hours\,, \quad t_5 = 26\ days\ 20\ hours\,. \end{aligned} \tag{8.4}$$

Define the local horizontal by the unit vectors $(\hat{e}_1, \hat{e}_2)$ such that $\hat{e}_2$ is along $r^0 \times v^0$ and $\hat{e}_1$ is along $\hat{e}_2 \times r^0$, then at every t_i, the configuration C_i is defined by the four following relative positions (or slots):

$$q^1 = 700\ m\ \hat{e}_1\,, \quad q^2 = -700\ m\ \hat{e}_1\,, \quad q^3 = 700\ m\ \hat{e}_2\,, \quad q^4 = -700\ m\ \hat{e}_2\,. \tag{8.5}$$

Note that at t_i, q^1 is in front of the reference state (in the local horizontal plane), q^2 is behind, q^3 is on the left and q^4 is on the right (see Fig. 8.1(a)). At each t_i, there must be one spacecraft per slot and we want to determine the sequence of reconfigurations that minimizes the total fuel expenditure (other cost functions such as equal fuel consumption for each spacecraft may be considered as well). For the first mission, there are $4!$ configurations (number of permutation of the set $\{1, 2, 3, 4\}$), for the second mission, for each of the previous $4!$ configurations, there are again $4!$ configurations, that is a total of $4!^2$ possibilities. Thus for 5 missions there are $4!^5 = 7,962,624$ possible configurations.

In this paper, we focus on impulsive controls, but the method we develop can equivalently apply to continuous thrust problems. Indeed, continuous thrust problems are usually solved using optimal control theory and reduce to a set of necessary conditions that are formulated as a Hamiltonian two-point boundary value problem. This boundary value problem can in turn be solved using the method we present in this paper [86]. Let us now design the above mission. We assume impulsive controls that consist of impulsive thrusts applied at $t_{i\in[0,5]}$. For each of the four spacecraft, we need to compute the velocity at t_i so that the spacecraft moves to its position specified at t_{i+1} under gravitational forces only. As a result, we must solve $5 \cdot 4! = 120$ position to position boundary value problems (given two positions at t_i and t_{i+1}, we need to compute the associated velocity). Using the generating functions, this problem can be handled at the cost of only 120 function evaluations. Then, we need to compute the cost function (sum of the norm of all the required impulses, assuming zero relative velocities at the initial and final times) for all the permutations (there are $7,962,624$ combinations) to find the sequence that minimizes the cost function. Fig. 8.4 represents the number of configurations as a function of the values of the cost function. We notice that most of the configurations require at least three times more fuel than the best configuration, and less than 6% yield values of the cost function that are less than twice the value associated with the best configuration. The cost function for the optimal sequence of reconfigurations is $0.00644\ km \cdot s^{-1}$ whereas it is $0.0396\ km \cdot s^{-1}$ in the least optimal design. In the optimal case, the four spacecraft have the following positions:

Spacecraft 1: (t_0, q^1), (t_1, q^2), (t_2, q^2), (t_3, q^2), (t_4, q^2), (t_5, q^2).
Spacecraft 2: (t_0, q^2), (t_1, q^1), (t_2, q^1), (t_3, q^1), (t_4, q^1), (t_5, q^1).
Spacecraft 3: (t_0, q^3), (t_1, q^4), (t_2, q^4) (t_3, q^4), (t_4, q^3), (t_5, q^4).
Spacecraft 4: (t_0, q^4), (t_1, q^3), (t_2, q^3) (t_3, q^3), (t_4, q^4), (t_5, q^3).

whereas the worst scenario corresponds to:

Spacecraft 1: (t_0, q^1), (t_1, q^1), (t_2, q^2), (t_3, q^2), (t_4, q^1), (t_5, q^2).
Spacecraft 2: (t_0, q^2), (t_1, q^2), (t_2, q^3), (t_3, q^4), (t_4, q^4), (t_5, q^3).
Spacecraft 3: (t_0, q^3), (t_1, q^3), (t_2, q^1) (t_3, q^3), (t_4, q^3), (t_5, q^4).
Spacecraft 4: (t_0, q^4), (t_1, q^4), (t_2, q^4) (t_3, q^1), (t_4, q^2), (t_5, q^1).

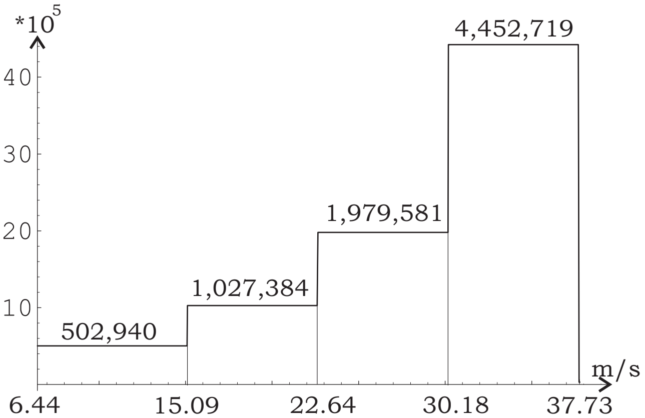


Figure 8.4: Number of configurations as a function of the value of the cost function

We may verify, *a posteriori*, if the solutions found meet the mission goals, i.e., if the order 4 approximation of the dynamics is sufficient to simulate the true dynamics.

Explicitly comparing the analytical solution with numerically integrated results shows that the spacecraft are at the desired positions at every t_i with a maximum error of $1.5{\cdot}10^{-8}\ km$.

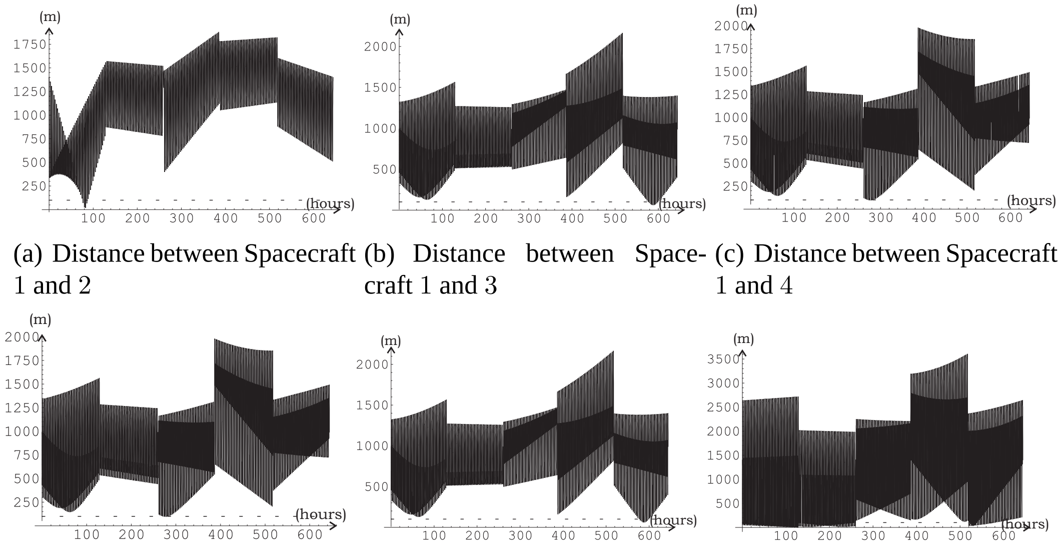
Considerations on collision management

Our algorithm does not consider the risk of collision in the design. However, it provides a simple way to check afterwards if there is collision. Recall the indirect method. It is based on the initial value problem and essentially consists in solving Hamilton's equations for an approximation of the flow. Once such a solution is found, we can generate any trajectory at the cost of a function evaluation, there is no need to integrate Hamilton's equations again. Checking for collisions is again a combinatorial problem and therefore our approach is particularly adapted to this. As an example let us verify if the design we proposed for the multi-task mission yields collisions. In figure 8.5 we plot the distance between each of the spacecraft. We remark that spacecraft 1 and 2, 1 and 3, 2 and 3, and 3 and 4 may collide (relative distance less than $100\ m$). A detail of the figure shows that spacecraft 3 and 4 collide whereas the other spacecraft have a relative distance larger than 40 meters.

It can be proven that for this specific mission, there is no design that prevents the relative motion of the spacecraft to be less than $100\ m$. In the best scenario, the smallest relative distance between the spacecraft is about $15\ m$, and is achieved in $3,360$ different designs. Among these $3,360$ possibilities, we represent in Fig. 8.6 the time history of the relative distance between the spacecraft for the design that achieves minimum fuel expenditure (the total fuel expenditure is $60\ \%$ larger than in the best case). This scenario corresponds to:

Spacecraft 1: (t_0, q^1), (t_1, q^2), (t_2, q^3), (t_3, q^3), (t_4, q^4), (t_5, q^3).

Spacecraft 2: (t_0, q^2), (t_1, q^3), (t_2, q^4), (t_3, q^4), (t_4, q^3), (t_5, q^4).



(a) Distance between Spacecraft 1 and 2 (b) Distance between Spacecraft 1 and 3 (c) Distance between Spacecraft 1 and 4

(d) Distance between Spacecraft 2 and 3 (e) Distance between Spacecraft 2 and 3 (f) Distance between Spacecraft 3 and 4

Figure 8.5: Distance between the spacecraft as a function of time for the best scenario

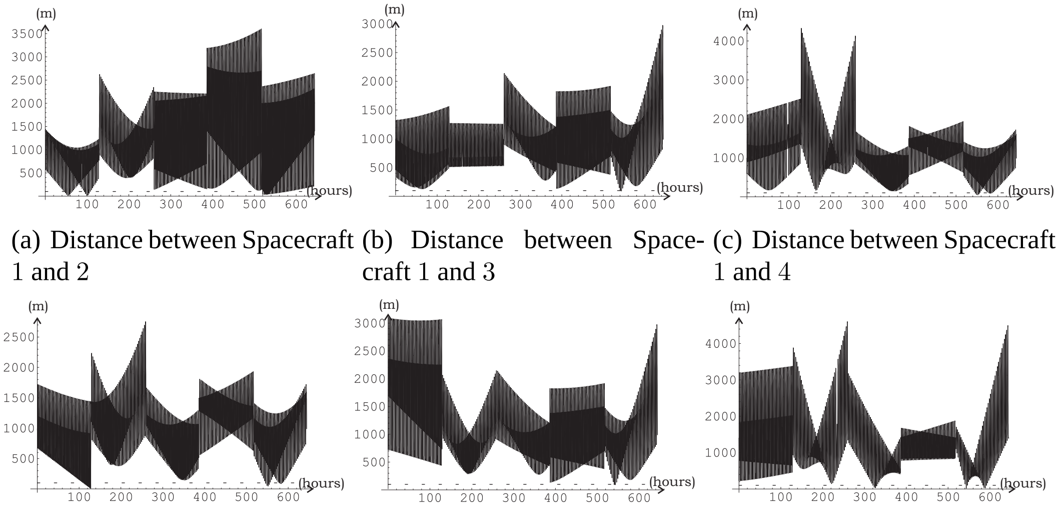
Spacecraft 3: (t_0, q^3), (t_1, q^4), (t_2, q^1) (t_3, q^2), (t_4, q^1), (t_5, q^2).

Spacecraft 4: (t_0, q^4), (t_1, q^1), (t_2, q^2) (t_3, q^1), (t_4, q^2), (t_5, q^1).

For times at which the spacecraft are close to each other, we may use some local control laws to perform small maneuvers for ensuring appropriate separation.

Another option consists of changing the configurations at t_i so that there exists a sequence of reconfigurations such that the relative distance between the spacecraft stay larger than $100\ m$. This can easily be done using our approach since F_1 is already known. Solving a new design would only require 120 evaluations of the gradient of F_1.

In the above example we take advantage of our algorithm to perform the required design, that is, we are able to plan missions involving several spacecraft over a month using non-trivial dynamics while minimizing a given cost function. Such a design is possible because we focus directly on specifying the problem as a series of boundary value problems. Solution of this problem using a more traditional approach to solving boundary



(a) Distance between Spacecraft 1 and 2

(b) Distance between Spacecraft 1 and 3

(c) Distance between Spacecraft 1 and 4

(d) Distance between Spacecraft 2 and 3

(e) Distance between Spacecraft 2 and 3

(f) Distance between Spacecraft 3 and 4

Figure 8.6: Distance between the spacecraft as a function of time

value problems would have required direct integration of the equations of motion for each of the 720 boundary value problems.

However, we have not taken full advantage of our algorithm yet, as the above example does not provide insight on the dynamics. We now consider a different mission to remedy this and show how our algorithm may be used for analytical studies.

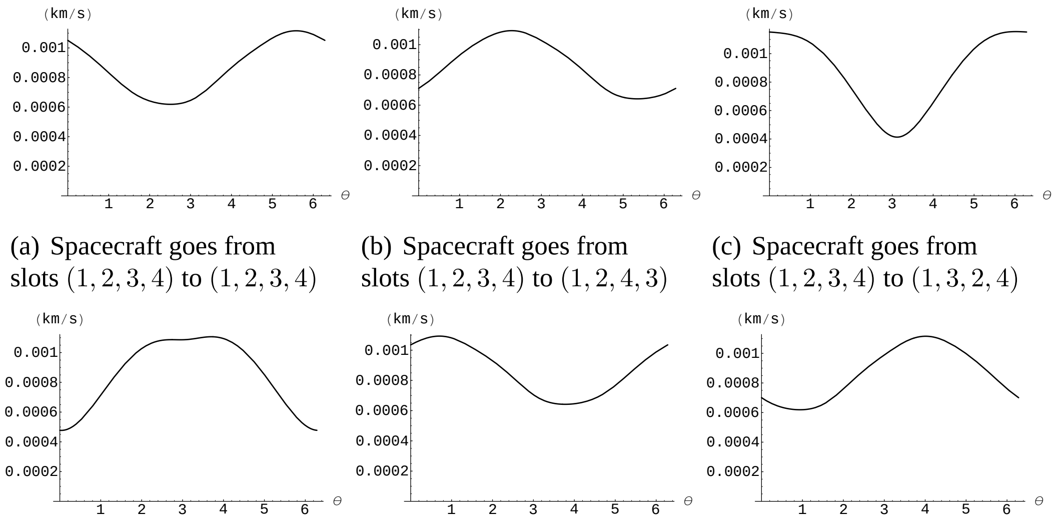
8.2.2 A different multi-task mission

For simplicity, we assume that the spacecraft must achieve only one task, that is we constrain the geometry of the formation at t_0 and t_1. However, instead of imposing absolute relative positions, we only require the spacecraft to be equally spaced on a circle of a given radius in the local horizontal plane at t_1. Such a constraint is more realistic, especially for imaging satellites as rotations of the formation about the local vertical should not influence performance. In this problem, combinatorics and smooth functional analysis are mixed together. Indeed, the positions of the four slots are given by a variable θ (θ indicates the position of the first slot, the other slots are determined from the constraint that they should be equally spaced). Then, we need to solve a combinatorial problem as in the previous case. To find the θ that minimizes the cost function, we use the polynomial approximation of the generating functions provided by our algorithm to express the cost function as a one dimensional polynomial in θ. Variations of the cost function are determined analytically by computing the derivative of the cost function.

We choose the initial position to be as in the previous example and require the spacecraft to be equally spaced at t_1 on a circle of radius $700\ meters$ in the local horizontal plane. In addition, we assume zero relative velocities at the initial and final times and again choose the cost function to be the sum of the norm of the required impulses. As before, $(\hat{e}_1, \hat{e}_2)$ span the local horizontal plane and we define θ as the angle between the relative position vector and $\hat{e}_1$. Since θ is allowed to vary from 0 to 2π (i.e., slot 1 describes the whole circle as θ goes from 0 to 2π), we may consider that spacecraft 1 always goes from slot 1 to slot 1. As a consequence, there are $3!$ free configurations. In Fig. 8.7, we plot the values of the cost function as a function of θ for each of the configurations. The best design is the one for which $\theta = 3.118\ rad$, spacecraft 1 goes from slot 1 to slot 1, spacecraft 2 from 2 to 3, spacecraft 3 from 3 to 2 and spacecraft 4 from 4 to 4.

If several missions need to be planned, then a new variable is introduced for each and a multi-variable polynomial must be studied. As a result, minima of the cost function are found by evaluating as many derivatives as there are missions.

Through this example, we have gained insight on the dynamics by using the analytical approximation of the generating function and were able to solve the fuel optimal reconfiguration problem. The method we use is very general and can be applied to solve any reconfiguration problem given that the constraints on the configurations are holonomic.



(a) Spacecraft goes from slots $(1, 2, 3, 4)$ to $(1, 2, 3, 4)$

(b) Spacecraft goes from slots $(1, 2, 3, 4)$ to $(1, 2, 4, 3)$

(c) Spacecraft goes from slots $(1, 2, 3, 4)$ to $(1, 3, 2, 4)$

(d) Spacecraft goes from slots $(1, 2, 3, 4)$ to $(1, 3, 4, 2)$

(e) Spacecraft goes from slots $(1, 2, 3, 4)$ to $(1, 4, 2, 3)$

(f) Spacecraft goes from slots $(1, 2, 3, 4)$ to $(1, 4, 3, 2)$

Figure 8.7: Fuel expenditure as a function of θ for each configuration

8.2.3 Stable trajectories

Now we focus on another crucial, but difficult, design issue for spacecraft formations. We search for configurations, called stable configurations, such that spacecraft stay close to the reference trajectory over a long time span.

Definitions

Let us first define the notion of stable formation more precisely. Let T be a given instant and M a real number.

Definition VIII.1 (Stable relative trajectory). *A relative trajectory between two spacecraft is (M, T)-stable if and only if their relative distance never exceeds M over the time span $[0, T]$.*

Definition VIII.2 (Stable formation). *A formation of spacecraft is (M, T)-stable if and only if all the spacecraft have (M, T)-stable relative trajectories with respect to the reference trajectory.*

Periodic formations are instances of stable formations, they are (M,∞)-stable. We also point out that our definition recovers the notion of Lyapunov stability: Lyapunov stable relative trajectories are (M,∞)-stable relative trajectories. In this paper, we focus on (M,T)-stable formations with T large but finite, the approach we present is not appropriate to find (M,∞)-stable configurations. However, when the reference trajectory is periodic Guibout and Scheeres[33] developed a technique based on generating functions and Hamilton-Jacobi theory to find periodic configurations.

Stable trajectories as solutions to two-point boundary value problems

In order to use the theory we have presented above, we formulate the search for stable trajectories as two-point boundary value problems.

Define the local vertical plane as the two-dimensional vector space perpendicular to the velocity vector of the reference trajectory. In other words, the local vertical is spanned by $(\hat{f}_1,\hat{f}_2)$ where $\hat{f}_1$ and $\hat{f}_2$ are two unit vectors along $r^0\times v^0$ and $v^0\times\hat{f}_1$ respectively. In the local vertical plane, we use polar coordinates, $(r-r^0,\theta)$, θ being the angle between $\hat{f}_1$ and the local relative position vector $r-r^0$. We denote by $\mathcal{C}_t^r$ the circle of radius r centered on the reference trajectory that lies in the local vertical plane at t. A position on this circle is fully determined by θ (see Fig. 8.8). Then, given an instant $t_f>t_0$ and a distance $r_f>0$, the circle $\mathcal{C}_{t_f}^{r_f}$ is defined.

Before searching for stable configurations, we first introduce a new methodology to find (M,T)-stable relative trajectories for a single spacecraft about the reference trajectory defined above. Consider the following two-point boundary value problem:
Find all trajectories going from the initial position of the spacecraft to any point on $\mathcal{C}_{t_f=T}^{r_f}$ in T units of time where $r_f<M$ (Fig. 8.9).

Solutions to this boundary value problem have the following properties:

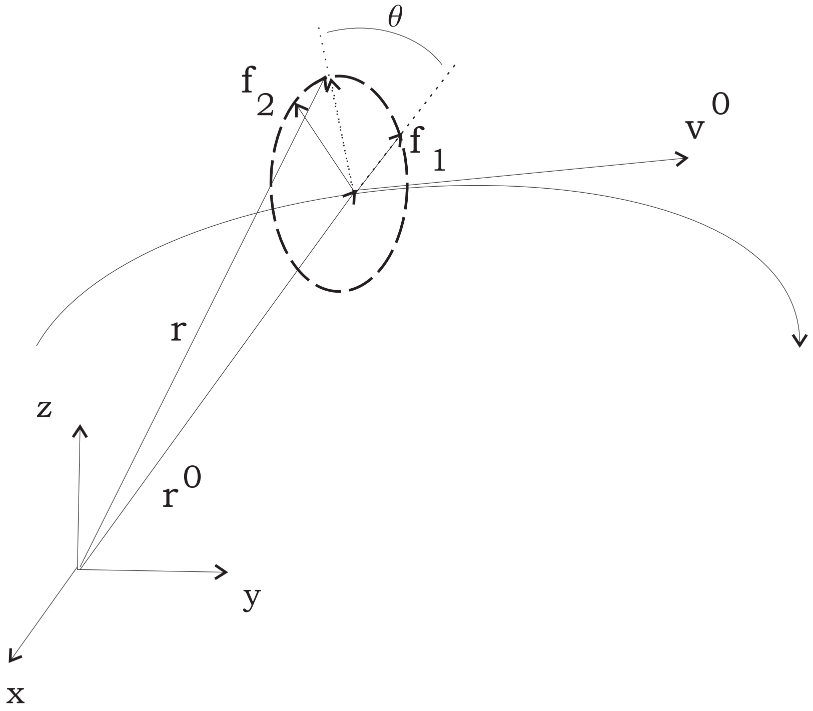


Figure 8.8: Representation of the local geometry

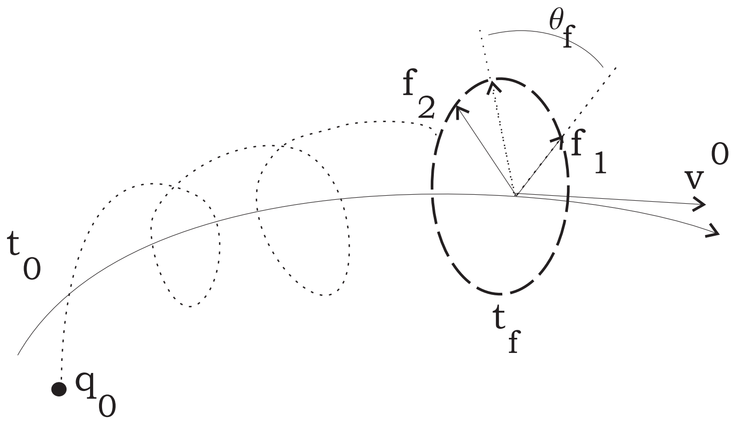


Figure 8.9: Boundary value problem

1. they contain (M, T)-stable relative trajectories.

2. they contain relative trajectories that are not (M, T)-stable, i.e., trajectories that go far from the reference trajectory in the time interval $(0, t_f = T)$ but come back close to the reference trajectory at t_f. We point out that many of these trajectories are ignored by our algorithm since it uses a local approximation of the dynamics.

On the other hand, we know that stable trajectories must have similar orbit elements as compared to the reference trajectory. Therefore, to discriminate between the solutions to the two-point boundary value problem we can use orbit elements, especially since we know, *a priori*, that the longitude of the ascending node and the argument of perigee have secular drifts. This leads us to define a cost function J as:

$$J = \frac{1}{4}\|\Delta\omega_{t_f}\| + \frac{1}{4}\|\Delta\omega_{t_f} - \Delta\omega_{t_0}\| + \frac{1}{4}\|\Delta\Omega_{t_f}\| + \frac{1}{4}\|\Delta\Omega_{t_f} - \Delta\Omega_{t_0}\| \,, \tag{8.6}$$

where $\|\Delta\omega_{t_f}\|$ corresponds to the relative argument of perigee at t_f, i.e, the difference at t_f between the argument of perigee of the spacecraft trajectory and the argument of perigee of the reference trajectory, $\|\Delta\omega_{t_f} - \Delta\omega_{t_0}\|$ characterizes the change in the relative argument of perigee between t_0 and t_f and the other terms are similar and involve the longitude of the ascending node instead.

Let us now consider the following boundary value problem: Find all trajectories going from the initial position of the spacecraft to any point on $\mathcal{C}_{t_f}^{r_f}$ in $t_f - t_0$ units of time that minimize J.

From the above discussion, we conclude that solutions to this boundary value problem characterize stable relative trajectories.

Methodology

We showed in the previous section that the search for stable trajectories reduces to solving a two-point boundary value problem while minimizing a given cost function. In this section, we solve this problem using generating functions.

First we notice that F_1 solves the boundary value problem that consists of going from an initial position q_0 to a position q_f in t_f units of time. Indeed, from Eqns. (3.7) and (3.8) we have:

$$p_0 = -\frac{\partial F_1}{\partial q_0}(q_f, q_0, t_f)\,, \tag{8.7}$$

$$p_f = \frac{\partial F_1}{\partial q}(q_f, q_0, t_f)\,. \tag{8.8}$$

Then we assume that q_f describes $\mathcal{C}_{t_f}^{r_f}$, that is, $q_f = r_f \cos(\theta_f)\hat{f}_1 + r_f \sin(\theta_f)\hat{f}_2$ where θ_f ranges from 0 to 2π. Since F_1 is approximated by a polynomial in (q_f, q_0) with time-dependent coefficients, Eqns. (8.7) and (8.8) allow us to express p_0 and p_f as polynomials in θ_f with time-dependent coefficients. Finally, with knowledge of $p_0(\theta_f)$, $p_f(\theta_f)$, q_0 and $q_f(\theta_f)$, we can express J as a function of θ_f and easily find its minima $\{\theta_f^1, \cdots, \theta_f^r\}$. Stable trajectories are then those that travel from q_0 to $q_f = r_f \cos(\theta_f^i)\hat{f}_1 + r_f \sin(\theta_f^i)\hat{f}_2$, $i \in [1, r]$ in t_f units of time.

Example

Let us illustrate this procedure by searching for stable trajectories for a spacecraft whose initial position relative to the reference trajectory at the initial time is $q_0 = (495, -428.6, 247.5)\ m$ in the inertial frame or equivalently $q_0 = 700\cos(\pi/4)\hat{f}_1 + 700\sin(\pi/4)\hat{f}_2\ \ m$. We use an order 4 approximation of the dynamics, $t_f = 10\ d\ 19\ h\ 13\ min$ and $r_f = 700\ m$. Then, using a symbolic manipulator, we express J as a function of θ_f and plot its values in Fig. 8.10. It has two local minima at $\theta_1 = 0.671503\ rad$ and $\theta_2 = 2.4006615\ rad$ that correspond to stable trajectories. The relative motions associated with these two trajectories are represented in Fig. 8.11 and 8.12 over time spans smaller and larger than t_f. We notice the excellent behavior of these trajectories, they remain stable over a time interval larger than the one initially considered.

We also point out that one of the trajectories (figure 8.11) is (r_f, t_f)-stable whereas the other one (figure 8.12) is $(3r_f, t_f)$-stable.

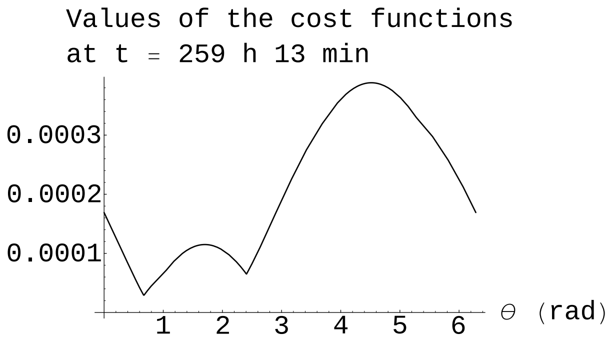
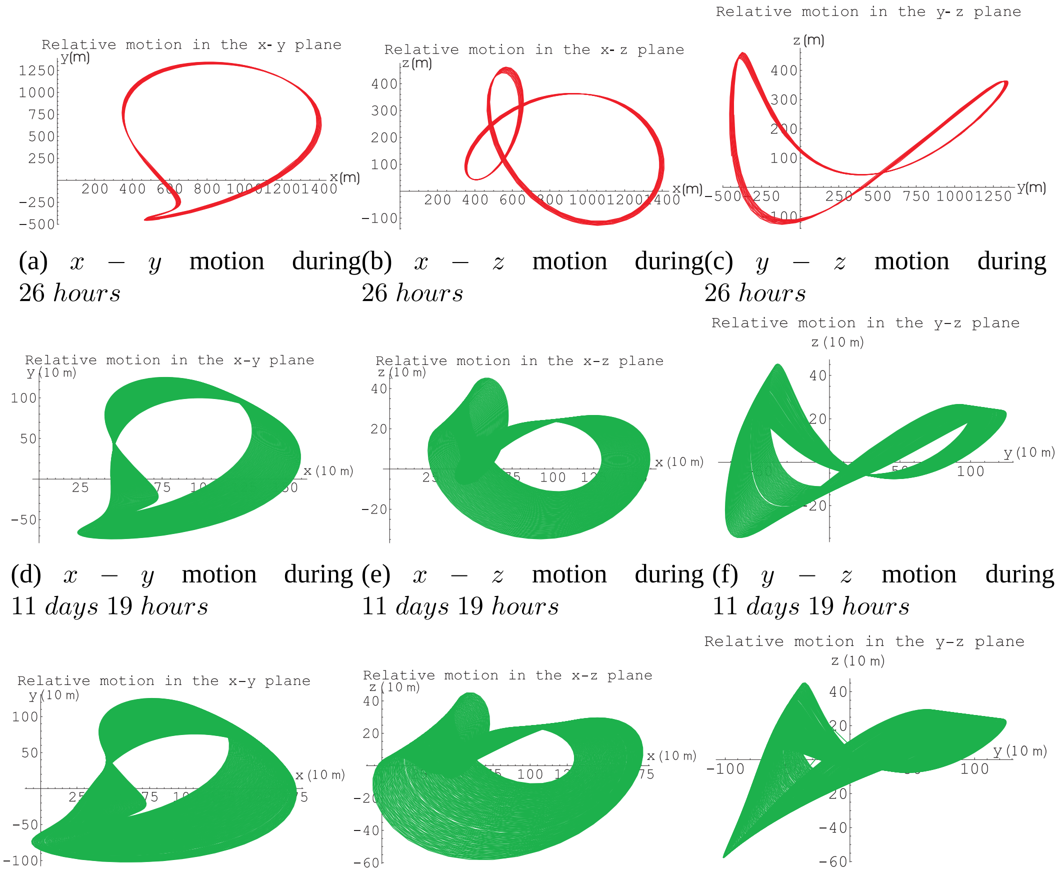


Figure 8.10: Cost function as a function of θ for $t_f = 10d19h13m$

Before going further, let us discuss the role played by t_f. We transformed the search for stable trajectories into a boundary value problem over a time span defined by t_f that we apparently chose arbitrarily. By varying t_f, we notice that minima of the cost function correspond to different stable trajectories. In Fig. 8.13 we plot the cost function as a function of θ_f for $t = t_f - 1\ h\ 6\ min = 10\ d\ 18\ h\ 19\ min$. In contrast to the previous case, the cost function has only one minimum at $\theta = 3.814575\ rad$. In Fig. 8.14 we represent the trajectory that corresponds to this minimum. It is stable but different from the previous ones (Fig. 8.11 and 8.12). This result was expected and makes our approach even more valuable. Indeed, since we reduced the search for stable trajectories to a boundary value problem, we completely ignore the behavior of the spacecraft at intermediary times $t \in [0, t_f]$, we only take into account the states of the spacecraft at the initial time and at t_f. As a result, short term oscillations play a major role and alter the locus of the minima of J. Thus, by varying t_f we are potentially able to find infinitely many stable trajectories going through q_0 at the initial time. This aspect allows us to design a deployment problem, for instance, where several spacecraft are at the same location at the initial time and we



(a) $x - y$ motion during 26 *hours*

(b) $x - z$ motion during 26 *hours*

(c) $y - z$ motion during 26 *hours*

(d) $x - y$ motion during 11 *days* 19 *hours*

(e) $x - z$ motion during 11 *days* 19 *hours*

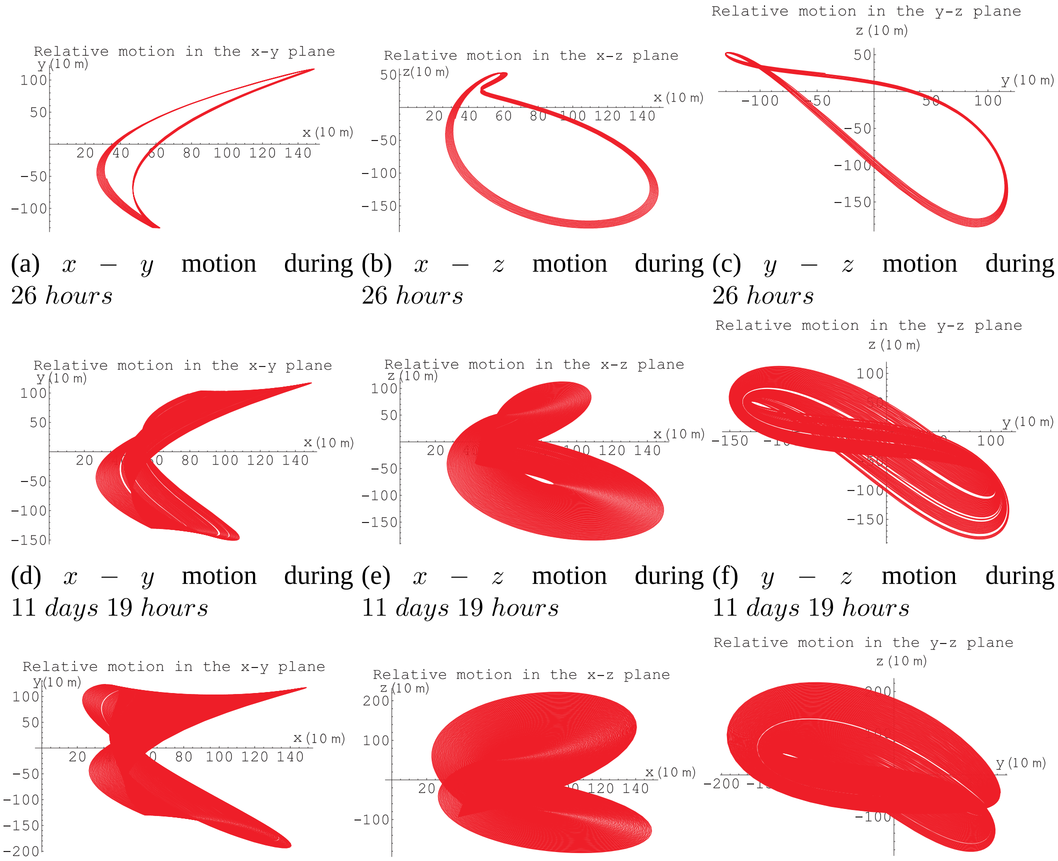
(f) $y - z$ motion during 11 *days* 19 *hours*

(g) $x - y$ motion during 21 *days* 11 *hours*

(h) $x - z$ motion during 21 *days* 11 *hours*

(i) $y - z$ motion during 21 *days* 11 *hours*

Figure 8.11: Trajectory associated with the minimum $\theta = 0.671503\ rad$, $t_f = 10\ d\ 19\ h\ 13\ min$



(a) $x - y$ motion during 26 *hours*

(b) $x - z$ motion during 26 *hours*

(c) $y - z$ motion during 26 *hours*

(d) $x - y$ motion during 11 *days* 19 *hours*

(e) $x - z$ motion during 11 *days* 19 *hours*

(f) $y - z$ motion during 11 *days* 19 *hours*

(g) $x - y$ motion during 21 *days* 11 *hours*

(h) $x - z$ motion during 21 *days* 11 *hours*

(i) $y - z$ motion during 21 *days* 11 *hours*

Figure 8.12: Trajectory associated with the minimum $\theta = 2.4006615\ rad$, $t_f = 10\ d\ 19\ h\ 13\ min$

want to place them on stable trajectories that do not collide.

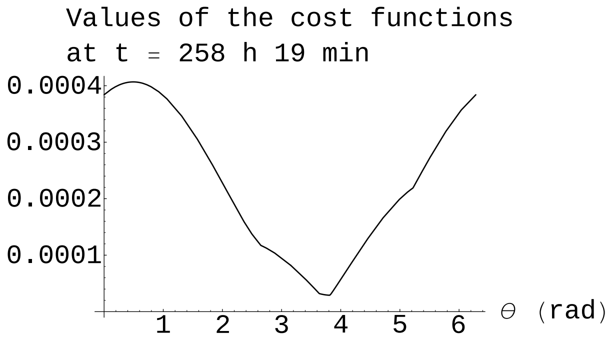
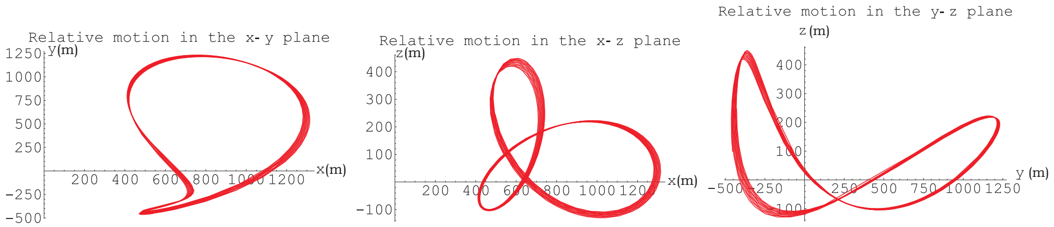


Figure 8.13: Cost function as a function of θ for $t_f = 10\ d\ 18\ h\ 19\ m$



(a) $x - y$ motion during 26 $hours$

(b) $x - z$ motion during 26 $hours$

(c) $y - z$ motion during 26 $hours$

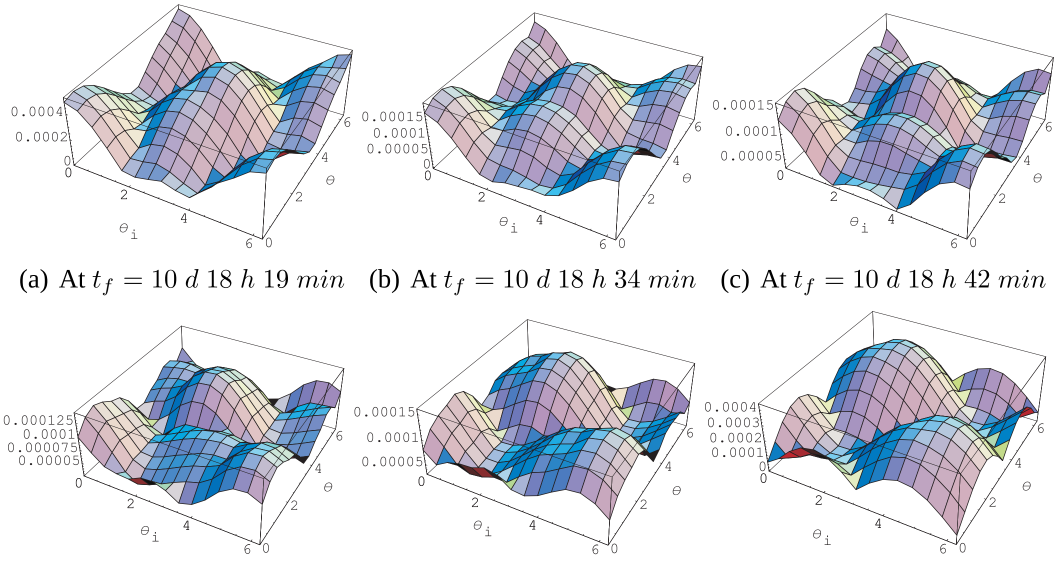
Figure 8.14: Trajectory associated with the minimum
$\theta = 3.814575\ rad$, $t_f = 10\ d\ 18\ h\ 19\ min$

Furthermore, larger or smaller values of t_f could have been chosen, however we must be aware that if t_f is too small, short term oscillations may be as large as the drift and in that case the cost function does not discriminate well; its minima do not necessarily correspond to stable trajectories. On the other hand, if t_f is very large, the minima correspond to (M, T)-stable relative trajectories with T increasing as t_f increases.

Finally, in the above example we selected trajectories that correspond to minima of J and let t_f vary to find several stable trajectories. However, trajectories that correspond to values of J close to the minimum may be stable trajectories as well. If we vary t_f, say from T^1 to T^2, we notice that the trajectory corresponding to the minimum of J at T^1 is different from the one corresponding to the minimum of J at T^2. Although the trajectory associated to T^1 does not correspond to a minimum of J at T^2, it is stable and therefore corresponds to a small value of J at T^2. As a result, we are able to identify regions in which there are no stable trajectories that go through an initial position q_0 and through the circle of radius r_f at t_f. For example, all stable trajectories that go through $q_0 = (495, -428.6, 247.5)\ m$ and $q_f = 700\cos(\theta_f)\hat{f}_1 + 700\sin(\theta_f)\hat{f}_2\ m$ at t_f are roughly localized on the arc defined by $\theta_f \in [0, \pi]$ when $t_f = 10\ d\ 19\ h\ 13\ min$ (Fig. 8.10) and by $\theta_f \in [2, 5]\ rad$ when $t_f = 10\ d\ 18\ h\ 19\ min$ (Fig. 8.13).

8.2.4 Stable configurations

In this section, we generalize the approach introduced above in order to design stable configurations. Without loss of generality, and for sake of simplicity, we assume that the formation is on $\mathcal{C}_{t_0}^{r_0}$ at the initial time so that the positions of the spacecraft are determined by the angle θ_0, the angle between $\hat{f}_1$ and the local relative position vector. As a result, the initial position may be regarded as a function of θ_0. Thus, Eqns. (8.7) and (8.8) provide a polynomial approximation of p_0 and p_f in the variables (θ_0, θ_f) (instead of θ_f only) with time-dependent coefficients. The procedure to find stable trajectories is the same as before but now we have an additional variable, θ_0. In Fig. 8.15 we represent the values of the cost function as a function of θ_i and θ_f for different times. We notice that if two out of the three variables $(\theta_f, \theta_0, t_f)$ are given, there exists a value of the third variable that minimizes the cost function. In other words, whatever θ_0 and t_f are, there exists a stable trajectory that goes through the initial position at the initial time and reaches $\mathcal{C}_{t_f}^{r_f}$ in t_f units of time. Moreover, if t_f varies, minima of the cost function correspond to different stable trajectories due to short term oscillations.



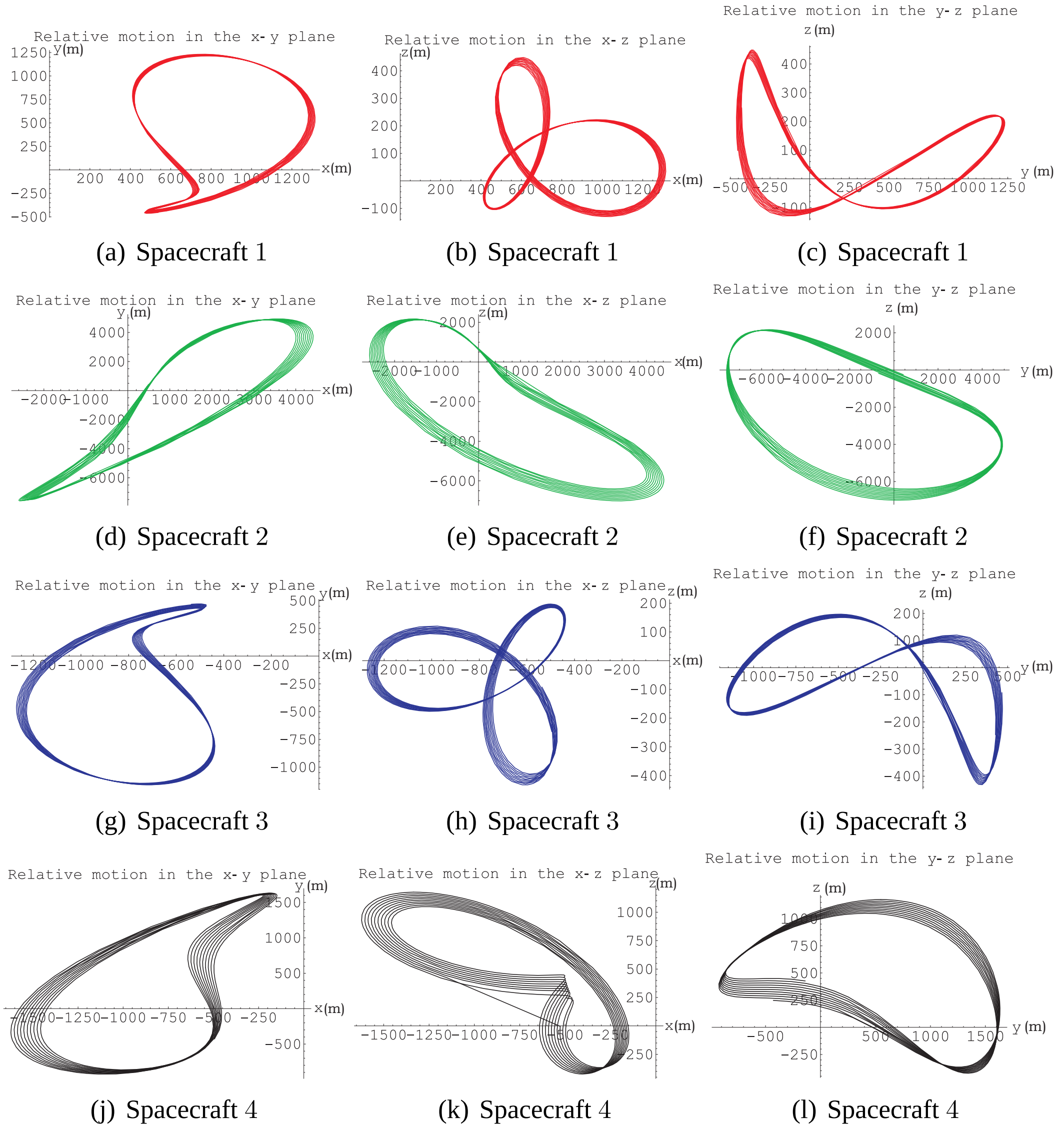
(a) At $t_f = 10\ d\ 18\ h\ 19\ min$ (b) At $t_f = 10\ d\ 18\ h\ 34\ min$ (c) At $t_f = 10\ d\ 18\ h\ 42\ min$

(d) At $t_f = 10\ d\ 18\ h\ 50\ min$ (e) At $t_f = 10\ d\ 18\ h\ 57\ min$ (f) At $t_f = 10\ d\ 19\ h\ 13\ min$

Figure 8.15: Cost function as a function of the initial and final positions for several t_f

Example

We consider a formation of four spacecraft equally spaced on a circle of radius $700\ m$ about the reference trajectory that lies in the local vertical plane at the initial time. Spacecraft k has its initial position defined by $\theta_i = \pi/4 + k\pi/2\,,\ k \in [0,3]$. Stable trajectories may be found by minimizing the cost function with respect to θ. For every choice of t_f there is a solution to the minimization problem (see Fig. 8.15). As a result, we are able to find infinitely many stable trajectories for each spacecraft. In Fig. 8.16 we plot the trajectories of the four spacecraft that are found by considering $t_f = 10\ d\ 18\ h\ 19\ m$ and in Fig. 8.17, $t_f = 10\ d\ 19\ h\ 13\ m$. The two solutions have very different properties; Even though the positions at the final time t_f are constrained to be at $700\ m$ from the reference trajectory in the local vertical plane, the relative distance may be large at intermediary times. For instance the solution found for $t_f = 10\ d\ 18\ h\ 19\ m$ yields a formation that is as large as $6\ km$. Such trajectories cannot be found using linear approximations of the relative motion.



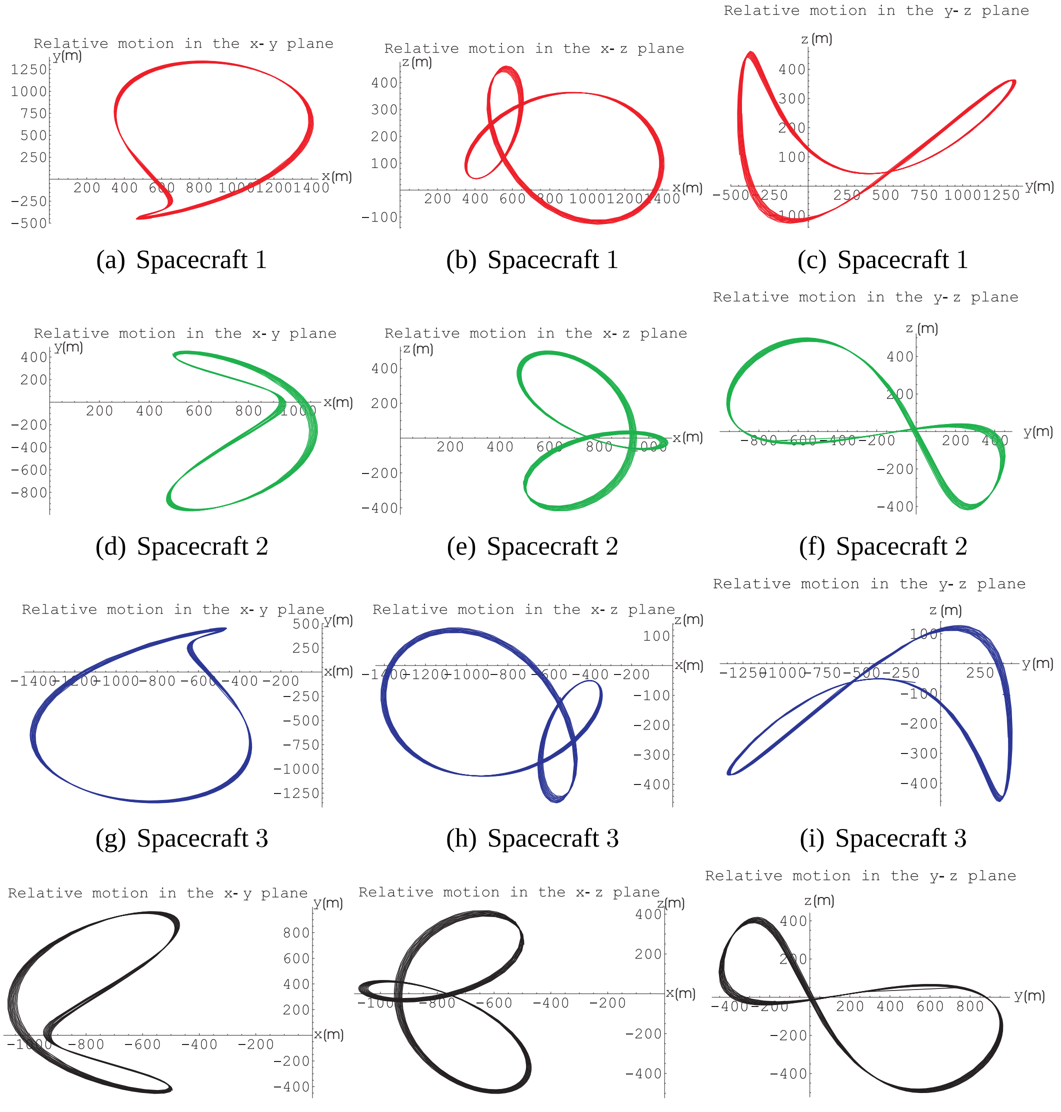
(a) Spacecraft 1 (b) Spacecraft 1 (c) Spacecraft 1

(d) Spacecraft 2 (e) Spacecraft 2 (f) Spacecraft 2

(g) Spacecraft 3 (h) Spacecraft 3 (i) Spacecraft 3

(j) Spacecraft 4 (k) Spacecraft 4 (l) Spacecraft 4

Figure 8.16: Trajectories of the four spacecraft for $t_f = 10\ d\ 18\ h\ 19\ m$



(a) Spacecraft 1

(b) Spacecraft 1

(c) Spacecraft 1

(d) Spacecraft 2

(e) Spacecraft 2

(f) Spacecraft 2

(g) Spacecraft 3

(h) Spacecraft 3

(i) Spacecraft 3

(j) Spacecraft 4

(k) Spacecraft 4

(l) Spacecraft 4

Figure 8.17: Trajectories of the four spacecraft for $t_f = 10\ d\ 19\ h\ 13\ m$

CHAPTER IX

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

9.1 Summary and contribution of the thesis

Starting from the observation that new methods are needed to address complex problems arising in spacecraft formation design and control, we have developed a novel approach for solving Hamiltonian two-point boundary value problems. The theoretical aspects of our approach make contributions to several fields, shed light on the properties of two-point boundary value problems and have found new results. The numerics of our approach is also very rich, and our study of it has led us to investigate and make contributions to the field of variational integrators. Finally, we presented several applications of our method. In particular, it allows us to develop innovative solution procedures to address difficult problems arising in a wide range of fields.

9.1.1 Theoretical aspects

The method we develop in this thesis is based on the Hamilton-Jacobi theory. We have observed that the generating functions associated with the phase flow readily solve any Hamiltonian two-point boundary value problem. This observation, that we believe no one has made before, has many consequences that we now re-state. Above all, it provides a very general methodology for solving boundary value problems for Hamiltonian systems.

Whereas traditional methods solve boundary value problems about an initial guess only, our approach gives a “full picture”. In particular, traditional methods completely ignore the number of solutions to the boundary value problem. Our approach, however, indicates the presence of multiple solutions as singularities of generating functions. In turn, we proved and illustrated that these singularities can be studied and the number of solutions may be determined.

In linear systems theory, it is well-known that perturbation matrices solve boundary value problems. These matrices have distinctive properties that are studied in the literature. Using generating functions we have recovered and extended some of these properties. Most importantly, we have proved that they correspond to coefficients of the generating functions. As a result, our approach naturally contains the theory of perturbation matrices. The relation between perturbation matrices and generating functions may also be investigated using the state transition matrix. In this respect, we have shown that the state transition matrix and generating functions are closely related. One of the main consequences of this allows us to predict singularities of the generating functions using the state transition matrix. This result broadens to nonlinear systems with polynomial Hamiltonian function.

In nonlinear systems theory, there is no equivalent of the perturbation matrices. Thus, the approach we have proposed is the first to define functions, namely the generating functions, that describe the phase flow as two-point boundary value problems. Obviously, no results as general as the ones derived for linear systems may be gleaned in this case. However, for polynomial generating functions we have established that singularities of the generating functions may still be predicted from the state transition matrix. As a result, the existence of multiple solutions to two-point boundary value problems is fully predicted by the linear dynamics. The number of solutions, however, depends on the nonlinear dynamics.

9.1.2 Numerics

To demonstrate the efficiency of our novel approach, we have proposed a robust algorithm to compute the generating functions. By combining two techniques (called direct and indirect), our algorithm allows one to approximate the generating functions locally in space and globally in time. We now briefly review its main characteristics:

- It handles initial conditions specified in terms of functions with parameters.
- It applies to any Hamiltonian system, independent of its complexity.
- It avoids or bypasses singularities.
- We believe (but have not proven yet) that the indirect method preserves the symplectic two-form if one uses a symplectic integrator. If a boundary value problem with long transfer time needs to be solved, this property is very valuable as long-term behaviors of nonlinear systems are better simulated by symplectic algorithms.
- The software is freely available upon request, from Daniel Scheeres and myself.

The necessity of using a symplectic algorithm in the indirect approach has led us to investigate geometric integrators. The research we have pursued in that direction went far beyond our objectives and contributed to advances in the field of variational integrators.

Specifically, we have presented a general framework to study discrete systems. We have introduced variational principles on the tangent and cotangent bundles that are the discrete counterpart of the known principles of critical action for Lagrangian and Hamiltonian dynamical systems. Our formulation has several important differences with previous works. One of its main advantages is its ability to work with both Lagrangian and Hamiltonian systems. Most of the work in the literature focuses on the Lagrangian point of view, and defines a discrete Legendre transformation to map the tangent bundle to the co-tangent bundle. In this manner, Hamiltonian systems that are also Lagrangian may be studied. However, this approach fails if the Hamiltonian system is not Lagrangian. As illustrated in Chapter IV, this particular case is often encountered, especially in optimal control theory. Furthermore, we have shown that our approach allows us to recover most of the classical symplectic algorithms. By increasing the dimensionality of the configuration space, it can also yield symplectic-energy conserving algorithms. When time is a generalized coordinate, the dynamical system is subject to an energy constraint, and we are able to adapt our variational principles to take such a constraint into account. In the same manner, our approach may be modified to derive symplectic algorithms to integrate non-autonomous dynamical systems with (non-holonomic) constraints.

In addition, we have given a discrete symplectic structure to the discrete phase space. For the first time, we have been able to extend the notions of symplectic two-form, canonical transformation and generating function to discrete settings. Once all these notions were introduced, we were able to develop a discrete Hamilton-Jacobi theory. This theory allows us to estimate the energy error in the integration using different set of coordinates related by discrete canonical transformations.

Finally, we have extended the above framework to optimal control problems and developed a unified theory to solve optimal control problems using symplectic integrators. Specifically, we have introduced a discrete maximum principle that yields discrete necessary conditions for optimality. These conditions are in agreement with the ones derived from the Pontryagin maximum principle and define symplectic integrators.

9.1.3 Applications

The approach we have presented to solve two-point boundary value problems applies to any Hamiltonian system. It is therefore not surprising that it has implications in several fields. In particular, it allows us to develop new solution procedures to study the phase space structure, solve optimal control problems and design spacecraft formations. These methods are all based on two important aspects of the present research:

1. Once the generating functions are known, we can solve any two-point boundary value problem at the cost of a single function evaluation; no initial guesses or iterations are required.

2. Using the algorithm we developed in Chapter V, we obtain a closed-form solution to two-point boundary value problems.

Spacecraft formation dynamics and design The first motivation for the present research was to address complex problems arising in spacecraft formation flight. We believe that the method we propose meets our expectations and objectives. Despite a complex dynamical model and an arbitrary reference trajectory, we have been able to obtain a semi-analytic description of the nonlinear relative phase flow as solutions to two-point boundary value problems. This representation allowed us to design two extremely difficult missions with little effort. Our approach, however, is not limited to these two missions and we recall its main features:

- The dynamical environment may be as complex as one wants, the only constraint being that the dynamical system must be Hamiltonian. In addition, the complexity of the dynamical system does not seriously impact the computation time.

- The reference trajectory may be arbitrary.

- The time span we consider may be very large, the larger it is the longer the ordinary differential equations obtained with the indirect algorithm should be integrated. The main advantage of describing the phase flow as two-point boundary value problems is that the time period we consider does not influence the accuracy of the results. This aspect is of major importance, especially as this is a weakness of traditional approaches based on the initial value problem.

- Our approach also allows one to deal with low-thrust spacecraft. In this case, the reconfiguration problem can be formulated as an optimal control problem whose necessary conditions for optimality are a Hamiltonian two-point boundary value problem. For these problems, the dynamical environment may not be Hamiltonian since the necessary conditions for optimality yield a Hamiltonian system. However, it should be emphasized that the dimensionality is double (because of the adjoint variables).

- There are no limitation on the complexity of the formation geometry in the reconfiguration problem as long as the geometry can be described with constraints on (q, p) only.

- From the semi-analytic expression of the generating functions, several problems may be addressed. We have seen how to solve the reconfiguration problem and the deployment problem, we have also been able to find stable configurations. One can readily apply Chapter VII to find periodic configurations.

Phase space structure By posing the search for periodic orbits as a two-point boundary value problem with constraints, we have reduced the search for periodic orbits to solving a few nonlinear equations. Through several examples, we have shown that our method recovers known periodic orbits and thus, captures the nonlinear dynamics. Compared to traditional methods, the technique we propose does not require initial guesses and/or iterations. It characterizes periodic orbits as solutions of nonlinear algebraic equations.

Optimal control theory Finally, we have proved that the method we developed for solving two-point boundary value problems has major implications in optimal control theory. Not only does it allow one to solve the necessary conditions for optimality, but it also overcomes barriers to truly reconfigurable control. Using traditional techniques, the optimal control law needs to be re-calculated as the boundary conditions and targets for the system change. Using the generating functions we have shown that if the boundary conditions change in values, the resulting optimal control law may be found instantaneously. Further, if the nature of the boundary conditions change, then we need to perform a Legendre transformation (i.e., a series of algebraic manipulations) to compute the new control law. These properties are specific to our approach and cannot be found in any other nonlinear methods.

9.2 Limitations and suggestions for further research

We now discuss the limitations of the present research and propose some ideas for future research.

9.2.1 On solving two-point boundary value problems

Computing the generating functions remains the main hurdle to successfully applying our work to any problem. The algorithm we present applies to polynomial Hamiltonian systems only. This is a severe restriction as it prevents us from solving the Hamilton-Jacobi equation for both non-polynomial Hamiltonian functions over a large spatial domain (we are restricted to study relative motion only) and non-analytic Hamiltonian functions. This latter case arises in optimal control problems involving control constraints or non-analytic cost functions, for instance.

In addition, some systems may be singular over a finite time span. For instance, in the three-dimensional two-body problem, transfers with the two radius vectors anti paralleled to each other have multiple solutions for all transfer times. Thus, we expect F_1 to be singular at all times in this geometry. Similarly, the Heisenberg optimal control problem yields a singular F_1 generating function. For such problems, certain classes of two-point boundary value problems can never be solved as the corresponding generating functions are always singular. This topic is still to be explored. To remove the singularities, one might need to consider generating functions with fewer variables (keeping only the independent ones), but subject to some constraints involving the missing variables.

9.2.2 In optimal control theory

A major implication of our work is in the field of optimal control. Using the generating functions, we can solve the Pontryagin necessary conditions for optimality for a large class of optimal control problems. However, if the cost function is not analytic, or if there exist control constraints, then the method we present must be altered. For instance, for time-optimal control problems, since we know that the control is either at its upper or lower bound, we can solve the Hamilton-Jacobi equation for both cases and then find the times at which the control shifts.

Furthermore, we previously pointed out that the Hamilton-Jacobi equation reduces to a set of matrix ordinary differential equations, one of them being a Riccati equation. We believe that the connection between the Hamilton-Jacobi equation and the Riccati equation is deeper. Its understanding could yield insights into nonlinear control theory. Sakamoto [83] started to analyze this link. In particular, he generalized properties of the Riccati equation to the Hamilton-Jacobi equation.

Finally, it should not surprise one that generating functions and cost functions are related, as they both verify similar equations and solve the optimal control problem. Park and Scheeres [76] have started to investigate this connection. They proved that “the cost function is related to a special kind of generating function, and that the optimal feedback control problem can be considered as part of a more comprehensive field of canonical transformations for Hamiltonian systems” (Park and Scheeres in [76]).

9.2.3 Variational integrators

We have presented a general framework for studying the discretization of certain dynamical systems. We believe that this framework may be extended to spacetime discretization. This would open the doors to variational principles for multi-symplectic algorithms. Such algorithms would allow one to develop efficient numerical techniques for simulating the motion of rigid bodies and complex interconnected systems, for instance.

In addition, the discrete maximum principle we have developed yields discrete necessary conditions for optimality under some smoothness conditions. Pontryagin’s maximum principle applies under far less severe regularity conditions. Its discrete counterpart has been studied by Jordan and Polak [55] for instance. However, the obtained discrete necessary conditions for optimality do not define a symplectic algorithm. It is not clear yet how one can remove the smoothness conditions in discrete settings while preserving the geometric features of the necessary conditions.

APPENDICES

APPENDIX A

THE DYNAMICS OF RELATIVE MOTION

In this appendix, we show that the dynamics of the relative motion of two particles in a Hamiltonian vector field is Hamiltonian.

Consider a Hamiltonian system with Hamiltonian function $H(q,p,t)$. Let (q_0^0, p_0^0) and (q_0^1, p_0^1) be two points in phase space such that:

$$q_0^1 \;=\; q_0^0 + \Delta q_0 \,, \tag{A.1}$$

$$p_0^1 \;=\; p_0^0 + \Delta p_0 \,, \tag{A.2}$$

where $(\Delta q_0, \Delta p_0)$ is small enough to guaranty the convergence of the Taylor series in Eq. (A.8). We denote by (q^i, p^i) the trajectory with initial conditions (q_0^i, p_0^i), i.e.,

$$q^1 = q(q_0^1, p_0^1, t)\,,\, p^1 = p(q_0^1, p_0^1, t)\,, \tag{A.3}$$

$$q^0 = q(q_0^0, p_0^0, t)\,,\, p^0 = p(q_0^0, p_0^0, t)\,. \tag{A.4}$$

and we define $X^h = \begin{pmatrix} \Delta q \\ \Delta p \end{pmatrix}$ the relative state vector by:

$$X^1 = X^0 + X^h \,, \tag{A.5}$$

where $X^i = \begin{pmatrix} q^i \\ p^i \end{pmatrix}$. For convenience we shall call (q^0, p^0) the reference trajectory and (q^1, p^1) the displaced trajectory .

Both trajectories verify the Hamilton equations of motion:

$$\dot{X}^i = J\nabla H^i \,, \tag{A.6}$$

where $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ and $\nabla H^i = \begin{pmatrix} \frac{\partial H}{\partial q} \\ \frac{\partial H}{\partial p} \end{pmatrix} (q^i, p^i, t)$. Using our previous notation, Eq. (A.6) reads, for $i = 1$:

$$\dot{X}^0 + \dot{X}^h = J\nabla H^1 \,. \tag{A.7}$$

We expand the right hand side of Eq. (A.7) about the nominal trajectory X^0, assuming $(\Delta q, \Delta p)$ small enough for convergence of the series:

$$\nabla H(q^1, p^1, t) = \nabla H(q^0, p^0, t) + \begin{pmatrix} \frac{\partial^2 H}{\partial q^2}(q^0, p^0, t)\Delta q + \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t)\Delta p \\ \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t)\Delta q + \frac{\partial^2 H}{\partial p^2}(q^0, p^0, t)\Delta p \end{pmatrix} + \cdots \tag{A.8}$$

Substituting this into Eq. (A.7) yields

$$\dot{X}^0 + \dot{X}^h = J\nabla H^0 + J \begin{pmatrix} \frac{\partial^2 H}{\partial q^2}(q^0, p^0, t)\Delta q + \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t)\Delta p \\ \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t)\Delta q + \frac{\partial^2 H}{\partial p^2}(q^0, p^0, t)\Delta p \end{pmatrix} + \cdots \tag{A.9}$$

Using equation (A.6), Eq. (A.9) simplifies to:

$$\dot{X}^h = J \begin{pmatrix} \frac{\partial^2 H}{\partial q^2}(q^0, p^0, t)\Delta q + \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t)\Delta p \\ \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t)\Delta q + \frac{\partial^2 H}{\partial p^2}(q^0, p^0, t)\Delta p \end{pmatrix} + \cdots \tag{A.10}$$

Therefore, the dynamics describing the relative motion of two particles in a Hamiltonian vector field is Hamiltonian if and only if there exists an Hamiltonian function H^h such that Eq. (A.10) can be written as Hamilton's equations. Let

$$H^h(X^h, t) = \frac{1}{2} X^h \begin{pmatrix} \frac{\partial^2 H}{\partial q^2}(q^0, p^0, t) & \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t) \\ \frac{\partial^2 H}{\partial q \partial p}(q^0, p^0, t) & \frac{\partial^2 H}{\partial q^2}(q^0, p^0, t) \end{pmatrix} X^h + \cdots \tag{A.11}$$

We can check that:

$$\dot{X}^h = J\nabla H^h(X^h, t) \,. \tag{A.12}$$

Without ignoring higher order terms, the expansion of the right hand side of Eq. (A.7) yields:

$$H^h(X^h,t) = \\ \sum_{p=2}^{\infty} \sum_{\substack{i_1,\cdots,i_{2n}=0\\ i_1+\cdots+i_{2n}=p}}^{p} \frac{1}{i_1!\cdots i_{2n}!} \frac{\partial^p H}{\partial q_1^{i_1}\cdots\partial q_n^{i_n}\partial p_1^{i_{n+1}}\cdots\partial p_n^{i_{2n}}}(q^0,p^0,t) {X_1^h}^{i_1}\ldots {X_{2n}^h}^{i_{2n}}\,. \tag{A.13}$$

Thus, the dynamics of a particle relative to a known trajectory is Hamiltonian with a Hamiltonian function $H^h(X^h,t) = H^h(\Delta q, \Delta p, t)$. The coefficients of the Taylor series $\frac{1}{i!j!}\frac{\partial^{i+j}H}{\partial q^i \partial p^j}(q^0,p^0,t)$ are time varying quantities and are easily evaluated for any Hamiltonian once the reference trajectory is known.

APPENDIX B

THE HAMILTON-JACOBI EQUATION AT HIGHER ORDERS

In this appendix, we give an explicit expression of P as defined by Eq. (5.7). This generalizes the approach developed in Section 3.2.2 for linear systems to nonlinear systems.

We assume a $2n$-dimensional Hamiltonian system with polynomial Hamiltonian function and polynomial generating functions. We have seen in Chapter V that the Hamilton-Jacobi partial differential equation reduces to an ordinary differential equation of the form

$$P(y, f^{p,r}_{i_1,\cdots,i_{2n}}(t), \dot{f}^{p,r}_{i_1,\cdots,i_{2n}}(t)) = 0\,. \tag{B.1}$$

In the following we use tensor notation in order to derive an explicit expression of P. In tensor notation, a Taylor series expansion writes as:

$$f(x,t) = f^0(t) + f^1(t)\cdot\vec{x} + (f^2(t)\cdot\vec{x})\cdot\vec{x} + ((f^3(t)\cdot\vec{x})\cdot\vec{x})\cdot\vec{x} + \cdots\,. \tag{B.2}$$

Applying this formula to $H(\vec{x},t)$ and $F_2 = F(\vec{y},t)$ yields:

$$\begin{aligned} H(\vec{x}) &= h_{i,j}(t)x_ix_j + h_{i,j,k}(t)x_ix_jx_k + \cdots\,, && \text{(B.3)}\\ F(\vec{y}) &= f_{i,j}(t)y_iy_j + f_{i,j,k}(t)y_iy_jy_k + \cdots\,. && \text{(B.4)} \end{aligned}$$

where we assume the summation convention. Let us now express $\vec{x} = (\Delta q, \Delta p)$ as a function of $\vec{y} = (\Delta q, \Delta p_0)$ (we drop the time dependence in the notation, i.e., we shall write $h_{i,j}$ instead of $h_{i,j}(t)$). For all $a \leq n$ and $j = n + a$

$$\begin{aligned} x_a &= y_a \,, && \text{(B.5)} \\ x_j &= \frac{\partial F}{\partial y_a} && \text{(B.6)} \\ &= f_{a,k} y_k + f_{k,a} y_k + f_{a,k,l} y_k y_l + f_{k,a,l} y_k y_l + f_{k,l,a} y_k y_l + \cdots \,, && \text{(B.7)} \end{aligned}$$

where n is the dimension of the configuration space. The Hamilton-Jacobi equation becomes:

$$\dot{f}_{i,j} y_i y_j + \dot{f}_{i,j,k} y_i y_j y_k + \cdots + h_{i,j} x_i x_j + h_{i,j,k} x_i x_j x_k + \cdots = 0 \,. \qquad \text{(B.8)}$$

Replacing $\vec{x}$ by $\vec{y}$ in Eq. (B.8) using Eq. (B.7), and keeping only terms of order less than 3 yields:

$$\begin{aligned} 0 = {} & \dot{f}_{i,j} y_i y_j + \dot{f}_{i,j,k} y_i y_j y_k + h_{a,b} y_a y_b + h_{a,b,c} y_a y_b y_c \\ & + (h_{a,n+b} + h_{n+b,a}) y_a (f_{b,k} y_k + f_{k,b} y_k + f_{b,k,l} y_k y_l + f_{l,b,k} y_k y_l + f_{k,l,b} y_k y_l) \\ & + h_{n+a,n+b} (f_{a,k} y_k + f_{k,a} y_k + f_{a,k,l} y_k y_l + f_{l,a,k} y_k y_l + f_{k,l,a} y_k y_l) \\ & (f_{b,m} y_m + f_{m,b} y_m + f_{b,m,p} y_m y_p + f_{p,b,m} y_m y_p + f_{m,p,b} y_m y_p) \\ & + (h_{n+a,b,c} + h_{c,n+a,b} + h_{b,c,n+a}) y_b y_c (f_{a,k} y_k + f_{k,a} y_k) \\ & + (h_{n+a,n+b,c} + h_{n+b,c,n+a} + h_{c,n+a,n+b}) y_c (f_{a,k} y_k + f_{k,a} y_k)(f_{b,l} y_l + f_{l,b} y_l) \\ & \qquad + h_{n+a,n+b,n+c} (f_{a,k} y_k + f_{k,a} y_k)(f_{b,l} y_l + f_{l,b} y_l)(f_{c,m} y_m + f_{m,c} y_m) \,. \qquad \text{(B.9)} \end{aligned}$$

Eq. (B.9) is the expression of P up to order 3 as defined by Eq. (5.7). It is a polynomial equation in the y_i variables with time dependent coefficients and holds if every coefficient is zero. We notice that the equations of order 2 (the one obtained by setting the coefficients of $y_i y_j$ to zero) are the same as the ones found previously in Section 3.2.2. The equations of order 3 reads:

$$\dot{f}_{i,j,k}y_iy_jy_k + (A_{i,j,k} + B_{i,j,k} + C_{i,j,k})y_iy_jy_k + (D_{a,i,j} + E_{a,i,j})y_ay_iy_j$$
$$+ G_{a,b,i}y_ay_by_i + h_{a,b,c}y_ay_by_c = 0\,, \tag{B.10}$$

where

$$A_{i,j,k} = h_{n+a,n+b,n+c}(f_{a,i} + f_{i,a})(f_{b,j} + f_{j,b})(f_{c,k} + f_{k,c})\,,$$
$$B_{i,j,k} = h_{n+a,n+b}(f_{a,i} + f_{i,a})(f_{b,j,k} + f_{j,k,b} + f_{k,b,j})\,,$$
$$C_{i,j,k} = h_{n+a,n+b}(f_{b,i} + f_{i,b})(f_{a,j,k} + f_{j,k,a} + f_{k,a,j})\,,$$
$$D_{a,i,j} = (h_{a,n+b,n+c} + h_{n+c,a,n+b} + h_{n+b,n+c,a})(f_{b,i} + f_{i,b})(f_{c,j} + f_{j,c})\,,$$
$$E_{a,i,j} = (h_{a,n+b} + h_{n+b,a})(f_{b,i,j} + f_{j,b,i} + f_{i,j,b})\,,$$
$$G_{a,b,i} = (h_{a,b,n+c} + h_{b,n+c,a} + h_{n+c,a,b})(f_{c,i} + f_{i,c})\,. \tag{B.11}$$

We deduce the coefficients of $y_iy_jy_k$:

- Coefficients of $y^3_{i\leq n}$

$$A_{i,i,i} + B_{i,i,i} + C_{i,i,i} + D_{i,i,i} + E_{i,i,i} + \dot{f}_{i,i,i} + G_{i,i,i} + h_{i,i,i} = 0\,. \tag{B.12}$$

- Coefficients of $y^3_{i>n}$

$$A_{i,i,i} + B_{i,i,i} + C_{i,i,i} + \dot{f}_{i,i,i} = 0\,. \tag{B.13}$$

- Coefficients of $y^2_{i\leq n}y_{j\leq n}$

$$(A + B + C + D + E + \dot{f} + G + h)_{\tau(i,i,j)} = 0\,. \tag{B.14}$$

where $\tau(i,j,k)$ represents all the distinct permutations of (i,j,k), that is

$$A_{\tau(i,j,k),l} = A_{i,j,k,l} + A_{i,k,j,l} + A_{k,i,j,l} + A_{k,j,i,l} + A_{j,k,i,l} + A_{j,i,k,l}$$

but

$$A_{\tau(i,i,j),l} = A_{i,i,j,l} + A_{i,j,i,l} + A_{j,i,i,l}\,.$$

- Coefficients of $y_{i\leq n}^2 y_{j>n}$

$$(A+B+C+\dot{f})_{\tau(i,i,j)} + (D+E)_{i,\tau(i,j)} + G_{i,i,j} = 0\,. \tag{B.15}$$

- Coefficients of $y_{i\leq n} y_{j\leq n} y_{k\leq n}$:

$$(A+B+C+D+E+\dot{f}+G+h)_{\tau(i,j,k)} = 0\,. \tag{B.16}$$

- Coefficients of $y_{i\leq n} y_{j\leq n} y_{k>n}$

$$(A+B+C+\dot{f})_{\tau(i,j,k)} + (D+E)_{i,\tau(j,k)} + (D+E)_{j,\tau(i,k)} + G_{\tau(i,j),k} = 0\,. \tag{B.17}$$

- Coefficients of $y_{i>n}^2 y_{j\leq n}$

$$(A+B+C+\dot{f})_{\tau(i,i,j)} + (E+D)_{j,i,i} = 0\,. \tag{B.18}$$

- Coefficients of $y_{i>n}^2 y_{j>n}$

$$(A+B+C+\dot{f})_{\tau(i,i,j)} = 0\,. \tag{B.19}$$

- Coefficients of $y_{i\leq n} y_{j>n} y_{k>n}$

$$(A+B+C+\dot{f})_{\tau(i,j,k)} + (D+E)_{i,\tau(j,k)} = 0\,. \tag{B.20}$$

- Coefficients of $y_{i>n} y_{j>n} y_{k>n}$

$$(A+B+C+\dot{f})_{\tau(i,j,k)} = 0\,. \tag{B.21}$$

Eqns. (B.12)-(B.21) allow us to solve for F_2 (and F_1 since they both verify the same Hamilton-Jacobi equation, only the initial conditions being different). The process of deriving equations for the generating functions can be continued to arbitrarily high order using a symbolic manipulation program.

APPENDIX C

THE HILL THREE-BODY PROBLEM

The three-body problem describes the motion of three point mass particles under their mutual gravitational interactions. This is a classical problem that covers a large range of situations in astrodynamics. An instance of such situations is the motion of the Moon about the Earth under the influence of the Sun. However, this problem does not have a general solution and thus we usually consider simplified formulations justified by physical reasoning. In this dissertation we consider three simplifications that yield two different models:

1. The circular restricted three-body problem: If one of the three bodies has negligible mass compared to the other two bodies (for instance a spacecraft under the influence of the Sun and the Earth), it is rather obvious that its gravitational attraction has very little effect on the motion of the other bodies. Ignoring the mass of this smaller body yields the restricted three-body problem. If in addition one of the two massive bodies is in a circular orbit about the other one, then we obtain the circular restricted three-body problem [7].

2. The Hill three-body problem: The Hill three-body problem can naturally be derived from the circular restricted three-body problem by assuming that one of the two

massive bodies has larger mass than the other one (the Sun compared to the Earth for instance).

C.1 The circular restricted three-body problem

We consider the planar motion of a massless body (a spacecraft for instance) under the influence of two massive bodies in circular orbit about each other. In the coordinate system centered at the center of mass of the two massive bodies the Hamiltonian function describing the dynamics of the massless body is:

$$\begin{aligned} &H(q_x, q_y, p_x, p_y) = \\ &\quad \frac{1}{2}(p_x^2 + p_y^2) + p_x q_y - q_x p_y - \frac{1-\mu}{\sqrt{(q_x+\mu)^2 + q_y^2}} - \frac{\mu}{\sqrt{(q_x - 1 + \mu)^2 + q_y^2}}\,, \end{aligned} \tag{C.1}$$

where $\mu = \frac{M_2}{M_1+M_2}$, M_1 and M_2 are the mass of the two bodies with $M_1 > M_2$, $q_x = x$, $q_y = y$, $p_x = \dot{x} - y$ and $p_y = \dot{y} + x$. In the above formulation we use normalized quantities, distances are normalized with respect to the two massive bodies relative distance and the time scale is such that the orbit period of one massive body with respect to the other one is 2π. Then Hamilton's equations of motion read:

$$\left\{ \begin{array}{rcl} \dot{q}_x & = & p_x + q_y \\ \dot{q}_y & = & p_y - q_x \\ \dot{p}_x & = & p_y - (1-\mu)\frac{q_x+\mu}{((q_x+\mu)^2+q_y^2)^{3/2}} - \mu\frac{x-1+\mu}{((q_x-1+\mu)^2+q_y^2)^{3/2}} \\ \dot{p}_y & = & -p_x - (1-\mu)\frac{q_y}{((q_x+\mu)^2+q_y^2)^{3/2}} - \mu\frac{y}{((q_x-1+\mu)^2+q_y^2)^{3/2}} \end{array} \right. \tag{C.2}$$

There are five equilibrium points for this system, called the Libration points.

C.2 The Hill three-body problem

If one body has a larger mass than the other one, we can expand the equations of motion about $\mu = 0$. Then, shifting the coordinate system so that its center is the body with mass

$M_2 << M_1$ yields Hill's formulation of the three-body problem. The Hamiltonian for this system reads:

$$H(q,p) = \frac{1}{2}(p_x^2 + p_y^2) + (q_y p_x - q_x p_y) - \frac{1}{\sqrt{q_x^2 + q_y^2}} + \frac{1}{2}(q_y^2 - 2q_x^2)\,, \tag{C.3}$$

and the equations of motion become:

$$\begin{cases} \dot{q}_x &= p_x + q_y\,, \\ \dot{q}_y &= p_y - q_x\,, \\ \dot{p}_x &= p_y + 2q_x - \frac{q_x}{(q_x^2+q_y^2)^{3/2}}\,, \\ \dot{p}_y &= -p_x - q_y - \frac{q_y}{(q_x^2+q_y^2)^{3/2}}\,. \end{cases} \tag{C.4}$$

Among the 5 equilibrium points identified in the circular three-body problem, only two survive in the planar Hill formulation. Their coordinates are

$$L_1(-\left(\frac{1}{3}\right)^{1/3}, 0) \text{ and } L_2(\left(\frac{1}{3}\right)^{1/3}, 0)\,.$$

Using linear systems theory, one can prove that the Libration have a stable, an unstable and two center manifolds (Fig. C.1).

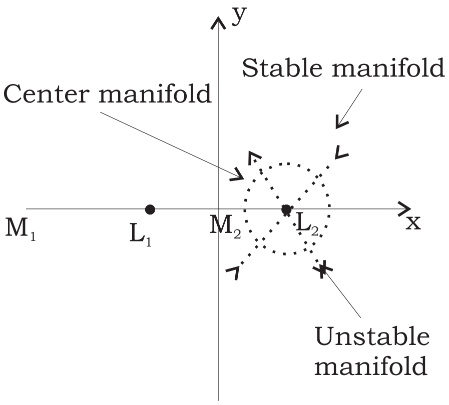


Figure C.1: The Libration points in the Hill three-body problem

To study the relative motion of a spacecraft about L_2, we use Eq. (A.13) to compute

H^h, the Hamiltonian function describing the relative motion dynamics.

$$H^h = \frac{1}{2} X_h^T \begin{pmatrix} H_{qq}(t) & H_{qp}(t) \\ H_{pq}(t) & H_{pp}(t) \end{pmatrix} X_h + \cdots \,, \tag{C.5}$$

where $X_h = \begin{pmatrix} q - q_0 \\ p - p_0 \end{pmatrix} = \begin{pmatrix} \Delta q_x \\ \Delta q_y \\ \Delta p_x \\ \Delta p_y \end{pmatrix}$, $(q_0, p_0) = (\left(\frac{1}{3}\right)^{1/3}, 0, 0, \left(\frac{1}{3}\right)^{1/3})$ refers to the state at the equilibrium point L_2 and,

$$H_{qq}(t) = \begin{pmatrix} \frac{1}{(q_{0x}^2+q_{0y}^2)^{3/2}} - \frac{3q_{0x}^2}{(q_{0x}^2+q_{0y}^2)^{5/2}} - 2 & -\frac{3q_{0x}q_{0y}}{(q_{0x}^2+q_{0y}^2)^{5/2}} \\ -\frac{3q_{0x}q_{0y}}{(q_{0x}^2+q_{0y}^2)^{5/2}} & \frac{1}{(q_{0x}^2+q_{0y}^2)^{3/2}} - \frac{3q_{0y}^2}{(q_{0x}^2+q_{0y}^2)^{5/2}} + 1 \end{pmatrix} , \tag{C.6}$$

$$H_{qp}(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} , \tag{C.7}$$

$$H_{pq}(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} , \tag{C.8}$$

$$H_{pp}(t) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{C.9}$$

Substituting (q_0, p_0) by its value yields the expression of H^h at second order:

$$H^h = \frac{1}{2} \begin{pmatrix} \Delta q_x & \Delta q_y & \Delta p_x & \Delta p_y \end{pmatrix} \begin{pmatrix} -8 & 0 & 0 & -1 \\ 0 & 4 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta q_x \\ \Delta q_y \\ \Delta p_x \\ \Delta p_y \end{pmatrix} . \tag{C.10}$$

At higher order, we find:

$$\begin{aligned} H^h &= \frac{1}{2}\begin{pmatrix}\Delta q_x & \Delta q_y & \Delta p_x & \Delta p_y\end{pmatrix}\begin{pmatrix}-8 & 0 & 0 & -1\\ 0 & 4 & 1 & 0\\ 0 & 1 & 1 & 0\\ -1 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\Delta q_x\\ \Delta q_y\\ \Delta p_x\\ \Delta p_y\end{pmatrix}\\ &+ 3^{4/3}\Delta q_x^3 - \frac{3^{7/3}}{2}\Delta q_x \Delta q_y^2 - 3^{5/3}\Delta q_x^4 + 3^{8/3}\Delta q_x^2\Delta q_y^2 - \frac{3^{8/3}}{8}\Delta q_y^4 \cdots \end{aligned} \quad \text{(C.11)}$$

We point out that H^h is time-independent.

Finally, we give in the following table the values of the normalized variables for the Earth-Sun system.

Normalized units		Earth-Sun system
0.01 unit of length	$\longleftrightarrow$	$21,500\ km$
1 unit of time	$\longleftrightarrow$	$58\ days\ 2\ hours$
1 unit of velocity	$\longleftrightarrow$	$428\ m/s$
1 unit of acceleration	$\longleftrightarrow$	$1.38 \cdot 10^{-5}\ m/s^2$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Ralph Abraham and Jerrold E. Marsden. *Foundations of mechanics*. W. A. Benjamin, 2nd edition, 1978.

[2] K. T. Alfriend and H. Schaub. Dynamics and control of spacecraft formations: challenges and some solutions. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, College Station, Texas. Paper AAS 00-259*. AAS, 2000.

[3] K. T. Alfriend, S. S. Vaddi, and T. A. Lovell. Formation maintenance for low Earth near-circular orbits. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-652*. AAS, 2003.

[4] Vladimir I. Arnold. *Geometrical methods in the theory of ordinary differential equations*. Springer-Verlag, 2nd edition, 1988.

[5] Vladimir I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, 2nd edition, 1988.

[6] Vladimir I. Arnold. *Catastrophe Theory*. Springer-Verlag, 3rd, revised and expanded edition, 1992.

[7] Vladimir I. Arnold, V. V. Kozlov, and A. I. Neishtadt. *Mathematical Aspects of Classical and Celestial Mechanics, Dynamical Systems III*. Springer-Verlag, 1988.

[8] Jean-Pierre Aubin. Boundary-value problems for systems of Hamilton-Jacobi-Bellman inclusions with constraints. *SIAM Journal on Control and Optimization*, 41(2):425–456, 2002.

[9] Giulio Avanzini, Davide Biamonti, and Edmondo A. Minisci. Minimum-fuel/minimum-time maneuvers of formation flying satellites. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-654*. AAS, 2003.

[10] Richard H. Battin. *An Introduction to the Mathematics and Methods of Astrodynamics*. American Institute of Aeronautics and Astronautics, revised edition, 1999.

[11] Jean-David Benamou. Direct computation of multivalued phase space solutions for Hamilton-Jacobi equations. *Communications on pure and applied mathematics*, 52(11):1443–1475, 1999.

[12] John T. Betts. Survey of numerical methods for trajectory optimization. *Journal of control, guidance, and dynamics*, 21(2):193–207, 1998.

[13] Gilbert Ames Bliss. *Lectures on the Calculus of Variations*. Chicago, University of Chicago Press, 1946.

[14] A. M. Bloch, J. Baillieul, P. E. Crouch, and J. E. Marsden. *Nonholonomic mechanics and control*. Springer, 2003.

[15] Anthony M. Bloch and Peter E Crouch. Constrained variational principles on manifolds. In *Proceedings of the 38th IEEE Conference on Decision and Control*, volume 1, pages 1–6, 1999.

[16] Anthony M. Bloch, Peter E. Crouch, Jerrold E. Marsden, and Tudor S. Ratiu. The symmetric representation of the rigid body equations and their discretization. *Nonlinearity*, 15:1309–1341, 2002.

[17] Frederic Bonnans, Philippe Chartier, and Housnaa Zidani. Discrete approximations of the Hamilton-Jacobi equation for an optimal control problem of a differential-algebraic system. *Institut National de Recherche en Informatique et en Automatique (INRIA), Rapport de Recherche*, 4265, 2001.

[18] R. W. Brockett. Control theory and singular Riemannian geometry. In *New directions in applied mathematics*, pages 11–27. Springer-Verlag, 1982.

[19] Arthur E. Bryson and Yu-Chi Ho. *Applied optimal control : optimization, estimation, and control*. Halsted Press, revised edition, 1975.

[20] C. J. Budd and A. Iserles. Geometric integration: Numerical solution of differential equations on manifolds. In *Phil. Trans Royal Soc. A*, volume 357, pages 945–956, 1999.

[21] P. J. Channell and J. C. Scovel. Symplectic integration of Hamiltonian systems. *Nonlinearity*, 3:231–259, 1990.

[22] Nikolai G. Chetaev. *Theoretical mechanics*. Mir Publishers, revised edition, 1989.

[23] Vladimir A. Chobotov. *Orbital mechanics*. American Institute of Aeronautics and Astronautics, 3nd edition, 2003.

[24] Peter Colwell. *Solving Kepler's equation over three centuries*. Richmond, Va. : Willmann-Bell, 1993.

[25] Juergen Ehlers and Ezra T. Newman. The theory of caustics and wavefront singularities with physical applications. *Journal of Mathematical Physics A*, 41(6):3344–3378, 2000.

[26] Zhong Ge and Dau-Liu Wang. On the invariance of generating functions for symplectic transformations. *Differential geometry and its applications*, 5:59–69, 1995.

[27] Herbert Goldstein. *Classical Mechanics*. Addison-Wesley, 2nd edition, 1980.

[28] Donald T. Greenwood. *Classical Dynamics*. Prentice-Hall, 1977.

[29] John Gregory and Cantian Lin. *Constrained optimization in the calculus of variations and optimal control theory*. New York: Van Nostrand Reinhold, 1992.

[30] Vincent M. Guibout and Anthony M. Bloch. A discrete maximum principle for solving optimal control problems. In *Proceedings of the $43rd$ IEEE Conference on Decision and Control, Bahamas, December 2004*, 2004.

[31] Vincent M. Guibout and Anthony M. Bloch. Discrete variational principles and Hamilton-Jacobi theory for mechanical systems and optimal control problems. *Physica D, submitted*, 2004.

[32] Vincent M. Guibout and Daniel J. Scheeres. Formation flight with generating functions: Solving the relative boundary value problem. In *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Monterey, California. Paper AIAA 2002-4639*. AIAA, 2002.

[33] Vincent M. Guibout and Daniel J. Scheeres. Periodic orbits from generating functions. In *Proceedings of the Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-566*. AAS, 2003.

[34] Vincent M. Guibout and Daniel J. Scheeres. Solving relative two-point boundary value problems: Spacecraft formation flight transfers application. *AIAA, Journal of Control, Guidance and Dynamics*, 27(4):693–704, 2003.

[35] Vincent M. Guibout and Daniel J. Scheeres. Computing the generating functions to solve two-point boundary value problems. *Submitted to the Journal of Aerospace Computing, Information, and Communication*, 2004.

[36] Vincent M. Guibout and Daniel J. Scheeres. New techniques for spacecraft formation design and control. In *Proceedings of the 18^th International Symposium on Space Flight Dynamics, Munich*. AIAA, 2004.

[37] Vincent M. Guibout and Daniel J. Scheeres. Solving two-point boundary value problems using generating functions: Theory and applications to optimal control and the study of Hamiltonian dynamical systems. *Submitted to the Journal of nonlinear science*, 2004.

[38] Vincent M. Guibout and Daniel J. Scheeres. Spacecraft formation dynamics and design. In *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Providence, Rhode Island*, 2004.

[39] H. Y. Guo, Y. Q. Li, K. Wu, and S. Wang. Difference discrete variational principle Euler-lagrange cohomology and symplectic, multisymplectic structures i: Difference discrete variational principle. *Communications in theoretical physics*, 37(1):1–10, 2002.

[40] H. Y. Guo, Y. Q. Li, K. Wu, and S. Wang. Difference discrete variational principle Euler-Lagrange cohomology and symplectic, multisymplectic structures ii: Euler-Lagrange cohomology. *Communications in theoretical physics*, 37(2):129–138, 2002.

[41] H. Y. Guo, Y. Q. Li, K. Wu, and S. Wang. Difference discrete variational principle Euler-Lagrange cohomology and symplectic, multisymplectic structures iii: Applications to symplectic and multisymplectic algorithms. *Communications in theoretical physics*, 37(3):257–264, 2002.

[42] Pini Gurfil and N. Jeremy Kasdin. Nonlinear modeling and control of spacecraft formation dynamics in the configuration space. *Submitted to the Journal of Guidance, Control and Dynamics*, 2002.

[43] Ernst Hairer and Christian Lubich. Energy conservation by Störmer-type numerical integrators. In D.F. Griffiths and G.A. Watson, editors, *Numerical analysis 1999*, volume 420 of *Research Notes in Mathematics Series*, pages 169–190. CRC Press LLC, 2000.

[44] Ernst Hairer and Christian Lubich. Long-time energy conservation of numerical methods for oscillatory differential equations. *SIAM journal on numerical analysis*, 38:414–441, 2001.

[45] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration. Structure-preserving algorithms for ordinary differential equations*. Springer, 2002.

[46] William Rowan Hamilton. On a general method in dynamics. *Philosophical Transactions of the Royal Society, Part II*, pages 247–308, 1834.

[47] William Rowan Hamilton. Second essay on a general method in dynamics. *Philosophical Transactions of the Royal Society, Part I*, pages 95–144, 1835.

[48] Michel Hénon. *Generating families in the restricted three-body problem*. Springer-Verlag, 1997.

[49] Magnus R. Hestenes. *Calculus of Variations and Optimal Control Theory*. New York, Wiley, 1966.

[50] Alan S. Hope and Aaron J. Trask. Pulsed thrust method for hover formation flying. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-655*. AAS, 2003.

[51] K. C. Howell and B. G. Marchand. Control strategies for formation flight in the vicinity of a libration point orbit. In *Proceedings of the* $13th$ *AAS/AIAA Space Flight Mechanics Meeting, Ponce, Puerto Rico, AAS Paper No. 03-113*, 2003.

[52] F. Y. Hsiao and Daniel J. Scheeres. Design of spacecraft formation orbits relative to a stabilized trajectory. In *Proceedings of the AAS/AIAA Space Flight Mechanics meeting, Ponce, Puerto Rico. Paper AAS 03-175*. AAS, 2003.

[53] Islam Hussein, Daniel J. Scheeres, and D. C. Hyland. Control of a satellite formation for imaging applications. In *Proceedings of the American Control Conference, Denver, Colorado*. AAS, 2003.

[54] S. Jalnapurkar, S. Pekarsky, and M. West. Discrete variational mechanics on cotangent bundles. *Unpublished working notes*, 2000.

[55] B. W. Jordan and E. Polak. Theory of a class of discrete optimal control systems. *Journal of Electronics and Control*, 17:697–711, 1964.

[56] C. Kane, Jerrold E. Marsden, and M. Ortiz. Symplectic-energy-momentum preserving variational integrators. *Journal of mathematical physics*, 40(7):3353–3371, 1999.

[57] Feng Kang. Difference schemes for Hamiltonian formalism and symplectic geometry. *Journal of Computational Mathematics*, 4(3):279–289, 1986.

[58] H. B. Keller. *Numerical methods for two-point boundary value problems*. Blaisdell, 1968.

[59] W. S. Koon, Jerrold E. Marsden, J. Masdemont, and R. M. Murray. J_2 dynamics and formation flight. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference, Montreal, Canada, August, AIAA 2001-4090*. AIAA, 2001.

[60] Cornelius Lanczos. *The variational principles of mechanics*. University of Toronto Press, 4th edition, 1977.

[61] Pierre-Louis Lions. *Generalized solutions of Hamilton-Jacobi equations*. Pitman, 1982.

[62] T. A. Lovell, K. R. Horneman, M. V. Tollefson, and S. G. Tragesser. A guidance algorithm for formation reconfiguration and maintenance based on perturbed Clohessy-Wiltshire equations. In *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-649*. AAS, 2003.

[63] R. S. MacKay. Some aspects of the dynamics and numerics of Hamiltonian systems. In *The dynamics of numerics and the numerics of dynamics*, pages 137–193. Oxford Univ. Press, 1990.

[64] Jerry Markman and I. Norman Katz. An iterative algorithm for solving Hamilton-Jacobi type equations. *SIAM Journal on Scientific Computing*, 22(1):312–329, 2000.

[65] Jerrold E. Marsden, George W. Patrick, and Steve Shkoller. Multisymplectic geometry, variational integrators, and nonlinear PDEs. *Comm. Math. Phys.*, 199:351–395, 1998.

[66] Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems*. Springer-Verlag, 2nd edition, 1998.

[67] Jerrold E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numerica*, pages 357–514, 2001.

[68] Robert McLachlan and Reinout Quispel. Six lectures on the geometric integration. In *Foundations of Computational Mathematics, ed. R. DeVore, A. Iserles, E. Sli*, pages 155–210. Cambridge University Press, 2001.

[69] David Mishne. Maintaining periodic relative trajectories of satellite formation by using power-limited thrusters. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-656*. AAS, 2003.

[70] F. H. Molzahn and T. A. Osborn. Tree graphs and the solution to the Hamilton-Jacobi equation. *Journal of mathematics physics*, 27(1):88–99, 1986.

[71] J. Moser and A. P. Veselov. Discrete versions of some classical integrable systems and factorization of matrix polynomials. *Comm. Math. Phys.*, 139:217–243, 1991.

[72] Forest R. Moulton. *Differential equations*. The Macmillan company, 1930.

[73] N. N. Newmark. A method of computation for structural dynamics. *ASCE Journal of the Engineering Mechanics Division*, 85:67–94, 1959.

[74] C. Park, Vincent M. Guibout, and Daniel J. Scheeres. Solving optimal low-thrust rendezvous problems with generating functions. *In preparation*, 2004.

[75] C. Park and Daniel J. Scheeres. Indirect solutions of the optimal feedback control using Hamiltonian dynamics and generating functions. In *Proceedings of the 2003 IEEE conference on Decision and Control, accepted, 2003. Maui, Hawaii*. IEEE, 2003.

[76] C. Park and Daniel J. Scheeres. Solutions of optimal feedback control problems with general boundary conditions using hamiltonian dynamics and generating functions. In *Proceedings of the American Control Conference, Boston, Massachusetts, June 2004. Paper WeM02.1*, 2004.

[77] Henri Poincaré. *Les Méthodes Nouvelles de la Mécanique Céleste*, volume 1,2,3. Gauthier-Villars, Paris; reprinted by Dover, New York (1957), 1892,1893,1899.

[78] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, and E.F. Mishchenko. *The mathematical theory of optimal processes*, volume 4. New York : Gordon and Breach Science Publishers, 1986.

[79] David L. Powers. *Boundary value problems*. San Diego : Harcourt Brace Jovanovich, 1987.

[80] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C, the art of scientific computing*. CAMBRIDGE UNIVERSITY PRESS, second edition, 1992.

[81] Hanno Rund. *The Hamilton-Jacobi Theory in the Calculus of Variations; its Role in Mathematics and Physics*. London, New York, Van Nostrand, 1966.

[82] Hans Sagan. *Introduction to the calculus of variations*. New York : Dover Publications, 1992.

[83] Noboru Sakamoto. Analysis of the hamilton-jacobi equation in nonlinear control theory by symplectic geometry. *SIAM journal on control and optimization*, 40(6):1924–1937, 2002.

[84] J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian Problems*. Chapman & Hall, 1994.

[85] Daniel J. Scheeres, F. Y. Hsiao, and N. X. Vinh. Stabilizing motion relative to an unstable orbit: Applications to spacecraft formation flight. *Journal of Guidance, Control, and Dynamics*, 26(1):62–73, 2003.

[86] Daniel J. Scheeres, C. Park, and Vincent M. Guibout. Solving optimal control problems with generating functions. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-575*. AAS, 2003.

[87] Y. Shibberu. *Discrete-time Hamiltonian dynamics*. PhD thesis, University of Texas at Arlington, 1992.

[88] J. C. Simo and O. Gonzalez. Assessment of energy-momentum and symplectic schemes for stiff dynamical systems. In *Proceedings of the ASME Winter Annual Meeting, New Orleans*, 1993.

[89] Jürgen Struckmeier and Claus Riedel. Canonical transformations and exact invariants for time-dependent hamiltonian systems. *Annalen der Physik (Leipzig)*, 11(1):15–38, 2002.

[90] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76:637–642, 1982.

[91] S. R. Vadali, H. Bae, and K. T. Alfriend. Control of libration point satellite formations. In *Proceedings of the $14th$ Space Flight Mechanics Meeting, Maui, Hawaii, AAS Paper No. 034-161*, 2004.

[92] S. R. Vadali, S. S. Vaddi, and K.T. Alfriend. An intelligent control concept for formation flying satellite constellations. *International journal of robust and nonlinear control*, 12(2-3):97–115, 2002.

[93] S. S. Vaddi, Kyle T. Alfriend, and S. R. Vadali. Sub-optimal formation establishment and reconfiguration using impulsive thrust. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-590*. AAS, 2003.

[94] P. K. C. Wang and F. Y. Hadaegh. Minimum-fuel formation reconfiguration of multiple free-flying spacecraft. *The Journal of the Astronautical Sciences*, 47(1-2):77–102, 1999.

[95] Alain Weinstein. Lectures on symplectic manifolds. *Regional conference series in mathematics*, 29, 1977.

[96] Jeffrey M. Wendlandt and Jerrold E. Marsden. Mechanical integrators derived from a discrete variational principle. *Physica D*, 106:223–246, 1997.

[97] Jack Wisdom and Matthew Holman. Symplectic maps for the n-body problem. *The Astronomical journal*, 102(4):1528–1538, 1991.

[98] Jack Wisdom and Matthew Holman. Symplectic maps for the n-body problem: Stability analysis. *The Astronomical journal*, 104(5):2022–2029, 1992.

[99] Y. Wu. The discrete variational approach to the Euler-Lagrange equation. *Computers Math. Applic.*, 20(8):63–75, 1987.

[100] Y. Xu and N. Fitz-Coy. Genetic algorithm based sliding method control in the leader / follower satellites pair maintenance. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-648*. AAS, 2003.

ABSTRACT

THE HAMILTON-JACOBI THEORY FOR SOLVING TWO-POINT BOUNDARY VALUE PROBLEMS: THEORY AND NUMERICS WITH APPLICATION TO SPACECRAFT FORMATION FLIGHT, OPTIMAL CONTROL AND THE STUDY OF PHASE SPACE STRUCTURE

by

Vincent M. Guibout

Co-Chairs: Daniel J. Scheeres and Anthony M. Bloch

This dissertation has been motivated by the need for new methods to address complex problems that arise in spacecraft formation design. As a direct result of this motivation, a general methodology for solving two-point boundary value problems for Hamiltonian systems has been found. Using the Hamilton-Jacobi theory in conjunction with the canonical transformation induced by the phase flow, it is shown that generating functions solve two-point boundary value problems. Traditional techniques for addressing these problems are iterative and require an initial guess. The method presented in this dissertation solves boundary value problems at the cost of a single function evaluation, although it requires knowledge of at least one generating function. Properties of this method are presented. Specifically, we show that it includes perturbation theory and generalizes it to nonlinear systems. Most importantly, it predicts the existence of multiple solutions and allows one to recover all of these solutions.

To demonstrate the efficiency of this approach, an algorithm for computing the generating functions is proposed and its convergence properties are studied. As the method developed in this work is based on the Hamiltonian structure of the problem, particular attention must be paid to the numerics of the algorithm. To address this, a general framework for studying the discretization of certain dynamical systems is developed. This framework generalizes earlier work on discretization of Lagrangian and Hamiltonian systems on tangent and cotangent bundles respectively. In addition, it provides new insights into some symplectic integrators and leads to a new discrete Hamilton-Jacobi theory. Most importantly, it allows one to discretize optimal control problems. In particular, a discrete maximum principle is presented.

This dissertation also investigates applications of the proposed method to solve two-point boundary value problems. In particular, new techniques for designing spacecraft formation flight, reconfiguring a formation, and searching for stable configurations in a general dynamical environment are presented. In addition, the present work allows one to reduce the search for periodic orbits with specified periods or locations to solving a set of nonlinear equations. Finally, a novel approach for solving optimal control problems is derived and applied.